Volume 6, December 2023 ISSN: 2591 - 801X

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

DOI: https://doi.org/10.37074/jalt.2023.6.2

Editorial Board

Contents

Editorial

Commentary

Research articles

Contents

Contents

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Editorial 6(2): Personal digital assistant or job killer? Generative AI and the teaching profession in higher education

Jürgen Rudolph[A]
Shannon Tan[B]
Tania Aspland[C]

A Director of Research, Kaplan Singapore
B Research Executive, Kaplan Singapore
C Professor Emerita & Vice President, Academic, Kaplan Australia and New Zealand

DOI: https://doi.org/10.37074/jalt.2023.6.2.1

Introduction

2023 was a busy year for our Journal of Applied Learning and Teaching (JALT). After our inclusion in Scopus in 2022, we were pleasantly surprised that, according to Scopus, we are in the top quartile and even the top ten per cent of education journals according to their CiteScore metric. A JALT article was cited an average of 6.3 times from 2019 - 2022 and, at the time of writing (27 November 2023), 9.2 times in the period of 2020 to 2023. While we are critical of the relentless neoliberal metrification of academic research (Fleming et al., 2021; Brookfield et al., 2023), it is nice to be the new popular kid on the block.

Fame, however, comes with its own problems. In the early years, Jürgen (the first author of this Editorial) distinctly remembers that, with varying success, he begged, pleaded, and cajoled authors into submitting an article to JALT. In these fledgling years, we often found the time to mentor and coach first-time authors and early career researchers extensively (which sometimes included proofreading and reference-fixing, occasionally almost amounting to co-authorship). But things have changed. In the second half of 2023, we have received more than two articles a day on average, and our rejection rate has shot up to more than 90 per cent.

We wish more authors would read and heed the advice on academic writing of excellent authors such as Helen Sword (2012, 2017, 2023) and Dannelle Stevens (2019). Stevens (2019) accurately outlined the primary reasons why journal editors reject articles, including (1) a mismatch with the journal’s scope or objectives, (2) inappropriate article format such as being overly journalistic, (3) unsuitable length, (4) non-adherence to journal or academic writing standards, (5) poor language usage including grammatical and punctuation errors, (6) lack of significant content or prolix elaboration of obvious points, (7) inadequate contextualisation for an international readership, (8) weak theoretical framework, (9) shoddy presentation with apparent lack of proofreading, and (10) inclusion of libellous or unethical content.

If 2000 was the year of COVID-19-related research in higher education, 2023 was the year of generative AI. In January, we published one of the first substantial journal articles on ChatGPT and higher education (Rudolph et al., 2023a). Since then, we have fostered a rich dialogue by publishing more than 20 substantial pieces on this pressing subject. Our journal has become a focal point for critical discussions encompassing the potential advantages, challenges, and actionable recommendations concerning the implementation of generative AI technologies such as ChatGPT, as documented in a series of articles (Firat, 2023; Gamage et al., 2023; Ifelebuegu et al., 2023; Limna et al., 2023; Rasul et al., 2023; Rudolph et al., 2023a, 2023b). Furthermore, our contributors have rigorously explored the ethical dimensions surrounding academic integrity and student engagements with AI tools during assessments, providing nuanced analyses of this complex issue (Chaka, 2023; Hassoulas et al., 2023; Ifelebuegu, 2023; Mohammadkarimi et al., 2023; Sullivan et al., 2023). Beyond these spheres, our repertoire has expanded to include discourse on the transformative role of AI in research (Khademi, 2023) and pedagogy (Xames & Shefa, 2023), its application in teaching numerically intensive subjects (Calonge et al., 2023), and the urgency of bridging the awareness gap on generative AI advancements in African countries such as Ghana (Adarkwah et al., 2023). Moreover, we have published Huang et al.’s (2023) framework for machine-human collaboration in educational settings, a methodological paper on an instrument measuring attitudes, benefits and threats toward using AI in higher education (Ahmad et al., 2023), an article on the transparency level of literature reviews on AI in education (Tlili et al., 2023) and two protocol papers for such literature reviews (Ismail et al., 2023; Stracke et al., 2023). In addition, we have published opinion pieces promoting open educational strategies for AI integration (Mills et al., 2023) and presenting critical perspectives that interrogate the influence and trajectory of AI in the higher education sector (Popenici, 2023b; see Popenici et al., 2023). Through these efforts, we are steadfast in our commitment to fostering a multifaceted scholarly conversation that assesses the evolving landscape of higher education in the AI era.

In our first Editorial this year, we explored graduate employability in the age of generative AI (Rudolph et al., 2023c). Throughout history, societal attitudes towards work have varied greatly, with the poor often subjected to harsh conditions and compulsory labour to maintain order and prevent idleness, as documented in medieval British laws and corroborated by thinkers such as George Orwell (1933). In stark contrast, the affluent classes have sometimes been discouraged from labour, with figures like Bertrand Russell (2004) advocating for reduced work hours to foster personal and civilisational growth through engagement in arts and sciences. In modern discourse, concepts like Fully Automated Luxury Communism (Bastani, 2020) foresee a future where technological advancements significantly reduce or eliminate the necessity for human labour, a vision echoed in various mythologies where work is seen as a form of punishment imposed following a fall from an idyllic state of existence (Rudolph et al., 2023c).

In our aforementioned Editorial, we traced the historical perceptions of work from the times of the Protestant work ethic to the 20th-century revaluation of labour as a form of self-sacrifice (Rudolph et al., 2023c). We spotlighted the stark income disparities of 'bullshit' and 'shit' jobs, as conceptualised by David Graeber (2018), amid rising automation and the advancements in generative AI technologies that threaten to reshape the global job market significantly. The analysis underscores the urgent need to rethink work and life paradigms in the face of potential mass job displacements, exploring alternatives such as Universal Basic Income while casting a critical eye on the impact of AI technologies on education and various professional sectors (Susskind, 2021; Rudolph et al., 2023c).

Large Language Models (LLMs), particularly since the release of ChatGPT in November 2022, have presented both unprecedented opportunities and challenges in knowledge-intensive professions. The latest generative AI systems have demonstrated rapid advancements, possessing unexpected and expanding capabilities, including proficiency in complex, creative, and analytical tasks, a domain previously preserved for highly skilled professionals (Eloundou et al., 2023; Geerling et al., 2023). These developments have intensified the urgency among scholars, organisations, and governments to comprehend the implications of integrating AI into existing work frameworks (Berg et al., 2023).

Dell'Acqua et al.'s (2023) study focuses on navigating the 'jagged frontier', a term coined to describe the unpredictable boundary where AI may or may not enhance human performance. The capabilities of AI resemble an invisible fortress wall with uneven battlements, where the AI effortlessly performs certain tasks perceived as complex, such as writing sonnets or idea generation, while surprisingly struggling with ostensibly simpler tasks, such as crafting a 50-word poem or basic mathematical calculations, illustrating a nuanced and unpredictable proficiency landscape within the current boundaries of generative AI (Dell'Acqua et al., 2023). This jagged frontier reveals that AI can significantly augment productivity and quality in certain tasks, reshaping traditional workflows of high human capital professionals. However, it is constantly evolving, making it challenging for professionals to accurately identify the tasks where AI can be beneficial. Furthermore, the opacity of these systems, including their unclear failure points and unexpected abilities, compounds the difficulty in fully grasping their potential and downsides for knowledge work (Dell'Acqua et al., 2023). In Dell'Acqua's (2023) empirical study, consultants using GPT-4 finished 12% more tasks on average, completed tasks 25% more quickly, and produced 40% higher quality results than those without. As generative AI's jagged frontier expands, it is anticipated that AI will have a substantial, albeit uneven, impact on work, necessitating ongoing research to understand how human-AI interaction dynamics will change over time (Noy & Zhang, 2023).

Will generative AI spell the end of the teaching profession?

Before we give our answer to this question, it is worthwhile noting that there have been various trends that have been eroding the teaching profession in higher education prior to the popularisation of ChatGPT and related generative AI. In *The fall of the faculty*, Ginsberg (2011, p. 2) describes the expansion of non-academic personnel via-a-vis academics and bemoans that universities are increasingly "filled with armies of functionaries – the vice presidents, associate vice presidents, assistant vice presidents, provosts, associate provosts, vice provosts, assistant provosts, deans, deanlets, deanlings, each commanding staffers and assistants – who, more and more, direct the operations of every school". Universities across the world in the early 21st century find themselves in a paradoxical position: "Never before in human history have they been so numerous or so important, yet never before have they suffered from such a disabling lack of confidence and loss of identity" (Collini, 2012, p. 5).

Fleming's *Dark academia. How universities die* (2021) discusses the demise of *homo academicus* also in a literal way. One particularly poignant example of the "proletariatisation of academic labour" is the death of 83-year-old adjunct professor Margaret Vojtko who had an onerous workload but barely earned US$25,000 with no healthcare benefits (Fleming, 2021, p. 92). After she was diagnosed with cancer, her health deteriorated, and her ostensibly Catholic employer dismissed her. Medical bills mounted, medicine and electricity ran out, and Prof Vojtko eventually died a lonely death (Fleming, 2021).

In *Artificial Intelligence and learning futures* – a book reviewed in this issue of JALT – Popenici (2023a) observes the denigration of the teaching profession in higher education that goes hand in hand with the devaluation of learning in the neoliberal paradigm, which is obsessed with performance-based accountability and return-on-investments. Consequently, the once cherished ethos of intellectual exploration is unceremoniously side-lined and supplanted with a myopic focus on test performances and tangible outcomes. Higher education's identity crisis is accompanied by an even graver crisis of learning, where students are driven not by the joyous quest for knowledge but by the pressures of conforming to metric determinants. This erosion of learning – where vibrant curiosity is replaced by crude instrumentalism and sloganeering – stands as a testament to the destructive potential of commodifying education (Popenici, 2023a).

Martin Andrew (2023) tells the story of a learner who discovered her online professor had been dead since 2019. Since the pandemic, there has been an increase in the precarity and casualisation of academic labour, which is undervalued, overused, and stigmatised (Solomon & Du Plessis, 2023). The above-described trends of the fall of the faculty (Ginsberg, 2011), the loss of faith in university education (Collini, 2012), the demise of *homo academicus* (Fleming, 2021), the denigration of the teaching profession and the devaluation of learning (Popenici, 2023a) all precede the rise of generative AI in higher education.

Predictions of technology such as (ro)bots or AI replacing teachers are not new. Throughout the history of educational technology, repeated cycles of high hopes followed by modest impacts underscore the importance of a grounded approach to educational innovation. From Edison's predictions about motion pictures replacing books (Terzian, 2019) to the envisioned role of radio as a 'Master Teacher' (Cook, 1938), television's educational promises (Terzian, 2019), and the advent and challenges of computers in classrooms (Watters, 2021), technology's touted transformative potential often outstripped its actual influence. This recurring optimism, combined with commercial interests, suggests a more complex, non-linear progression of educational technology. Bror Saxberg once quipped that "Technology is just technology" (cited in Rudolph, 2014). We need to avoid both uncritical adoption and outright rejection and acknowledge that no technological 'miracle cure' for higher education exists (Rudolph, 2018; Kefalaki et al., 2021).

Even before the current generative AI hype, there have been predictions that robots (taking the form of AI software programs or humanoid machines) will replace human teachers by 2027 (Houser, 2017). With the recent launches of ChatGPT and a host of other generative AI software, the capabilities of AI technologies appear to be quickly increasing. The debate on AI potentially substituting teachers is intensifying, with the prospect appearing increasingly likely and the media actively discussing this potential shift (Chan & Tsi, 2023; Devlin, 2023). Replacing higher education teachers with machines could be motivated by financial difficulties faced by universities, caused, for instance, by "eye-watering mortgages for shiny new teaching buildings" (Haw, 2019) – Haw (2019) worried that "swapping expensive lecturers for cheap, versatile machines that don't go on strike, don't need sleep, and respond to students within nanoseconds will be hard to resist".

In contrast, a study by the World Economic Forum (2023, p. 6) predicts that by 2027, jobs in the education industry are "expected to grow by about 10%, leading to 3 million additional jobs" for teachers in vocational education and higher education. Predictions about the future are notoriously unreliable. While the above WEF forecast heartens us, we reckon that higher education teachers' full benefit from generative AI will depend on their access to good-enough digital devices, fast internet access, educational technology training and institutional policies. In the near future, many knowledge workers may have an AI 'co-pilot'. Perhaps Stephen Brookfield's tongue-in-cheek 'law of employment' will continue to apply: "act as if you assume you're going to be fired – and you probably won't be" (Brookfield et al., 2023, p. 185).

Generative AI as teacher's assistant

The developments in the generative AI space are progressing at a dizzying speed, and the following thoughts about how teachers can benefit from using generative AI will consequently require constant updating. It should, however, be obvious that provided that teachers know how to use generative AI in a critically informed way, substantial productivity gains are possible.

Higher education teachers can use generative AI for brainstorming like other knowledge workers. Depending on the appropriateness of the results for the teacher's purposes, prompts can be refined and repeated, thus churning out multiple ideas within a few minutes. It is noteworthy that GPT-4, in particular, scored very highly on various creativity tests, exceeding 91% of humans on an Alternative Uses Test for creativity and 99% on the Torrance Tests of Creative Thinking (Haase & Hanel, 2023; Shimek, 2023).

When getting ready to delve into a new subject or staying current with recent literature, tools like Chat PDF and Claude 2 may be beneficial for summarising and analysing articles or books. Claude 2 users can input up to 100,000 tokens (equivalent to 75,000 words or hundreds of pages of technical documentation or a book) in a single prompt (Anthropic, 2023). While in the ideal world, we may prefer to read everything by ourselves, sometimes it may not be possible, and then these tools provide a solution superior to not reading. For instance, Jürgen has asked GPT-4 to organise and summarise student feedback based on Stephen Brookfield's Critical Incident Questionnaire (Brookfield et al., 2023). While it may take a human teacher an hour to organise and summarise 100 student responses, GPT-4 can do this in less than a minute. Then, the teacher can spend a couple of minutes editing and double-checking the text before sharing it with the students.

Another popular use of GPT-4 and other chatbots is the drafting of coherent text based on one's own notes or extracts from other sources. GPT-4 and other chatbots can produce drafts for blog posts, essays, speeches, lectures, scripts, and other texts. These texts can improve with a good prompt (we have found it quite useful to tell GPT-4 that it is a Professor of Higher Education when asking it to draft text). Generative AI can be used to suggest how to improve our texts and employ different academic writing styles. It can also be used to combat writer's block and to draft emails.

Combining the functions of brainstorming, summarising, and drafting may lead to significant productivity gains not only in administrative and research-related work but also in teaching-related processes such as creating teaching and learning activities, lesson plans and curricula. Generative AI can also be used for data visualisation – creating figures, charts and graphs. Various AI-driven tools, such as Appy Pie's Free AI Graph Maker, Chartify.ai, and Graph Maker, allow users to produce custom graphs quickly without the need for coding skills. Another tool, VizGPT, provides a chat interface for users to generate and modify data visualisations using natural language queries (Mikami, 2023). We should also mention that there are more than 700 plugins for GPT-4 – while we have not tried them all, there are a few which appear to be particularly useful: for instance, Wolfram for mathematics, Vox Script and Video Insight for summarising long YouTube videos, Show Me for creating diagrams, Zapier for automating workflows, and Ask Your Pdf for analysing long texts.

For PowerPoint presentations, instead of time-consuming searches for Creative Commons images, one can use Midjourney's "describe" feature. By uploading a base image to Midjourney, the software suggests prompts for similar visuals. Tweaking these prompts can produce unique and intriguing images quickly, enhancing the presentation's appeal (Mollick, 2023a). To exemplify productivity gains with another Microsoft application, Excel, GPT-4's Code Interpreter tool can be used to craft a five-year revenue projection for a hypothetical startup in a usable CSV file that is easy to verify (Mollick, 2023a). Microsoft has recently incorporated Copilots into the premium version of its office-work software, Microsoft 365, and the lines between what humans and AI do will blur further. They may transform their users into virtual cyborgs.

There is also a fast-increasing number of AI tools for video creation. They can be differentiated into three broad categories: Video editors with AI editing tools, generative text-to-video apps, and video productivity apps (that create content for multiple marketing channels and platforms) (Rebelo, 2023). For instance, Runway can be used to experiment with generative AI and Visla to turn a script into a video (Rebelo, 2023).

Teachers can consider using generative AI intentionally as a 24/7 virtual tutor, which helps students practise their skills, for instance, when learning a new language (Ifelebuegu et al., 2023). More generally, Mollick (2023b) has seen seven different types of generative AI applications in the classroom: "AI-tutor, for increasing knowledge, AI-coach for increasing metacognition, AI-mentor to provide balanced, ongoing feedback, AI-teammate to increase collaborative intelligence, AI-tool for extending student performance, AI-simulator to help with practice, and AI-student to check for understanding" (see Mollick & Mollick, 2023).

A generative AI function that we remain sceptical about is grading students' assignments and providing feedback (Baidoo-Anu & Owusu Ansah, 2023). While students can use generative AI, such as GPT-4, to seek feedback on their work, it could also be used for formative assessments. However, we think that marking and grading students' work (unless it is programmable tasks such as multiple-choice questions) should remain the domain of human teachers. Are these the famous last words?

Finally, the conversational character of generative AI, such as GPT-4, may be helpful as it is good for dialogically thinking through one's ideas. While we are fully aware that generative AI is *not* sentient and of our tendency to anthropomorphise chatbots, it is ironic that precisely this anthropomorphising with a 24/7 digital personal assistant can be fruitful. Higher education teachers may achieve significant productivity gains provided they have access to the right hardware and software, good Internet speeds and training and tech support. However, if we blindly take what generative AI offers, there is a high chance that it will be bad or mediocre at best. Teachers' and students' critical thinking remains of key importance. We must never outsource critical thinking to generative AI.

Generative AI tools, like ChatGPT, threaten to disrupt education. However, this may not be due to their intelligence but rather our flawed education systems that undervalue human intelligence (Luckin, 2023). Mistakenly, generative AI is perceived as more intelligent than it is; it lacks understanding, merely producing text based on probabilities (Chomsky et al., 2023). Its assessment performance reveals the tests' focus on information memorisation over knowledge comprehension. To outpace AI, education must evolve, emphasising human intelligence's uniqueness. Instead of mere rote learning, curricula should prioritise critical thinking and interpretation within traditional subjects while integrating critical AI literacy. To ensure AI enhances our lives, we must challenge tech giants' profit motives, discern which intellectual tasks we delegate to AI, and safeguard our unique human traits for future generations.

Overview of the issue

Our issue kicks off with Martin Andrew's invited Commentary "Come to the Cabaret: Voices from the modern university". In his creative contribution that pushes the boundaries of traditional academic writing conventions, Andrew's article creates his own cabaret songs that reflect on the modern university, drawing inspiration from the satirical Kabarett performances of the Weimar Republic in 1920s Germany. Using poetic enquiry, the study contrasts the university's contemporary culture with historical expressions, particularly the subversive tones of the Kabarett. The research employs a critical lens reminiscent of Puck from Shakespeare's *A Midsummer Night's Dream*, highlighting the absurdities of today's higher education institutions.

Andrew's Commentary is followed by 24 research articles, ranging from topics such as the impact of (generative) AI on higher education to contract cheating, academic dishonesty, student resilience, international student employability, learning styles, teaching method preferences, a critically reflective teacher journey, a combination of design thinking and project-based learning, explorations of NVIVO (a qualitative data analysis software) and UTAUT2 (the second iteration of the Unified Theory of Acceptance and Use of Technology) and students' multidimensional learning outcomes.

We start the research article section with nine manuscripts on the hot topic of AI in higher education. First, Huang et al.'s paper, "Educational futures of intelligent synergies between humans, digital twins, avatars, and robots – The iSTAR framework", presents the innovative iSTAR framework, a pivotal model for human-machine collaboration in education. This framework, standing for Intelligent human-machine Synergy in collaborative teaching with digital Twins, Avatars/Agents, and Robots, introduces the DELTA dimensions — Design, Ethics, Learning, Teaching, and Assessments — which are instrumental in forging safe, ethical, and responsible learning environments. The iSTAR framework reimagines the relationship between humans and AI in education as a dynamic ecosystem, offering comprehensive guidelines for synergistic interactions between educators and machines.

Second, Tlili et al.'s "Speaking of transparency: Are all Artificial Intelligence (AI) literature reviews in education transparent?" evaluates the transparency of AI-in-education-specific literature reviews. Literature reviews are crucial for generating new theories and trend identification, and their lack of transparency might compromise findings. Tlili et al.'s findings expose methodological gaps and aim to improve AI education research transparency, trustworthiness, and efficacy.

Ismail et al.'s study, "Artificial Intelligence in higher education: A protocol paper for a systematic literature review," complements Tlili et al.'s work by proposing a longitudinal review method for generative AI chatbot research in higher education. This method aims to develop an open-access database for academic use and adaptability across various fields. Similarly, Stracke et al.'s paper aligns with these approaches, introducing a standardised protocol for AI in education (AI&ED) reviews. This protocol, which, like Ismail et al.'s contribution, is grounded in PRISMA guidelines, enhances the reliability and replicability of reviews, focusing on AI's role in learning, teaching, and literacy. It's exemplified through a review of ethical and trustworthy AI&ED literature, with future applications planned for diverse AI&ED areas and longitudinal trend analysis.

A fifth AI-specific study by Hassoulas et al., "Investigating marker accuracy in differentiating between university scripts written by students and those produced using ChatGPT", investigates marker accuracy in differentiating student work from ChatGPT-generated content. OpenAI's ChatGPT is reshaping higher education assessment, prompting varied institutional responses. Markers from a medical school struggled to recognise ChatGPT-generated scripts. Hassoulas et al.'s research underscores the need for responsible AI integration in assessment and redefining academic misconduct.

A sixth AI-related article by Adarkwah et al., "Awareness and acceptance of ChatGPT as a generative conversational AI for transforming education by Ghanaian academics: A two-phase study", highlights the importance of increasing technology awareness among African scholars to harness innovative tech for efficiency. It examines the slow adoption of digital transformation in Ghanaian education using ChatGPT as a case study. The study reveals limited knowledge among Ghanaian academics about ChatGPT and AI-powered chatbots, emphasising the need to promote tech awareness in African countries like Ghana to transition from 'laggards' to 'early adopters' in line with innovation theory. Policymakers and educators are urged to play a role in fostering technological awareness.

In a seventh AI-related study, "Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools", Chaka tested five AI content tools on ChatGPT, YouChat, and Chatsonic responses. Copyleaks AI Content Detector and OpenAI's AI Text Classifier performed best, but GPTZero misidentified translated ChatGPT responses as human responses. Current tools struggle to detect AI-generated content effectively, adversely impacting efforts to combat AI-generated plagiarism.

In an eighth research article on AI and higher education, Mohammadkarimi examines "Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence". The study found mixed perceptions. While some participants acknowledged AI benefits, concerns about academic integrity prevailed. Teachers saw AI negatively affecting honesty and skill growth. The need to detect AI-generated work and address ethics was stressed. Training and support were highlighted to manage AI-related dishonesty, urging institutions and policymakers to establish ethical AI guidelines for higher education.

To assess attitudes towards AI in higher education, a ninth article by Ahmad et al. developed an ABT (Attitudes, Benefits, Threats) instrument and surveyed students and teachers in 11 Asian and African countries. Using Google Forms for data collection, they analysed responses through factor analysis. The preferred model, out of six, explained 55.6% variance and comprised three factors: Attitude (15 items), Benefits (6), and Threats (14). The model's reliability and validity were confirmed for evaluating attitudes towards AI tools in an educated demographic.

Lawson and Martella's article, "Critically reflecting on the use of Immersive Virtual Reality in educational settings: What is known and what has yet to be shown?" shifts the focus from AI to the burgeoning field of immersive virtual reality (IVR) in education. The authors delve into the increasing global interest in IVR, highlighting its affective impact, notably in boosting student motivation and its debated cognitive benefits, with mixed results in learning effectiveness. Lawson and Martella's reflective piece underscores the research void in IVR's pedagogical applications and advocates for more comprehensive studies to resolve its inconsistent educational outcomes. The goal is to refine IVR's integration into educational frameworks.

The next article by Gamage et al., "Contract cheating in higher education: Impacts on academic standards and quality," switches the focus from using AI for cheating purposes (with reference to Mohammadkarimi's earlier-discussed piece in this issue) to that of humans. Gamage et al. explore students' motivations, deterrents for contract cheating, and assessment design's impact on authentic learning. As universities shifted to online learning and assessments during the recent COVID-19 pandemic, global contract cheating has been on the rise. Despite preventive measures like authentic assessments and tools, no single solution is able to guarantee academic integrity. Gamage and co-authors argue that a global movement is needed to address this ongoing issue. Relatedly, Goegan et al.'s study "Preservice secondary teachers' beliefs about academic dishonesty: An attribution theory lens to causal search" delves into academic dishonesty among preservice secondary teachers. Context influences their perceptions of dishonesty, with descriptive scenarios evoking more robust responses than isolated behaviours. Goegan and co-authors' research sheds light on the less-explored 'why' behind academic dishonesty.

Owan et al.'s contribution discusses their Persistence to Publish Questionnaire (PPQ), which offers a reliable measure of academics' persistence in publishing in Scopus-indexed journals. Created through a thorough process including content validity and pilot testing, the PPQ was tested with 262 academics from various fields at two Nigerian universities. It uses Exploratory and Confirmatory Factor Analysis to identify five factors: manuscript preparation, submission, revisions, handling rejections, and dealing with delays. Demonstrating strong internal consistency and construct validity, the PPQ may serve as a valuable tool for enhancing research productivity and quality in academia.

Ahmed-Shafi et al.'s research article, titled "Learning in a disrupted environment: Exploring higher education student resilience using the dynamic interactive model of resilience", explores the insights gained into how systems (people, institutions, and societies) cope during disruption (COVID-19). The research was conducted at a university in the South West of England. It employed a mixed-methods approach to examine students' responses and coping strategies amid COVID-19 disruption and factors influencing their resilience.

Calonge et al.'s "Should I stay or should I go? International students' challenges and opportunities to secure employment in their host country after graduation. A scoping review using PRISMA" examines challenges and opportunities for international students seeking post-graduation employment. International students are often motivated to seek job opportunities yet frequently struggle to find employment in host countries, facing lower rates as compared to their local peers.

Alptekin et al.'s research, "An analysis of the learning styles in online environments of graduate students studying distance education," analysed learning styles in the context of a Turkish university's distance education non-thesis Master's programme. Their findings suggest that learning styles do not significantly differ based on sex, income, or device use. Age influences visual, aural, and active learning levels. Retired students showed lower audio-visual and active learning levels. Higher technology use efficacies correlated with increased logical learning levels. Students with extensive daily device use exhibited higher independent learning levels.

Freire et al.'s "A systematic review of graduate training on cultural competence" examines scholarship from the past decade on graduate training for culturally competent mental health care, focusing on marginalised individuals (based on race, gender, and sexual orientation). The review adopts a holistic view of cultural competence, acknowledging clients within their cultural contexts and recognising power dynamics. Recommendations include refining cultural competence concepts, developing innovative training methods, and enhancing evaluation tools.

Ambe et al.'s "Sociodemographic factors and teaching method preferences among university academics: Implications for effective curriculum implementation" explores the teaching method preferences of 400 university academics and their sociodemographic factors' influence on these preferences in Nigeria. Results showed no significant connections between factors like gender, academic faculty, and teaching experience with teaching method preferences.

Lorenz and König's study "Engaging students through messaging applications in foreign language learning", investigates undergraduates' experiences with eStudentMentors using WhatsApp and Telegram for German language learning at a Singaporean university. Lorenz and König's research found that social perceptions and pressures outweighed perceived benefits, challenging Social Exchange Theory.

Kamali's autoethnographic narrative, "Metamorphosis of a teacher educator: A journey towards a more critical self", traces the author's journey from a non-critical to a critical teacher educator. Data from personal sources like diaries and feedback highlight the factors shaping thoughts and practices. The study underscores how voice, agency, and transformation into critical teacher educators are achieved.

Amaral and Gamez's article "Exploring the synergistic effects of combining design thinking and project-based learning in a blended course" details the creation and execution of a Brazilian project that blends design thinking and project-based learning. Data from multiple sources revealed that design thinking helped address community challenges, motivating learning and problem-solving, and combining approaches fostered project management and interdisciplinary learning.

Limna's study examines "The impact of NVivo in qualitative research: Perspectives from graduate students". Qualitative interviews reveal that NVivo, a qualitative data analysis software, enhances research efficiency, collaboration, and outcomes. Or's "Examining Unified Theory of Acceptance and Use of Technology 2 through Meta-analytic Structural Equation Modelling" analyses empirical studies in education using One-stage Meta-Analysis and Structural Equation Modelling (OSMASEM). OSMASEM enables researchers to explore UTAUT2's technology acceptance and use trends without replicating studies.

While JALT focuses on higher and adult education, we occasionally publish other educational research as an exception. Owan et al.'s study "Predicting students' multidimensional learning outcomes in public secondary schools: The roles of school facilities, administrative expenses and curriculum", builds on prior research by examining how school facilities, administrative expenses, and curriculum impact students' cognitive, affective, and psychomotor learning outcomes in Nigeria. Their findings inform educational quality enhancement strategies.

Owan et al.'s study concludes the research section, which is followed by an interview with an educational thought leader, "A critical perspective on generative AI and learning futures. An interview with Stefan Popenici". The interview focuses on Popenici's discussion of his research on AI's impact on higher education. Themes from his book *Artificial Intelligence and learning futures*, including eugenics and systemic racism, are explored. Popenici critiques the power of technology and its role in higher education's identity crisis. Amongst other things, Popenici and his interviewers explore the challenges and opportunities of higher education brought upon by AI.

Three EdTech articles bring us back to this issue's dominant AI theme, which has so far been exemplified by nine research articles and the educational thought leader interview. First, Ifelebuegu et al.'s contribution examines the role of AI in education, particularly chatbots, highlighting their benefits, like personalised learning and administrative ease, alongside challenges such as job displacement and misinformation. It explores AI's impact on research and collaboration. Ethical concerns, including data privacy and the digital divide, are also addressed. The paper emphasises the need to balance AI and human elements in education and calls for robust ethical frameworks for AI use in educational settings.

Second, Calonge et al.'s EdTech article "Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics" returns us again to the topic of generative AI and higher education. The authors compare AI chatbots (ChatGPT, GPT-4, Bard, and LLaMA) for mathematics and statistics education. Their research highlights chatbots' potential positive impact on higher education transformation. Third, Gamage et al.'s contribution, "ChatGPT and higher education assessments: More opportunities than concerns?", addresses the increasing use of AI tools like ChatGPT and their near-human writing capabilities. This has raised concerns about student cheating in assessments. The paper investigates why students are tempted to cheat, the challenges in detecting AI-generated content, and the potential of AI to improve the assessment of higher-order thinking skills among academics.

The EdTech section is followed by Chen's case study, "Mentoring international postgraduate students and early career researchers through transnational telecollaboration: a supervisor's autoethnography". He discusses the challenges international students face pursuing Higher Degrees by Research (HDR) in Australia.

Furthermore, the issue contains four opinion pieces. The first opinion piece is Popenici's paper titled "The critique of AI as a foundation for judicious use in higher education," which addresses the challenges posed by integrating Artificial Intelligence (AI) in education after the launch of ChatGPT. Beyond the AI hype and marketing, it critically examines potential risks, ethical considerations, and practical applications. The analysis encompasses AI's ethical implications, effects on higher education teachers, students, and learning, and long-term societal consequences, seeking ways to utilise AI beneficially.

As AI and chatbots like ChatGPT advance, educators assess their benefits and risks in online assessment. While AI offers personalised learning, its use challenges assessment legitimacy and integrity. Ifelebuegu's opinion piece "Rethinking online assessment strategies: Authenticity versus AI chatbot intervention" examines AI's impact on authentic online assessments, highlighting issues with current testing validity due to AI misuse. He emphasises the importance of authentic assessments that foster higher-order skills, resisting AI influence. However, AI can aid assessment automation, personalisation, and collaboration. Ifelebuegu's contribution advocates rethinking and improving online assessments in the AI era for greater authenticity and resilience against malpractice.

Next, Martin Andrew's "Neo-neoliberalist capitalism, intensification by stealth and campus real estate in the modern university in Aotearoa/New Zealand" critiques higher education and vocational training reforms in Aotearoa/New Zealand. Andrew explores the origins and manifestations of neoliberal ideology in the country's tertiary education. Neoliberal policy's work intensification and *responsibilisation* impacts academics adversely. Examining universities' finances and property portfolios, the article explores higher education issues in the age of 'neo-neoliberalist capitalism'.

In the issue's final opinion piece, Gilmore critically reflects on his personal educational experiences, recalling supportive and unsupportive educators. Gilmore's opinion piece "The (academic) road less travelled: From dropout to recovery" offers personal academic redemption and hope for those who have faced similar challenges. It emphasises recovery and eventual success as a response to those who doubted our potential. Rahimi's brief paper "Developing and analysing an authentic technical proposal writing assignment through the lens of an authenticity framework: Implications for practice" explores the use of an authentic assessment framework in the analysis of a technical proposal writing assignment in an undergraduate engineering course.

Finally, we come to the book review section, which contains two detailed discussions by Rudolph. The first review assesses Popenici's *Artificial Intelligence and learning futures.* Popenici challenges the idea that AI is a universal solution. Although the book predates the generative AI craze, Rudolph argues that this is an important, rich and challenging book as it discusses 'intelligence' and 'artificial intelligence' in a historical and critical higher education context. Rudolph's second review is about *Learning intelligence: Innovative and digital transformative learning strategies*, edited by Kumaran Rajaram and co-authored with Samson Tan. The book focuses on guiding complex learning in the digital transformation and innovation era. It is recommended for its thought-provoking content and broad coverage of higher education teaching and learning topics in the digital age.

References

Adarkwah, M. A., Amponsah, S., van Wyk, M. M., Huang, R., Tlili, A., Shehata, B., Metwally, A. H. S., & Wang, H. (2023). Awareness and acceptance of ChatGPT as a generative conversational AI for transforming education by Ghanaian academics: A two-phase study. *Journal of Applied Learning and Teaching, 6*(2), 78-93. https://doi.org/10.37074/jalt.2023.6.2.13

Ahmad, M., Alhalaiqa, F., & Subih, M. (2023). Constructing and testing the psychometrics of an instrument to measure the attitudes, benefits, and threats associated with the use of Artificial Intelligence tools in higher education. *Journal of Applied Learning and Teaching, 6*(2), 114-120. https://doi.org/10.37074/jalt.2023.6.2.36

Andrew, M. (2023). Neo-neoliberalist capitalism, intensification by stealth and campus real estate in the modern university in Aotearoa/New Zealand. Journal of Applied Learning & Teaching, 6(2), 393-401. https://doi.org/10.37074/jalt.2023.6.2.16

Anthropic. (2023, May 11). *Introducing 100K context windows*. https://www.anthropic.com/index/100k-context-windows

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52-62.

Bastani, A. (2020). *Fully automated luxury communism*. Verso Books.

Berg, J., Raj, M., & Seamans, R. (2023). *Capturing value from Artificial Intelligence.* Academy of Management Discoveries.

Brookfield, S. D., Rudolph, J., & Tan, S. (2023). *Teaching well: Understanding key dynamics of learning-centered classrooms*. Routledge.

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching, 6*(2), 94-104. https://doi.org/10.37074/jalt.2023.6.2.12

Chan, C. K. Y., & Tsi, L. H. (2023). *The AI revolution in education: Will AI replace or assist teachers in higher education?*. arXiv preprint arXiv:2305.01185.

Chomsky, N., Roberts, I., & Watumull, J. (2023, March 8). Noam Chomsky: The false promise of ChatGPT. *The New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Collini, S. (2012). *What are universities for?*. Penguin.

Cook, L. (1938). *Community backgrounds of education: A textbook and educational sociology.* Macmillan.

Devlin, H. (2023, July 7). AI likely to spell end of traditional school classroom, leading expert says. *The Guardian.* https://www.theguardian.com/

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). *Gpts are gpts: An early look at the labor market impact potential of large language models*. arXiv preprint arXiv:2303.10130, 2023.

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching, 6*(1), 57-63. https://doi.org/10.37074/jalt.2023.6.1.22

Fleming, P. (2021). *Dark academia. How universities die*. Pluto Press.

Fleming, P., Rudolph, J., & Tan, S. (2021). 'Never let a good crisis go to waste'. An interview with Professor Peter Fleming on dark academia, the pandemic and neoliberalism. *Journal of Applied Learning and Teaching, 4*(2), 110-120. https://doi.org/10.37074/jalt.2021.4.2.14

Gamage, K. A., Dehideniya, S. C., Xu, Z., & Tang, X. (2023). ChatGPT and higher education assessments: more opportunities than concerns?. *Journal of Applied Learning and Teaching, 6*(2), 358-369. https://doi.org/10.37074/jalt.2023.6.2.24

Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). ChatGPT has aced the test of understanding in college economics: Now what?. *The American Economist,* 05694345231169654.

Ginsberg, B. (2011). *The fall of the faculty.* Oxford University Press.

Graeber, D. (2018). *Bullshit jobs. A theory.* Allen Lane.

Haase, J., & Hanel, P. H. (2023). *Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity*. arXiv preprint arXiv:2303.12003.

Hassoulas, A., Powell, N., Roberts, L., Umla-Runge, K., Gray, L., & Coffey, M. (2023). Investigating marker accuracy in differentiating between university scripts written by students and those produced using ChatGPT. *Journal of Applied Learning & Teaching, 6*(2), 71-77. https://doi.org/10.37074/jalt.2023.6.2.13

Haw, M. (2019, September 6). Will AI replace university lecturers? Not if we make it clear why humans matter. *The Guardian.* https://www.theguardian.com/education/2019/sep/06/will-ai-replace-university-lecturers-not-if-we-make-it-clear-why-humans-matter

Houser, K. (2017, December 12). The solution to our education crisis might be AI. *Futurism*. https://futurism.com/ai-teachers-education-crisis.

Huang, R., Tlili, A., Xu, L., Ying, C., Zheng, L., Metwally, A. H. S., ... & Bonk, C. (2023). Educational futures of intelligent synergies between humans, digital twins, avatars, and robots-the iSTAR framework. *Journal of Applied Learning & Teaching, 6*(2), 28-43. https://doi.org/10.37074/jalt.2023.6.2.33

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. Journal of Applied Learning and Teaching, 6(2), 385-392. https://doi.org/10.37074/jalt.2023.6.2.2

Ifelebuegu, A. O., Kulume, P., & Cherukut, P. (2023). Chatbots and AI in Education (AIEd) tools: The good, the bad, and the ugly. *Journal of Applied Learning and Teaching, 6*(2), 332-345. https://doi.org/10.37074/jalt.2023.6.2.29

Ismail, F., Tan, E., Rudolph, J., Crawford, J., & Tan, S. (2023). Artificial intelligence in higher education. A protocol paper for a systematic literature review. Journal of Applied Learning and Teaching, 6(2), 56-63. https://doi.org/10.37074/jalt.2023.6.2.34

Kefalaki, M., Diamantidaki, F., & Rudolph, J. (2022). Editorial 5(SI1): Technology and education. Innovation and hindrances. *Journal of Applied Learning & Teaching, 5*(S1), 6-11. https://doi.org/10.37074/jalt.2022.5.s1.1

Luckin, R. (2023, July 14). Yes, AI could profoundly disrupt education. But maybe that's not a bad thing. *The Guardian*. https://www.theguardian.com/

Mikami, K. (2023, August 17). Is there AI that can make charts and graphs? *EcoAGI*, https://ecoagi.ai/topics/AIGC/is-there-ai-that-can-make-charts-and-graphs

Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1), 16-30. https://doi.org/10.37074/jalt.2023.6.1.34

Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *Journal of Applied Learning and Teaching, 6*(2), 105-113. https://doi.org/10.37074/jalt.2023.6.2.10

Mollick, E. R. (2023a, August 13). Automating creativity: There is now strong evidence that AI can help make us more innovative. *One useful thing*. https://www.oneusefulthing.org/p/automating-creativity

Mollick, E. (2023b, June 12). Assigning AI: Seven ways of using AI in class. *One useful thing*. https://www.oneusefulthing.org/p/assigning-ai-seven-ways-of-using

Mollick, E. R., & Mollick, L. (2023, March 17). *Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts*. http://dx.doi.org/10.2139/ssrn.4391243

Noy, S., & Zhang, W. (2023). *Experimental evidence on the productivity effects of generative artificial intelligence*. SSRN: https://ssrn.com/ abstract=4375283

Orwell, G. (1933). *Down and out in Paris and London*. Penguin

Popenici, S. (2023a). *Artificial Intelligence and learning futures: Critical narratives of technology and imagination in higher education*. Routledge.

Popenici, S. (2023b). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching, 6*(2), 378-384. https://doi.org/10.37074/jalt.2023.6.2.4

Popenici, S., Rudolph, J., Tan, S., & Tan, S. (2023). A critical perspective on generative AI and learning futures. An interview with Stefan Popenici. *Journal of Applied Learning and Teaching, 6*(2), 311-331. https://doi.org/10.37074/jalt.2023.6.2.5

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Rebelo, M. (2023, August 2). The best AI video generators in 2023: Use these tools to create, edit, and enhance videos with AI. *Zapier*. https://zapier.com/blog/best-ai-video-generator/

Rudolph, J. (2014). *Massive Open Online Courses (MOOCs) as a disruptive innovation in higher education?* M.Ed. dissertation, University of Adelaide.

Rudolph, J. (2018). A brief review of Mentimeter – A student response system. *Journal of Applied Learning & Teaching, 1*(1), 35-37. https://doi.org/10.37074/jalt.2018.1.1.5

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning & Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Rudolph, J., Tan, S., & Aspland, T. (2023c). JALT Editorial 6(1): Fully automated luxury communism or Turing trap? Graduate employability in the generative AI age. *Journal of Applied Learning and Teaching, 6*(1), 7-15. https://doi.org/10.37074/jalt.2023.6.1.35

Russell, B. (2004). *In praise of idleness and other essays*. Psychology Press.

Santandreu, D. C., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching, 6*(2), 346-357. https://doi.org/10.37074/jalt.2023.6.2.22

Shimek, C. (2023, July 6). AI outperforms humans in creativity test. *Neuroscience News*. https://neurosciencenews.com/ai-creativity-23585/

Solomon, S., & Du Plessis. M. (2023). Experiences of precarious work within higher education institutions: A qualitative evidence synthesis. *Frontiers in Education, 8,* https://doi.org/10.3389/feduc.2023.960649

Stevens, D. (2019). *Write more, publish more, stress less!* Stylus.

Stracke, C. M., Chounta, I.-A., Holmes, W., Tlili, A., & Bozkurt, A. (2023). A standardised PRISMA-based protocol for systematic reviews of the scientific literature on Artificial Intelligence and education (AI&ED). *Journal of Applied Learning and Teaching, 6*(2), 64-70. https://doi.org/10.37074/jalt.2023.6.2.38

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching, 6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Susskind, D. A. (2021). *World without work: Technology, automation and how we should respond.* Penguin.

Sword, H. (2012). *Stylish academic writing*. Harvard University Press.

Sword, H. (2017). *Air & light & time & space. How successful academics write.* Harvard University Press.

Sword, H. (2023). *Writing with pleasure.* Princeton University Press.

Terzian, S. (2019). The history of technology and education. In J. L. Rury & E. H. Tamura (Eds.), *The Oxford handbook of the history of education* (pp. 554-567).

Thio, S. Y. (2023, February 24). ChatGPT: Has artificial intelligence come for our jobs?. *The Business Times*, https://www.businesstimes.com.sg/opinion-features/chatgpt-has-artificial-intelligence-come-our-jobs

Tlili, A., Huang, R., Mustafa, M. Y., Zhao, J., Bozkurt, A., Xu, L., ... & Burgos, D. (2023). Speaking of transparency: Are all Artificial Intelligence (AI) literature reviews in education transparent?. *Journal of Applied Learning and Teaching, 6*(2), 44-55. https://doi.org/10.37074/jalt.2023.6.2.15

Watters, A. (2021). *Teaching machines: The history of personalized learning.* MIT Press.

World Economic Forum. (2023). *Future of jobs report 2023.* https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf

Xames, M. D., & Shefa, J. (2023). ChatGPT for research and publication: Opportunities and challenges. *Journal of Applied Learning and Teaching, 6*(1), 390-395. https://doi.org/10.37074/jalt.2023.6.1.20

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Come to the cabaret: Voices from the modern university

Martin Benedict Andrew[A]

A Associate Professor of Work-Based Learning, Otago Polytechnic, Dunedin, Aotearoa/New Zealand

DOI: https://doi.org/10.37074/jalt.2023.6.2.19

Abstract

This article creates and curates a sequence of cabaret songs to represent recognizable characters or types from the scholarship and literature of the modern, neoliberalised university. Using poetic enquiry as an ethnographic technique, I stand at the border of practice-based qualitative and performative research paradigms and re-imagine such figures as the ninja, the nervous wreck, the precariat, the zombie and the activist as characters in the cabaret of the modern university. The enquiry has two primary groundings: the comprehensive literature of the increasingly toxic modern university and a sociocultural critique of the *Kabarett*, the underground cabarets of the Weimar Republic of the 1920s and early '30s. The art of the cabaret involved writing and performing satirical portraits of familiar types seen in society and the world.

The expressionistic era of the *Kabarett* was a time of decadent creativity and unashamed freedom of voice, an era that reached eventual termination after the rise of Nazism. Foregrounded by an exploration of the features of the modern university, including its corporatism, its performative audit culture, its governmentalised hegemonic regimes of truth and its climate of anxiety and fear, the study traces the origins of its types and aims to replicate aspects of their identities. The research examines both the origin stories of ninjas, nervous wrecks, zombies, precariats and activists and interprets their identities in the satiric form of subversive *Kabarett* songs. The findings of this study can literally be performed. The study also reaches back to medieval and Renaissance drama to examine and critique my dramatic narrative voice. The ironic eye with which I view this cabaret aligns with that of Puck/Robin Goodfellow in Shakespeare's *A midsummer night's dream*. This ironic eye examines the denizens of the neoliberalised universities and declares, "Lord, what fools these mortals be!"

Keywords: Academic identity; cabaret; free speech; neoliberalism; performative research; poetic enquiry; satire.

Introduction

This study originates from my deep immersion in the literature of the 'modern' university with its roots in German exchange (Levine, 2021), the 'university' itself, of course, dating back to 11th century Bologna and grounded in law and, later, theology. The 'modern university' is known by other names: the 'modernized' (i.e. messed around with) university (Soares, 2002), the neoliberalised university with its audit culture (Sparkes, 2007), zombie economics (Quiggan, 2010) and capitalistic cult of performance (Micali, 2010); the university in ruins (Readings, 1996); the multiverse 'schizophrenic university' (Shore, 2010); the 'toxic university' (Smythe, 2017); 'dark academia' (Fleming, 2021), 'dystopia' (McBride, 2022), 'the troubled university' (Hil, 2012) and most comical of all, 'whackademia' (Hil, 2012). The latter is close to the university depicted in David Lodge's satirical trilogy of astute campus novels such as *Nice Work* (1988, subtitled *An academic romance*), presciently played out against the Thatcherite corporate cost-cutting that we still experience today (Blackham, 2020; Solomon & Du Plessis, 2023). Before I invite you to come to the cabaret to meet the characters of the modern university, I will, in lieu of the literature review of ordinary research, take you on an artistic audit (Hasemann, 2006) of the terrain.

Artistic audit

The modern university

Reviving the term 'modern university', Levine has a current project entitled *Rehabilitating the university as a public good* (Stanford, 2021), arguing that leaders of the modern university need a chameleonic entrepreneurial self that encompasses the scholarly, the social and both the political and the economic. Such a protean figure may not yet exist in the *dramatis personae* of the modern university, but it is a pleasing thought to imagine a return to the fruitful utopia of academia, where the key figures were the professor and the student; where there may be a theoretical space for the 'ethical academic' (Barrow & Xu, 2021) and where the 'pipeline' from academic to professor (Spina et al., 2022) might exist if the path were not littered with human and technocratic obstacles.

However, regardless of how we negotiate conflicting discourses, frame our agency and leverage entrepreneurial technologies of the self (Barker, 2017; Barrow & Xu, 2021; Varea et al., 2021) or care of the self (Foucault, 1986; see also Ball & Olmedo, 2012), make no mistake: this is a *war* on higher education (Giroux, 2019). Put bluntly, universities are increasingly managed and neoliberalised, corporatising and commercialising universities and bringing with it corporate cost-cutting (Blackham, 2020) and building real estate empires instead of funding academic positions (Andrew, 2023). Baumann (1999) taught us that climates of fear and uncertainty are not inevitable results of time cycles but creations of human power. Foucault (1980) taught us about a society's regimes of truth: the mechanisms determining and sanctioning truth and falsehood depend on the degree of privilege and status of those "charged with saying what counts as true" (Foucault, 1980, p. 131). Andrew's (2023) case study of a neoliberalised university in New Zealand suggests multiple stakeholders close to the university have vested interests in the full truth remaining hidden. As Spina et al. (2022) wrote in their analysis of the subaltern precariat: "discourses privilege certain perspectives, while creating discursive prohibitions around other points of view in specific times and places" (p. 535). It is with these discourses and perspectives that I play, inviting you to the cabaret.

Types of the modern university

Scholars within the subgenre of university critique have developed their own *dramatis personae*, a list of types who dwell in the fictional domain of *Academe*, a term which is itself evokes that quaint and nostalgic sense of an Arcadian past. Despite ultimately being a story of agency and self-governance, Parker's 2017 article "Ninjas, zombies and nervous wrecks?" in particular offers three clearly recognizable types of the modernized university. Indicating the two main interpretations of the zombie, Ryan (2012) had already asked whether academic zombies are a failure of resistance or a means of survival. Acquiescent, they push on because they need the work, conforming with compliance and audit. The human anxieties embodied in the zombie are globalisation anxieties and fear of mass control. This is because, as Molpeceres (2017) wrote, "the zombie is a brainless being unable to understand, explain or judge the surrounding world" (p. 161). As Deslandes and Adamson (2013) note, the zombie in the literature of the modern university represents the deadly hand of capitalism and the "life sucking practices of institutional bureaucracy" (p. 69) but remains cognizant. Zombiedom infuses all aspects of the modern university (Whelan et al., 2013). Like all the types in the cabaret, zombies evoke "myths [that] appeal to our unconscious cognitive frames" (Molpeceres, 2014, p. 86). Similarly, the jaded nervous wreck is a composite of the figures described in the literature of increasingly serious sickness (Gill, 2009; Fanghanel, 2012; Acton & Glasgow, 2015), even death and suicide (Fleming, 2021; Treagar et al., 2022). Literature also describes those in the third space, more professional than academic but synergetic just the same (Whitchurch, 2013) and also often precariat. Academic developers, digital education specialists and student support are increasingly shoehorned in this subaltern space.

Speaking of the subaltern, literature characterizes the impermanent, insecure, powerless, precariat (Standing, 2011; Blackham, 2020), doing the 'housework' and always on the edge of obtaining or losing tenure or work itself (Solomon & Du Plessis, 2023). They are typically "undervalued, overused, and stigmatized" (Solomon & Du Plessis, 2023). Barcan (2013) and Spina et al. (2022) suggest that the disillusioned mid-career precariat may be more impacted by a neoliberalist ethos than a more entrepreneurial 2020s post-doctorate precariat. Precarity was once a stepping stone to academia, but latterly it has become a prison (Spina et al., 2022) with sessionals "a legion of lost souls" (Whelan et al., 2013, p. 69). The young precariats, the new graduates with one foot on the ladder, are still driven, enticed and aspirational, even hungry for tenure. They are given titles like 'Teaching Fellow' and they still negotiate, even embrace the challenges and contradictions (Barrow & Xu, 2021; Varea, et al., 2021). Some of these early career researchers are characterized as superheroes like the Powderpuff Girls with their 'liquid' subjectivities (Varea et al., 2021) – at least in nations such as Argentina, where such agency might still exist. They are agile bodies, "robust, resilient, responsive, flexible, innovative, and adaptable" (Gillies, 2011, p. 210). Such young precariats need to be extraordinary to secure ongoing work. Such academics are 'superheroes' (Pitt & Mewburn, 2016) or 'rockstars' (Smyth, 2017) but still types of the toxic university.

Types

Carl Jung (1970) saw archetypes as personages who embody the universal traits of readily identifiable characters. He initially identified a group of four, which I present here with how I realise them in parentheses: the persona *(our mask, how we present ourselves to the world)*, the shadow *(our repressions such as secret desires and weaknesses and our instincts and orientations)*, the anima/animus *(our projected – and gendered – self and our gateway to collective consciousness)*, and the self *(the individuated site of the unification of our unconscious and conscious where the ego meets the personality)*. Jung famously declared that all the most powerful ideas in history go back to archetypes (Jung, 1970). The modern university has its fair share of these: the fastidious micromanager, marked by an inward deep feel of failure; the tweedy professor, spurting forth either archaic or arcane discourse in words of no fewer than four syllables; the owl-like office administrator, characterized by a hawkish eye for detail and a marked efficiency shown in their clipped voice. I just made these examples up out of thin air, except it is not thin air, but from our shared experience of people seen in the modernized university.

Situated at the intersection of the mystic and the pseudoscientific, I need to go no closer to illustrating the legitimacy of the concepts of collective unconscious and the archetype. These types are memories from my own personal unconscious, grounded in my own experience. They are products of my experiential learning, presented as personages in an evocative autoethnography. In Jungian terms, my own consciousness recognizes and represents figures from the external world into visible reality. Those specific but overlapping and unfixed figures identified by Jung may be listed as Ruler, Creator/artist, Sage, Innocent, Explorer, Rebel, Hero, Wizard, Jester, Everyman, Lover and Caregiver. Molpeceres (2017) links Jungian types and myth: "If… myths and archetypes are activated when needed in a particular society, then the study of myths will allow us to discover and understand the fears, worries, desires, and aspirations of that society" (p. 152). For Jung, an individual's goal was to achieve a cohesive self, similar to Maslow's 1943 needs-based concept of self-actualization. Late-era Maslow (1987) emphasized that human behaviour has multiple motivation points: "any behavior tends to be determined by several of the basic needs simultaneously rather than by only one of them" (p. 71). The characters in my cabaret actualize themselves, and perhaps their multiple needs, through their songs, but I intend them to be recognizable as types of the modern university, made up, perhaps, of parts of any of the above, and all with penetrable personae, animus, shadows and selves.

Drama and irony

The 15th century medieval morality play, epitomised by *Everyman*, predates Jung by centuries but contains a trope that informs how the characters of my cabaret represent themselves. In *Mankind*, for instance, there appear Mercy, Mischief, Nought, New Guise, Nowadays, Mankind, and Titivillus (the devil). By way of both characterization and exposition, characters typical of vices and virtues soliloquise their identities to an audience already literate in the tropes and typologies of the drama. Collective literacy enables audiences not only to recognize the physical and rhetorical characteristics of types but also to understand the drama as an allegory; that is, they are invited to see themselves in Everyman as he comes to understand the death and fate of his human soul (cue Ball, 2003). Everyman is torn between the Seven Deadly Sins pulling him one way and the Four Daughters of God (Mercy, Justice, Temperance, and Truth) pulling him the other. It says much about the world today that few of us could name God's imagined daughters, but most could muster the seven deadly sins, largely due to our uncanny familiarity with the 1995 David Fincher film *Se7en*. The moral trait of moderation, the key message of many morality plays, is followed by few in our age of the neoliberalised university, powered by corporate greed and populated by academics vainly seeking to publish, not perish, and vice chancellors coveting another university's position on the league table or more acquisition of real estate (Baldwin, 2021; Andrew, 2023).

Of the many vernacular morality plays known to have existed in English, few survive, and today what we chiefly know are the Seven Deadly Sins and Faust. The Faust story, allegorized in the literature of the modernized university as the scholar's sacrifice to business greed (Ball, 2003; Shore, 2010), remains the best-known morality, stimulating generations of audiences, including Marlowe's *Doctor Faustus* (1604). *Mephistopheles* in Goethe's *Faust* (trans. George Madison) disguises himself as a starveling and embarks on a moral tale about greed. Before it turns into a misogynistic tale, the Starveling speaks like the stock character of Greed (Avaritia):

> When hearth and home were women's zone,
> As *Avaritia* I was known.
> Then did our household thrive throughout,
> For much came in and naught went out!
> Zealous was I for chest and bin;
> 'Twas even said my zeal was sin.
> *(Faust, II.iv)*

In drawing on the morality play, I return to a dark place in academia visited previously (Andrew, 2019), where I use the tropes of the morality play *The castle of perseverance* as a method of presenting my autoethnography and the metaphor of the post-mortem as a forensic method of analysing findings like entrails. I used a *Hamlet-like* persona to evoke the ills of the ruined university. A different Shakespearean persona is at play in the present article. In 1595/6's *A midsummer night's dream* (III.ii), Puck, the play's narrator, a sprite-cum-henchman for Oberon, King of the Fairies, observes the romantic confusions of two bickering pairs of human lovers and declares, "Lord, what fools those mortals be!" (l.117). Like a morality play figure, Puck is also known as Robin Goodfellow, so his basic goodness, despite his playfulness, is signalled to the audience early on. Yet his status as a supernatural is shown in his eye-of-God envisioning of the stupid humans from an otherworldly domain, resulting in a good-natured satire. Like King Lear's fool, we know he is the speaker of truths, both jester and sage. Channelling mischievous Puck, but not leaving Hamlet entirely behind, I invite you to come to the cabaret.

The all-knowing Master of Ceremonies of *Cabaret* (such as Fosse's 1972 film) curates the presentation of the show, just as I, as a researcher, present a sequence of the types of the modernized university, playing chess with the protagonists. The metaphor of the chess game as an allegory for action playing out for protagonists is a recognizable trope of allegory in popular culture. The credits for the television series *The Aphrodite inheritance* (Bird, 1979) showed somehow omniscient hands moving chess pieces, hinting that the three protagonists are types of Greek gods. The trope had been cleverly used by Thomas Middleton in both *A game at chess* (1624) and the Jacobean tragedy, *Women beware women* (1621/1657; II.ii) as an allegory of the battle of virtue and vice in the character of Bianca, who turns out not to be so white. At the time of encountering revenge tragedy, I was well into a combined arts and social sciences degree. An inverted *Midsummer night's dream*, where the masque had been a celebration of the normalizing power of marriage, *Women beware women* contains a marital anti-masque of mass murder where the lustful and borderline incestuous Hippolito is shot by cupids bearing poisoned arrows before falling on his own sword. The character's name recalls Hippolyta, Queen of the Amazons in *A midsummer night's dream*, and the sinister cupids seem like blackly comical parodies of Puck. All is not as it may seem in these human chess games, so come to the cabaret.

Das Kabarett

Das Kabarett, cabaret, was an expressionistic place for subversive identities in the Weimar Republic of 1920s Germany (Wouilloz-Boutrois, 2021). The late '20s Weimar was a rare moment in time and place where glimpses of real people could be had, in this case, in lyrical underground satires. This was an era of unprecedented freedom of expression and speech, but it was very place-specific. The epicentre of this Bohemian cliqué was the monocled lyricist Marcellus Schiffer who "targeted the snobbish, upper-class bourgeois, whose weaknesses he revealed bluntly and humorously with wicked charm" (Wieland, 2011, p. 71) but not ideology. The fact that these kinds of freedoms are under threat as the 2020s cave in serves as a reminder to us that 100 years on, we have learned precisely nothing (Cantu & Lambert, 2023).

Simultaneous with the Kabarett, but in a slightly parallel universe, Bertolt Brecht and Kurt Weill were collaborating on theatrically innovative socio-political satires like *Die Dreigroschenoper* (1928, *The Threepenny Opera*). Socio-political satire, represented with vivid expressionism and Lotte Lenya, was a powerful mix. By 1933, Friedrich Hollaender, one of several masters of the Kabarett, was one of the many Jewish artists of the period who fled Germany to write songs for Hollywood films. Austrian Jew Max Reinhardt, director of the satiric cabaret revue *Schall und Rauch* (*Sound and smoke*, 1929) and of annual productions of *Everyman*, fled too, making his renowned *A midsummer night's dream* (1927, 1934) into a film (1935). Another fled Jew, Erich Wolfgang Korngold, composed the score. Mickey Rooney was Puck, wondering at what fools these mortals be.

As the roaring 20s depressed their way into the 30s, we now can hear the imminent danger of Nazism in increasingly allegorical songs like Frederick Hollaender's terrifying composed lyric "Münchausen" (1931), which captures the transition from days of freedom to the coming of the swastika. In strophes, it relates anecdotes about the impact of an authoritarian culture steeped in lies on ordinary people. Narratively, this elegiac song epitomises one key feature of the cabaret songs of this period: the use of first-person personae ("I" and "we") to record experiences and impressions and to capture identities.

Singer (2000) reflects that for a brief window in '20s Berlin, you could enter Kurt Robischek's Cabaret of Comedians (*Kabarett der Komiker, The Comedian Cabaret*, popularly called *KadeKo*). It was the embodiment of big-city attitudes towards life (Wieland, 2011), and Schöneberg's queer district still dominates today. Lareau (2011) writes that it embodied the cosmopolitan spirit of the Kurfürstendamm, with a whiff of Viennese elegance. Here, the music of Mischa Spoliansky rang out in clever collaborations with the scriptwriter and lyricist Schiffer (Wouilloz-Boutrois, 2021). Song after song satirised, often in *Sprechstimme (spoken voice)*, identity-unfurling first person soliloquies, often mocking super cool Berlinites and members of the smart set of the Spoliansky song "(Das) Gesellschaftslied" ("The social song"/ "The smart set"). A satirical lyric soliloquy, Spoliansky's "Ich bin ein Vamp!" ("I'm a vamp!") satirises the movie-struck fantasies of an average young woman who aspires to be like Theda Bara, while "Sex appeal" (no translation needed) does the same for a Garbo aspirant.

Spoliansky's remarkable "Das lila Lied"/ "The lavender song" (lyrics: Kurt Schwabach), steeped in joyful defiance, is a '20s version of a gay pride march/anthem: "Und dennoch sind die Meisten stolz, daß sie von ander'm Holz!" *("And most of us are nevertheless proud to be cut from different cloth")*. The story of such jazz clubs, their artists, such as the Weintraub Syncopators (who appeared in the 1930 Dietrich film *The blue angel*) and their denizens (including the Nazi Ernst Röhm, whose queerness was an open secret initially tolerated by Hitler) is told in the 2023 documentary *Eldorado* (Cantu & Lambert, 2023). Hitler encouraged Röhm to commit suicide. Refusing, he was executed in a Munich cell. In 1933, Hanns Eisler and Bertolt Brecht briefly curated *Cabaret in the exile*, in non-Nazi occupied territories. It is amazing that much of the *Entartete Musik (forbidden music)*, utterly outlawed by 1934, survived, most of the operatic revues having been consumed in the fires of Nazism (Lareau, 2011).

Singer (2000) also writes that lyric writers like Schiffer and Kurt Tucholsky, Walter Mehring and Erich Kästner were word-perfect social critics. Their persona-based songs were interpreted by sometimes cross-dressing artists like Max Hansen, Trude Hesterberg, Curt Bois, Kurt Gerron, Eva Busch and, momentarily, Dietrich. Nowhere is the gender-bending clearer than in the Spoliansky song "Maskulinum-Femininum", and those seeking evidence of pro-queer discourse look to "Wenn die beste Freundin" ("When the special girlfriend") as well as "Das lila Lied" ("The Lavender Song"). "Wenn die beste Freundin" was Dietrich's duet with pencil-thin Margot Lion, Schiffer's lover, and seems to use the female propensity for shopping to mask a lesbian relationship from the 'sweet little man', played by Oscar Karlweiss in the song's 1928 debut in the revue *Es liegt in der Luft/ It's in the air* set in a department store. Wieland (2011) writes that Spoliansky claimed he was satirising the Dolly Sisters and their materialism and engagement with Mr Selfridge, hence the department store. If that is the case, he would be referencing a 1925 incident, old hat by 1928 when the gambling Dollies were in decline. Wieland (2011) claims that sexual references were added by later biographers, an issue of retrospective denialist erasure.

Repopularising the forbidden music of the Weimar Republic in the collection "Berlin Cabaret Songs" (1996, from translations by Alan Lareau and Kathleen Komar and arranged by Robert Ziegler English version 1997, translated by Jeremy Lawrence, in Decca's *Entartete Musik*, degenerate music, series), Ute Lemper became the modern empress of Kabarett. Her work not only brought back these underground songs; it also interpreted and winkingly recontextualised them for modern audiences. She said in an interview: "this is cabaret, political satire, where you deal directly with the audience. You give a message about society" (Clarke, 1997, p. 32). Nearly 30 years on, Lemper continues to re-present this restored repertoire for audiences, drawing ever more pertinent parallels between Nazism and Trumpist populism. The German 20s/30s oeuvre, which Lemper makes her own, also has a strong morality play element, as I indicated in Korngold's love of *Everyman*. Kurt Weill's hybrid ballet-vocal score *Die sieben Todsünden/ The seven deadly sins* (1933), which she recorded in 1991, channels Kabarett tropes to create a lyrically bizarre but harmonically expansive modern morality play. There is something in Lemper's interpretations that taps into a Jungian universalism, perhaps because, as she points out, everything is about telling a story (Clarke, 1997).

There is one more important fact about Lemper. Along with Liza Minnelli in the 1972 Bob Fosse film, she embodies Kit Kat Kabarett chanteuse Sally Bowles, the protagonist of *Cabaret* (music John Kander; lyrics Fred Ebb' book by Joe Masteroff, based in turn on Christopher Isherwood's semi-autobiographies including *Goodbye to Berlin*, 1938, featuring Sally). Lemper was a celebrated Sally in Paris in 1997. *Cabaret* is set in 1929/30 Berlin during the twilight of the Jazz Age as the Nazis are ascending to power. Of this period, Lemper said in an article entitled 'The corrupt world of then, so similar to now', today's Germany is "almost without any memory" (Holden, 2017). Works like *Eldorado*, along with Lemper's work, reveal the fools. Scholars write about the modernized university to ensure this does not happen to our own personal apocalypse.

Methodological notes

My methodological approach owes a debt to the poetic contributors to volume 41 of the journal *Social Alternatives* (2022), subtitled *It's time: the re-form of Australian public university*. Here, the poetic voice is an evocatively autoethnographic one where the activist purpose of the research can be captured best or only by poetic form. Standing at the borderline of qualitative and performative research paradigms, I re-imagine the types who populate the modern university, all of whom originate in existing scholarship, in many cases utilising the device of rescued speech where fragments remembered verbatim appear in my lyrics (Butler-Kisber, 2020). Both found/remembered and practice-led poetry may be either narrative (story-telling), or lyrical, emphasizing subjective feeling and emotion (Butler-Kisber, 2020).

Poetic enquiry is a powerful ethnographic technique (Galvin & Prendergast, 2012) and a practice for engaging in and with the world (Rapport & Hartill, 2012). Viewing such poetics as inductive and iterative creative analytic practice (Richardson, 2000) and "ways of being" (Wiebe, 2015, p. 155) rather than tools or methods, I also wonder "what can be learned from the poet's fierce/mischievous openness to the aesthetic qualities of human experience?" (Wiebe, 2015, p. 153). I wonder this in the context of presenting the types of the modern university as lyrical characters in a cabaret. In the spirit of Rapport and Hartill (2012), my mischievous/ fierce/ tender cabaret attempts to present the human experience I know more deeply. I write, then, "in a blend of fierce/ tender, attending to the humanness of the participants and holding firm the intention to consider how their words are not just findings but "disclosure[s] of the individual" (Rapport & Hartill, 2012, p. 18). Having called my subject matter "our own personal apocalypse" above, I am conscious of my Jacobean horrid laughter both dealing with and reimagining things traumatic. Rapport and Hartill (2012) considered the use of personal testimony of the Holocaust narrative in poetic enquiry and the potential to retraumatise. In light of this, with Wiebe (2015, p. 157), I believe that "balanced by tenderness, a poetic inquiry that is fierce [should] not fall into being ferocious or intimidating".

Hasemann (2006, p. 3) sees the need for a performative paradigm because practice-led researchers "do not commence a research project with a sense of 'a problem'. Indeed they may be led by what is best described as 'an enthusiasm of practice': something which is exciting, something which may be unruly". The research I present here does not originate from a problem which leads to a question; it is rather a way of using poetic enquiry in much the same way as narrative enquiry is often used: to tell stories elucidating a phenomenon or a range of experiences, impressions and perceptions (Butler-Kisber, 2020). This paper is ethnographic, grounded in both experience and data, but re-presents its data in an alternative form or text; in this case, cabaret songs. I exemplify the performance turn in qualitative research, multi-method led by embodied autoethnographic practice (Sparkes, 2007), using material forms of practice in place of 'findings' and an 'artistic audit' in place of a literature review (Hasemann, 2006).

My multiple roles as cultural historian, university academic and lyricist can, hence, blend in a research-informed performative text. I intend readers to pry below the surface to see, for instance, the hidden ego behind the projected personality. Though any resemblance to people living or dead may be purely coincidental, I believe that this approach places the reader in a triangulatory epistemological place. If you see anything familiar in these portraits, it is your recognition that validates my portraits into a liminal place between fiction and knowledge. Who do you recognize? I invite you to come to the cabaret.

Willkommen

The researcher turns Master of Ceremonies to introduce our cast of characters, types from the decadent modernized university. He functions as the Perlocutor in morality plays, laying out the play to lords and ladies. You may see me, at this juncture as Joel Gray as the Emcee in *Cabaret*, presenting "Willkommen."

Willkommen! And bienvenue! Welcome! Fremder, étranger, stranger. You are welcomed to this darkened once hallowed halls of the modernized university. Here in this place of fallen ivory we meet the powerful and the disempowered, the enterprising and the neurotic, the defiantly conformist and the equally defiant non-conformist. In the midst, we find the denizen of the third space, neither fully professional nor fully academic, and the sessional precariat, willing even to betray colleagues with tenure in order to get their teeth on the ladder. Here, ladies and gentlemen, we give you the types of the modernized university. But wait, here comes a man in an expensive suit, and who is the person in Birkenstocks trailing behind him? Ladies and gentlemen, I give you, your manager!

The manager (A tango duet with a staff member)
I'm the greatest manager
My skill is mystic
You micromanage meanly
Narcissistic

I've risen to the top now
Dressed in Prada
But what have you achieved here?
Precisely nada.

They chose me from the hundreds
I'm an idol
You silence the department
With a bridle

I audit your performance
And survey you
And all my loyal colleagues
Want to slay you

I even own your soul now,
My possession
Keep your evil for your priest now
In confession.

But even though my powers take me far beyond the top
We're sure your boundless ego will not tell you when to stop
I'll get the best of all of you and measure and observe you
Assaulting all the women staff, you disgusting perve, you
The women all adore me, fawning, seeking my approval
They fear and hate the sight of you, demanding your removal
I'll get the very best of you, you'll toil till you're tired
And we will not give up our fight. Until. You. Are. Fired.

Stand clear now, ladies and gentlemen, as the ruthless, all-taking wannabe alpha researcher, the ninja approaches, an agile body, arms akimbo, energies effervescent. Stand clear please.

The ninja *(Staccato, dagger-stiletto rhythms)*
I'll cut out your eyes if you stand in my way
Out-publishing you, I'm out-citing, you're brooding.
Soaring my scoring, outplaying your plays
Outsourcing my data and yours I'm excluding.

I'm burning the candle out at either end
Hiding my rivals, my colleagues as friends,
Sweating on studies, on chapters, and then
Excising their authorship when they pretend
My work was their work and if they contend
My credits are theirs, then I'll always depend
On my dean or Vice Chancellor bound to extend *(vile laughter)*
Disciplinary warnings. They count as my friends
Whom I cite without end.

People trust to my face never seeing my stealth
And nobody knows me, just my achievements,
My prize-winning papers, much cited; my wealth's
My own. You can have your bereavements,
Failed promotions, grants always declined,
And abstracts rejected. Reviewers are blind
But not as blind as you, my friends.
A word from me and your career *(horrid laughter)* ends.

Our university is people with both big winners and, sadly with those consumed by the system, and, on that note, here comes the nervous wreck.

The nervous wreck (Attuned to Leonard Bernstein's *America* from *West Side Story*)
I want to sleep but I just can't rest
Doctors say diazapam is best
I think that I wake but I'm still so stressed
I can't catch a breath and my work's regressed

Staff glare at me like my name is mud
My managers think I'm a dreadful dud
My career has crashed with a thunderous thud
And all they want is my sweat and blood

All of the technocrats think I shirk
Yet all that I do counts as admin work
Performance assessors lie in wait and lurk
I can't research while I'm just a clerk

My analyst tells me I can't survive
Yet every day all I do is give
I've forsaken my soul and my mind's a sieve
All of those pricks have no right to live

It's so unfair that those bastards gain
I attend every day but it's all in vain
My resources, my soul and my spirits all drain
All I want is my sleep and release for my brain.

Brain rest is not on the mind of our next character, our wide-eyed zombie, aware of immortality but moving endlessly onward out of fear or economic need. This zombie sings a grunty metallic horror song.

The Zombie
I have survived the apocalypse,
You all lost your jobs and I endured.
Every 'YES' defiles my lips;
You hope someday I might be cured
From comfortable conformity
And realise things that I've denied—
The grossness and enormity.
And times I've lied, or have I died?

I walk the darkened corridors
Of the exploded college walls
Finding my survival horrid or
Vile, while what's left of my spirit calls
Me to recall when we all thrived
And lived in close collegiality
Before mad management thieved
Our souls, replaced them with banality:

Forms, spreadsheets, applications
Documents, pointless white papers:
Critical thought's above our stations;

Research burned to smoke and vapour.
No politics can eat my brain
Resistance frightens every cell
Of my being; I've much to gain
From management, though it be hell.

Our corridors are also populated with third space academics, professional staff who are stuck in a rut and who often harbour less than secret desires to join the faculty. More fool them. Here's one now.

The third space precariat *(Fast, spiky, shaky rhythms)*
And I sit in my office partitioned by glass
Dreaming my days as a lackey will pass
I mentor the students and sit on my ass
And wait to become a full member of staff

I'm not just a starlet, I shine like Lee Remick
I'm called a professional, not academic
I supported IT throughout the pandemic
I've been here so long that they think I'm endemic

I don't sit in the limelight and I deserve better
Hidden, forgotten, a pathetic regretter
(Whimpering) Lost in the post like my employment letter
Overlooked, I'm never going to be a go-getter.

I've completed my doctorate, worked here for years
Assisted and proofread, massaged egos and fears
Made myself indispensable, at least somewhere near
And yet they still treat me as if I'm a spare

Another minion seeking belonging but not finding it is our early-career precariat. This precariat is an urgently wannabe mini-professor, keeping their dagger behind them, casting eyeballs on the jobs of others. We meet our precariat having another job interview with a panel.

The early career precariat (Antiphon)
I am new to this place but I'm keen and I'm clever
(*They are new to this job but they're keen and they're clever*)
I do not have a contract but I'll never say never
(*They do not have a contract but they never say never*)
(*Ritornello*)

I'll conform and comply and obey
(*They'll conform and comply, they'll obey*)
I will do what my managers say
(*They will do what their managers say*)
(*Ritornello*)

If you authorize me, I'll make life hell for the tenured
(*If we authorize them, they'll make hell for the tenured*)
Experience comes when you put in trust in go-getters
(*They're experience comes since they're trusty go-getters*)
(*Ritornello*)

I don't care what they say to my face
(Losing rhythm)(They've no care what staff say to their face)
For its them that I aim to replace
*(Finally disgusted) (This little c*** doesn't have any grace)*
(Ritornello)

Loudly, now, approaches the activist, surrounded by other activists, marching with placards and protesting cuts. Their march is a brassy unionist's anthem.

The activist (tribute to Jeremy Lawrence's translation of Spoliansky's "Das lila Lied"; Sprechstimme)

What makes them think they have the right
To set us tasks from dawn to night?
What makes them think they have the right
To pustulate our academic paradise?
They make our lives hell underground
Forcing us to comply with still more cuts
If we resist, sacking awaits
We know that what they say is total nuts
The crime's to conform to compliance,
Together we're an underground alliance.

Refrain
We're not afraid to resist and suffer
If that means hell then hell we'll take the chance
They're all so straight, uptight, upright and rigid,
Standing together, we'll maintain our stance.
We can see a world of hopeful education
All they can see is sheer banality
We stand for the future and Treaty integration
Resolved to reclaim the university!

Round us all up, make us redundant,
That's what you'd really like to do
But we're too strong, proud, and abundant,
In fact we defy and pity you
You act from fear, why should that be?
What is it that you're frightened of?
A huge drop in pay? A power decline?
The fact we fight for every student's rights?
We're going to win our fight
To work by day and sleep by night

(Repeat refrain)

Encore

Butler-Kisber (2020) sees poetic enquiry as inductive, leading to questions rather than emanating from them. The key question is, however did we come to this juncture? However did it happen that the lunatics took over the asylum? How did corporatisation colonize the place of learning and create the narcissistic faculty manager or the real estate-greedy financial manager? These people are far from being leaders by any measure. What affordances of neoliberalism led some academics to conformist zombiedom, desperate wretchedness or underground resistance? What affordances led others to leverage their technologies of the self and regimes of truth (Foucault, 1982) and thrive as ninjas or alphas and potentially as powderpuff girls: homo oeconomicus. What facets of the modern university generate the perpetual precariat and kept aspirational third-space academics downtrodden? I might ask, with Spina et al. (2022, p. 546), "what are the costs, to academia and to society, of the career patterns we identified in this paper?" It is true that "discourses privilege certain perspectives while creating discursive prohibitions around other points of view in specific times and places" (Spina et al., 2022, p. 535). *Lord, what fools these mortals be!*

I have used my typical characterizations as figures in an ever-evolving landscape or map of resistance. In 2005, we felt a strong feeling of being absolutely forestalled (Davies & Petersen (2005). With more understanding of how neoliberalist ideology from corporatism is operationalized, we developed resolved sites of resistance (Anderson, 2008; Peleas & Peleas, 2011; Mountz et al., 2015; Tett & Hamilton, 2021). These sites might lead us back to self-care and hope (Giroux, 2003; hooks, 2003) and even, perhaps, re-formation (Stanford, 2021; Hil et al., 2023).

Our cabaret opened with a responsory tango duet between a manager and a staff member, and, as with all of these songs, any genders you imagine are unintended, for these people might be anyone or no-one. This is an enactment of regimes of truth. The arrogant self-aggrandisement of the manager, delighting in performative and audit control over subaltern bodies, is sadistic in its cruelty, and there is more than an undertone of sexual harassment to extend the theme of bullying. This is imagined as having the bitter passion of Jenny and Macheath's "Zuhälterballade"/ "Tango ballad" in *The Threepenny Opera*, a sex worker managed by her pimp. We then met the all-focussed ninja, publishing, not perishing, stealing others' work and opportunities and taking all the credit, allying with those corruptibly in power to leverage self-support strategies and ensure hypercitation and impact. Such ninjas share the sociopathy of the manager and will knife a colleague in the back at any remove: the Mack the Knife of the modernized university.

We move, then, to victims of the neoliberalised university. The nervous wreck is imagined fluttering to the hemiola-rich, huapango-based dance. As in Bernstein's *America* from *West-side story* (1957, lyrics: Stephen Sondheim), there is a disjuncture between the content and the metrical order with which it is presented and the sustained rhymes and consonances. The result is a portrayal of anxious, dispelled energy. Through the song, the nervous wreck's desperation and forced actions are revealed, as a cabaret character reveals themselves to the audience with dramatic irony. The zombie is also self-revelatory and shows a shadowy knowledge of who they have been forced to become in a world of blind number-crunching and compliance in the name of quality. Their disjected identity is shown in the frequent enjambment and caesurae.

The super-hard-working third space academic is loyal, conscientious and competent. This person longs for a break and thinks of being stellar like Lee Remick (1935-1991). The reference to Hollywood evokes the Berliner types with their Garbo and vamp aspirations as well as offering a rhyme for 'academic'. The degenerating metrics characterize their sadness. As a modulated construction of university power, Giroux (2014) wrote, precarity defeats dissent by keeping workers preoccupied with the fear of redundancy or loss of identity. Features of the mid-career precariat appear in this character, too, while the early career precariat is a ninja in the making, interviewed promisingly by a job panel until their ruthlessness becomes apparent. The use of the choric, antiphonic responses and playful *ritornelli* between stanzas evokes the Byzantine jazz songs of '50s Greece while paying tribute to the playful and satirical dialogues of Berlin cabaret. The euphony masks a selfish personality, another trope of the cabaret.

The Kabarett spirit is evocative in the parody, 'The Activist', recognizable as 'Das lila Lied.' The proud sexual underground of Berlin become the proud activist underground of the modern university, vowing to save the university as they know it, which is, of course, impossible; but the world needs ardent and honest idealists to remind us of the world we have lost and the compensatory actions we take (Bottrell & Manathunga, 2019). Public displays of intellectualism in the form of resistance are, in the populist world, likely to be seen themselves as zombification (Deslandes & Adamson, 2013). Everyone is someone else's zombie (Whelan et al., 2013). Ball and Olmedo (2012) remind us that we can practice passive resistance: re-imagine power, invalidate audit metrics, redefine productivity and value, and esteem being human. Rebuild a public domain (McBride, 2022). Yet, still, our characters suffer under the authoritarian eyes of Foucault's panopticon (Lorenz, 2012) and performativity (Roberts, 2007; Sparkes, 2008; Ball, 2003, 2012; Craig et al., 2014). Even for the entrepreneurial self, Bröckling (2016) identified just three options: exhaustion, irony or *passive* resistance. In this light (or is it darkness?), the goal of the neoliberalised university is transforming faculty into "an army of temporary subaltern labour" (Giroux, 2014, p. 38). Another army. Another war. Welcome to the cabaret.

References

Acton, R., & Glasgow, P. (2015). Teacher wellbeing in neoliberal contexts: A review of the literature. *Australian Journal of Teacher Education, 40*(8), 99-114.

Anderson, G. (2008). Mapping academic resistance in the managerial university, *Organization, 15*(2), 251-270.

Andrew, M. (2019). Double negative: When the neoliberal meets the toxic. In D. Bottrell. & C. Manathunga. (Eds.), *Resisting neoliberalism in higher education: Volume 1. Seeing through the cracks* (pp. 59–81). Palgrave Macmillan.

Andrew, M. B. (2023). Neo-neoliberalist capitalism, intensification by stealth and campus real estate in the modern university in Aotearoa/New Zealand. *Journal of Applied Learning and Teaching, 6*(2). Advanced online publication.

Baldwin, D. L. (2021). *In the shadow of the ivory tower: How universities are plundering our cities.* Bold Type Books.

Ball, S. J. (2003). The teacher's soul and the terrors of performativity, *Journal of Education Policy, 18*(2), 215-228.

Ball, S. J. (2012). Performativity, commodification and commitment: An I-spy guide to the neoliberal university, *British Journal of Educational Studies, 60*(1), 17-28.

Ball, S. J. & Olmedo, A. (2012). Care of the self, resistance and subjectivity under neoliberal governmentalities. *Critical Studies in Education, 54*(1), 85-96.

Barcan, R. (2013). *Academic life and labour in the new university: Hope and other choices.* Ashgate.

Barker, D. (2017). Ninjas, zombies and nervous wrecks? Academics in the neoliberal world of physical education and sport pedagogy, *Sport, Education and Society, 22*(1), 87–104, https://doi. org/10.1080/13573322.2016.1195360

Barrow, M., & Xu, L. (2021). Constituting ethical academics in teacher education: Navigating multiple and conflicting discourses. *Teaching and Teacher Education, 109*. https://doi.org/10.1016/j.tate.2021.103572

Bauman, Z. (1999). *In search of politics.* Polity Press.

Bird, M. J. (1979). *The Aphrodite inheritance.* TV Series. BBC.

Blackham, A. (2020). Unpacking precarious academic work in legal education. *Law Teacher, 54*, 426–442. https://doi.org/10.1080/03069400.2020.1714276

Bottrell, D., & Manathunga, C. (Eds.) (2019). *Resisting neoliberalism in higher education: Volumes 1 & 2. Seeing through the cracks.* Palgrave Macmillan.

Brecht B., & Weill, K. (1928). *Die Dreigroschenoper/ The Threepenny Opera*. Theater am Schiffbauerdamm.

Bröckling, U. (2016). *The entrepreneurial self.* SAGE.

Butler-Kisber, L. (2020). Poetic enquiry. In E. Fitzpatrick & K. Fitzpatrick. (Eds.), *Poetry, method and education research* (pp. 21-40). Routledge. https://doi.org/10.4324/9780429202117

Clarke, K. (1997). Class act. *Music Magazine (November),* 30-35.

Craig, R., Amernic, J., & Tourish, D. (2014). Perverse audit culture and the modern public university, *Financial Accountability and Management, 30*, 1-24.

Davies, B., & Petersen, E. B. (2005). Neoliberal discourse in the academy: The forestalling of collective resistance. *Learning and Teaching in the Social Sciences*. http://ezproxy.uws.edu.au/login?url=http://dx.doi.org/10.1386/ltss.2.2.77/1

Cantu, C., & Lambert, M. (2013). *Eldorado: Everything the Nazis hate* (Documentary). Netflix.

Deslandes A., & Adamson, K. (2013). Zombie solidarity. In A. Whelan., R. Walker., & C. Moore. (Eds.), *Zombies in the academy* (pp. 67-78). Intellect Ltd.

Fanghanel, J. (2012). *Being an academic.* Routledge.

Fincher, D. (1997). *Se7en*. New Line Cinema.

Fleming, P. (2021). *Dark academia: How universities die.* Pluto Books.

Fosse, B. (1972). *Cabaret*. 20th Century.

Foucault, M. (1980). Truth and power. In C. Gordon (Ed.), *Power/knowledge: Selected interviews and other writings, 1972–1977* (pp. 109–133). Pantheon Books

Foucault, M. (1982). Technologies of the self. In L. H. Martin, H. Gutman., & P. H. Hutton. (Eds.), *Technologies of the self: A Seminar with Michel Foucault* (pp. 16–49). Tavistock Publications.

Foucault, M. (1986). *The care of the self.* Pantheon.

Foucault, M. (1995). *Discipline and punish: The birth of the prison. A. Sheridan* (trans). Vintage.

Galvin, K. T., & Prendergast, M. (2015). *Poetic inquiry II: Seeing, understanding, caring: Using poetry as and for inquiry.* Sense Publishers.

Gill, R. (2009). Breaking the silence: The hidden injuries of neo-liberal academia. In R. Flood & R. Gill (Eds.), *Secrecy and silence in the research process: Feminist reflections* (pp. 228–244). Routledge.

Gillies, D. (2011). Agile bodies: A new imperative in neoliberal governance. *Journal of Education Policy, 26*(2), 207-223.

Giroux, H. (2003). Utopian thinking under the sign of neoliberalism: Towards a critical pedagogy of educated hope. *Democracy and Nature, 9*(1), 91-105. https://doi.org/10.1080/1085566032000074968

Giroux, H. (2014). Public intellectuals against the neoliberal university. In N. K. Denzin & M. D. Giardina (Eds.), *Qualitative inquiry outside the academy* (pp. 35–60). Left Coast.

Giroux, H. A. (2019). *Neoliberalism's war on higher education.* Haymarket.

Goethe, W. (1808). *Faust I & II* (Trans. George Madison). http://goethe.holtof.com/faust/Faust_II_04.htm

Haseman, B. (2006). A manifesto for performative research. Media *International Australia incorporating Culture and Policy (Special issue Practice-led Research), 118,* 98-106.

Hil, R. (2012). *Whackademia: An insider's account of the troubled university.* University of Sydney Press.

Hil, R., Thompsett, F., Lyons, K., Joannes-Boyau, R., Lake, S., Lucas, A., McCallum, A., O'Connor, J., Pelizzon, A., Tregear, P. & Vodeb, O. (2023). Over the horizon: is there an alternative to neoliberal university governance? *Social Alternatives, 41*(1). https://socialalternatives.com/issue/its-time-the-re-form-of-australian-public-universities/

Holden, S. (2013, August 16). The corrupt world of then, so similar to now. *The New York Times*, https://www.nytimes.com/2013/08/17/arts/music/ute-lemper-in-cabaret-at-54-below.html

hooks, b. (2003). *Teaching community: A pedagogy of hope.*

Routledge.

Isherwood, C. (1939). *Goodbye to Berlin.* Hogarth Press.

Jung, J. C. (1970). Collected works of C. G. Jung. In G. Adler G. & R. Hull. (Eds.), *Structure & dynamics of the psyche* (vol. 8). Princeton University Press.

Lareau, A. (2011). Kabarett der Komiker. Berlin 1924–1950. *Monatshefte, 103*(3), 469-471. Https://doi:10.1353/mon.2011.0087

Lawrence, J. (1997). "*The lavender song." Translation of "Das lila Lied" by M. Spoliansky. Berlin Cabaret Songs (English version).* Decca Records. CD.

Lemper, U. (1996). *Berlin cabaret songs.* Decca Records. CD.

Lemper, U. (1997). *Berlin cabaret songs* (English version). Decca Records. CD.

Lemper, U. (1991). *The seven deadly sins/Mahagonny Songspiel by Kurt Weill.* Decca Records. CD

Lester, G. A. (Ed.) (2014). *Three late medieval morality plays: Everyman, mankind and Mundus et Infans: A new mermaids anthology.* Methuen.

Levine, E. (2021). *Allies and rivals: German-American exchange and the rise of the modern research university.* University of Chicago Press.

Lodge, D. (1988). *Nice work.* King Penguin.

Lorenz, C. (2012). 'If you're so smart, why are you under surveillance'? Universities, neoliberalism, and New Public Management. *Critical Enquiry, 38*, 599-629.

Marlowe, C. (1604/2018). *Doctor Faustus*. Methuen

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370-96.

Maslow, A. H. (1987). *Motivation and personality* (3rd ed.). Pearson Education.

McBride, S. (2022). *Escaping dystopia: Rebuilding a public domain.* The Polity Press.

Micali, S. (2010). The capitalistic cult of performance. *Philosophy Today, 54*(4), 379–391.

Middleton, T. (1623/1975). *Women beware women*. Methuen.

Molpeceres, S. (2014). *Mito persuasivo y mito literario: Bases para un análisis retórico-mítico del discurso.* Servicio de Publicaciones de la Universidad de Valladolid.

Molpeceres, S. (2017). The zombie: A new myth in the making. A political and social metaphor. *Journal of Comparative Literature and Aesthetics, 40*(2), 151-167.

Mountz, A., Bonds, A., Mansfield, B., Loyd, J., Hyndman, J., Walton-Roberts, M., Basu, R., Whitson, R., Hawkins, R., Hamilton, T., & Curran, W. (2015). For slow scholarship: A feminist politics of resistance through collective action in the neoliberal university. *ACME: An International E-Journal for Critical Geographies, 14*(4), 1235-1259.

Pitt, R., & Mewburn, I. (2016). Academic superheroes? A critical analysis of academic job descriptions. *Journal of Higher Education Policy and Management, 38*(1), 88–101. doi:10.1080/ 1360080X.2015.1126896.

Peleas, R. J., & Peleas, M. (2011). A Screed, a surrender, and a summons: Facing the political surround. *Cultural Studies Critical Methodologies, 11*(4), 418–420.

Quiggan, J. (2010). *Zombie economics: How dead ideas still walk among us.* Princeton University Press.

Rapport, F., & Hartill, G. (2012). Crossing disciplines with ethnographic poetic representation. *Creative Approaches to Research, 5*(2), 11–25.

Readings, B. (1996). *The university in ruins.* Harvard University Press.

Richardson, L. (2000). Writing: A method of inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The handbook of qualitative research* (2nd ed., pp. 923–949). Sage.

Roberts, P. (2007.) Neoliberalism, performativity and research. *International Review of Education, 53*(4), 349–365.

Ryan, S. (2012). Academic zombies: a failure of resistance or a means of survival? *Australian Universities Review, 5*4, 3–11.

Shakespeare, W. (2005). *William Shakespeare: The complete works.* Oxford University Press.

Shore, C. (2010). Beyond the multiversity: Neoliberalism and the rise of the schizophrenic university. *Social Anthropology, 18*(1), 15-29.

Singer, B. (2000). When cabaret had an edge. *The New York Times*. https://www.nytimes.com/2000/02/27/arts/music-when-cabaret-had-an-edge.html

Skea, C. (2021). Emerging neoliberal academic identities: Looking beyond Homo economicus. *Studies in Philosophy and Education, 40*, 399–414. https://doi.org/10.1007/s11217-021-09768-7

Smyth, J. (2017). *The toxic university.* Palgrave.

Soars, A. (2002). *The decline of privilege: The modernization of Oxford University.* Stanford University Press.

Solomon, S., & Du Plessis. M. (2023). Experiences of precarious work within higher education institutions: A qualitative evidence synthesis. *Frontiers in Education, 8.* https://doi.org/10.3389/feduc.2023.960649

Sparkes A. C. (2007). Embodiment, academics, and the audit culture: A story seeking consideration. *Qualitative Research,* *7*(4), 521-550. https://doi.org/10.1177/1468794107082

Spina, N., Smithers, K., Harris, J. & Mewburn, I. (2022). Back to zero? Precarious employment in academia amongst 'older' early career researchers, a life-course approach. *British Journal of Sociology of Education, 43*(4), 534-549, https://doi.org/10.1080/01425692.2022.2057925

Spoliansky, M. (1996). *"Das Lila Lied." Berlin cabaret songs. Ute Lemper*. Decca Records. CD.

Spoliansky, M. (1996). *"Wenn die beste Freundin." Berlin cabaret songs. Ute Lemper.* Decca Records. CD.

Standing, G. (2011). *The precariat: The new dangerous class*. Bloomsbury Academic.

Stanford University (2021). *The story behind the modern university* (Podcast). https://ed.stanford.edu/news/story-behind-modern-university

Tett, L., & Hamilton, M. (2021). *Resisting neoliberalism in education: local, national and transnational perspectives*. Policy Press.

Tregear, P., Guthrie, J., Lake, S., Lucas, A., O'Connor, J., Pellzzon, A., & Vodeb, B. (2022). 'Enough to make you sick!' Pathological characteristics of the Australian academic workplace. *Social Alternatives, 41*(1), 44-51.

Varea, V., Riccetti, A., & Siracusa. M. (2021). The Powerpuff Girls: Making it as early career academics in physical education. *Journal of Gender Studies, 30*(6), 687-698, https://doi.org/10.1080/09589236.2021.1937081

Whelan, A., Walker, R., & Moore, C. (2013). *Zombies in the academy* (pp. 67-78). Intellect Ltd.

Whitchurch, C. (2013). *Reconstructing identities in higher education: The rise of third space professionals.* New Routledge.

Wiebe, S. (2015). Poetic inquiry: A fierce, tender, and mischievous relationship with lived experience. *Writing and Literacy.* https://doi:10.20360/G2VP4N

Wieland, K. (2011). *Dietrich & Riefenstahl. Hollywood, Berlin and a century in two lives.* Liveright.

Wouilloz-Boutrois, F. (2021). *La fabuleuse histoire des cabarets Berlinois.* L'Harmattan.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Educational futures of intelligent synergies between humans, digital twins, avatars, and robots - the iSTAR framework

Ronghuai Huang[A]
Ahmed Tlili[B]
Lin Xu[C]
Ying Chen[D]
Lanqin Zheng[E]
Ahmed Hosny Saleh Metwally[F]
Ting Da[G]
TingWen Chang[H]
Huanhuan Wang[I]
Jon Mason[J]
Christian M. Stracke[K]
Demetrios Sampson[L]
Curtis J. Bonk[M]

A *Professor, Smart Learning Institute of Beijing Normal University, Beijing, China*
B *Associate Professor, Smart Learning Institute of Beijing Normal University, Beijing, China*
C *Research Assistant, Smart Learning Institute of Beijing Normal University, Beijing, China*
D *Postdoctoral Researcher, Smart Learning Institute of Beijing Normal University, China*
E *Associate Professor, Faculty of Education, Beijing Normal University, China*
F *Postdoctoral Researcher, Smart Learning Institute of Beijing Normal University, Beijing, China*
G *Postdoctoral Researcher, Smart Learning Institute of Beijing Normal University, Beijing, China*
H *Associate Research fellow, Smart Learning Institute of Beijing Normal University, China*
I *Assistant Professor, Smart Learning Institute of Beijing Normal University, China*
J *Associate Professor, Faculty of Arts & Society, Charles Darwin University, Darwin, Australia*
K *Coordinator Virtual Collaboration & Cloud, University of Bonn, Bonn, Germany*
L *Professor, Department of Digital Systems, University of Piraeus, Piraeus, Greece*
M *Professor, Instructional Systems Technology Department, School of Education, Indiana University, Bloomington, USA*

Keywords

Artificial Intelligence (AI);
future education;
human-machine collaboration;
human-machine interaction;
team roles.

Correspondence

ahmed.tlili23@yahoo.com [B]

Article Info

Received 18 September 2023
Received in revised form 16 October 2023
Accepted 16 October 2023
Available online 18 October 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.33

Abstract

With the rapid advances of Artificial Intelligence (AI) and its technologies, human teachers and machines are now capable of collaborating to effectively achieve specified outcomes. In educational settings, such collaboration requires consideration of several dimensions to ensure safe, responsible, and ethical usage. While various research studies have discussed human-machine collaboration or cooperation in education, a framework is now needed that aligns with contemporary affordances. Providing such a framework can help to better understand how human teachers and machines can team up in education and what should be considered while doing so. To address this gap, this paper outlines the iSTAR (Intelligent human-machine Synergy in collaborative teaching: utilizing the digital Twins, Avatars/Agents and Robots) framework. iSTAR represents human-machine collaboration as an ecosystem that goes beyond the simple collaboration between human teachers and machines in education. Therefore, it presents core dimensions of DELTA (design, ethics, learning, teaching and assessments) that should be considered in designing safe, responsible, and ethical learning opportunities.

Introduction

Can machines think? is a simple yet sophisticated question (Turing, 1950). In response to this question, a scholarly event was organized in 1955, where the term "artificial intelligence (AI)" was coined to refer to machines and processes that imitate human cognition and make decisions like humans (McCarthy et al., 2006). The Turing Test was proposed, originally known as the imitation game, as a protocol to determine whether a machine can exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human. Such developments proved pivotal in the emergence of cognitive science and debates about 'what computers could or could not do' that shaped much of the early research in this interdisciplinary field (Dreyfus, 1992). A few decades earlier, the term [ro]bot had also been articulated for the first time in Čapek's (1921) science fiction play; however, it was Asimov (1942; 1950) who visioned that these machines could transform into intelligent forms which led him to introduce the *three laws of robotics* to set the rules that bots should stick to.

Not so long ago, current advancements were depicted as science fiction. With the rapid evolution of computational power and access to massive data, however, such capabilities are being realized through Large Language Models (LLMs). AI machines, such as ChatGPT, are now capable of maintaining conversations like humans. As 'conversational agents', these capabilities represent a major innovation extending beyond the information processing of search engines (Mason, 2023). These technological advancements are finding application in various domains, including education. For instance, Intelligent Tutoring Systems (ITS) have matured as a potent educational tool, harnessing AI technologies to deliver personalized and adaptive learning experiences for individuals or groups of students. By integrating sophisticated algorithms and cognitive models, ITS can assess students' level of knowledge in a given subject domain, monitor their progress over a course of learning actions, and provide targeted instructional interventions (including the recommendation of resources to study and practice, guidance and feedback) (Koedinger & Aleven, 2016).

Alongside ITS, computer-based teaching systems like Plato (Programmed Logic for Automated Teaching Operations), originally developed in 1960 (Dear, 2017), have garnered significant recognition in educational contexts. Plato acts as a comprehensive platform for managing courses, delivering content, and conducting assessments, offering teachers a centralized hub to organize and disseminate instructional materials effectively (Dear, 2017; Jones, 2015). Furthermore, the integration of AI companions in education has demonstrated tremendous potential (Sharples, 2022). AI companions, exemplified by ChatGPT, can engage in conversations with students, offer explanations, address queries, and provide guidance; thereby emulating human-like interactions and supporting learners throughout their educational journey (Tlili et al., 2023a; Adarkwah et al., 2023). These remarkable technological advancements have the capacity to transform education by enhancing personalized learning experiences, fostering student engagement, and providing timely support and feedback.

Such technological advancements also raise questions (e.g., Selwyn, 2023; Tlili et al., 2023a) on how human teachers and machines could work together to achieve an educational objective, as well as the meaningful, transformative changes brought to education (e.g., evolutionary or revolutionary). Schmidtler et al. (2015) observed that the relationship between humans and intelligent machines has shifted from human–machine co-existence and cooperation to human–machine collaboration. In business contexts, the notion of 'collaborative intelligence' has also been used to describe how "(h)umans and machines can enhance each other's strengths" (Wilson & Daugherty, 2018, p. 114). Moreover, Schmidtler et al. (2015) used the term 'synergy' as they believed that education is a complex task that requires more than simple collaboration. Such synergy between the human teacher and the machine (i.e., their combined effect is greater than the sum of their separate effects) is crucial to achieving the desired learning objective.

In this context, several studies pointed out that "humans and machines have complementary capabilities that can be combined to augment each other" (Dellermann et al., 2019, p. 4). Scholars (e.g., Gerber et al., 2020) working on hybrid intelligence or human-machine symbiosis further pointed out that excellent outcomes are possible when the abilities of humans and machines are combined in a mutually beneficial exchange (Dellermann et al., 2019). Consequently, human-machine collaboration is referred to in this present study as 'intelligent human-machine synergy during collaborative teaching'. We define human-machine synergy during collaborative teaching *as the way human teachers and machines interact and work together in several educational settings as a team to achieve a common objective, resulting in enhanced learning outcomes. This synergy combines Human Intelligence (HI) and Artificial Intelligence (AI) to achieve Collaborative Intelligence (CI) in education*. Thus, in this era where IT can be depicted as 'intelligent technology', it is crucial to explore and investigate how human teachers and machines could work together to achieve this synergy and collaborative intelligence for future education.

Similarly, the Beijing Consensus on AI (UNESCO, 2019) calls for using AI to empower teaching and teachers. It suggests that related bodies should dynamically review and redefine teachers' roles and required competences in the context of teachers' training policies and capacity-building programmes for better preparation of teachers to work effectively in AI-rich education settings. Kaber (2018) mentioned that one of the core questions in human-machine collaboration is 'Who does what?'. In the same vein, Vuorikari et al. (2020), through the analysis of eight future-oriented scenarios, highlight 'the ethical considerations (including the balance between human autonomy and machines) and the evolving competence requirements of teaching professionals.' It is, therefore, important to further investigate the different roles that human teachers can take in collaborative teaching with machines.

Furthermore, given the rapid progress in AI development and application, it is most important to address ethical questions and issues on the usage of AI and machines in educational settings and systems (Holmes et al., 2023). In particular, the legal and societal responsibility of human control over AI and machines (that is always a human one) needs to be reflected in introducing and using terms such as human-machine collaboration and synergy. We discuss these urgent and challenging aspects in full detail later.

To sum up, human-machine collaboration can change how we live and do daily tasks and activities. Aspects of human-machine collaboration have been investigated in several fields, such as economy (Bolton et al., 2018), managerial decision making (Haesevoets et al., 2021), and health monitoring (Muin & Mosalam, 2021). However, very few studies have proposed a viable model or comprehensive framework for human-machine collaboration in education. There is still a lack of information on what types of machines teachers could collaborate with in education and how to ensure an effective collaboration for a safe learning experience and enhanced learning outcomes. To close the various research gaps, this study presents a conceptual analysis and proposes an innovative human-machine collaboration framework, namely the iSTAR framework. The iSTAR framework focuses on intelligent human-machine Synergy in collaborative teaching through utilizing Digital Twins (DT), Avatars, and Robots (iSTAR). It aims to present the different types of machines that human teachers could collaborate with in education, as well as different levels of possible collaboration. Additionally, the iSTAR framework signals how human teachers could work with machines and how their roles can be reformed in the era of so-called 'intelligent technologies', keeping in mind different dimensions for a safe and effective learning environment. The iSTAR framework was developed based on a rapid review, which is defined as "a type of knowledge synthesis in which components of the systematic review process are simplified or omitted to produce information in a short period of time" (Tricco et al., 2015, p. 2).

Theoretical background: Human-machine collaboration

Several theoretical foundations can be identified in the emergence of human-machine collaboration. Hoc (2000) observed that the trend toward increased complexity and coupling of Information Technology (IT) systems required a new conception of human-computer interaction (HCI) that signalled the role of interfacing with automated systems. He argued that a "human-machine cooperation (HMC) approach is necessary to address the new stakes introduced by this trend" (Hoc, 2000, p. 833). This construct can also be found in earlier literature (Vanderhaegen et al., 1994). Likewise, the foundational terminology of 'socio-technical systems' was coined around 1960 in the context of labor studies "to stress the reciprocal relationship between humans and machines" (Ropohl, 1999, p. 186). The terminology of socio-technical systems (STS) and human-machine cooperation (HMC) is now embedded in the literature, and both place emphasis on systems interoperability. Thus, systems science can be regarded as a pivotal foundation of the more recent construct of 'human-machine collaboration'.

Further investigation of the origins of human-machine collaboration are revealed in the development of Man-Machine-Environment System Engineering (MMESE), a fundamental principle of human-centered system design. This principle was initially introduced by Professor Shengzhao Long in 1981, with the influential support of esteemed Chinese scientist Xuesen Qian (Guo et al., 2022). MMESE has developed as a research field that uses system science theory and system engineering methods to efficiently handle the relations between humans and machines with a view to achieving an "optimal combination of man-machine-environment system" (Long & Huang, 2022, p. 3). During the past 40 years, MMESE has been developed and applied to many areas, such as automation systems, shipboard equipment, aircraft systems, finance, etc. Notably, the three goals of the optimization of MMESE are safety, efficiency, and economy (Long & Huang, 2022).

Card (2018) observed that human-machine collaboration as a research field is different from human-machine interaction because it goes beyond interaction and information presentation theories to include team- and group work. Human-machine collaboration has been tackled in the literature from various perspectives, most commonly from the different levels of automation (Vagia et al., 2016), where fully manual implies that humans are fully in control, while fully automated implies that humans, as operators, are completely out of the loop (Parasuraman et al., 2000). However, less consensus exists in the literature about the scales between 'fully manual' and 'fully automated'. Consequently, several taxonomies have been put forward discussing the various automation levels (Saurin & Patriarca, 2020; Simmler & Frischknecht, 2021).

Most of the aforementioned frameworks are technical-focused, neglecting the importance of effective synergy between humans (in this study, human teachers) and machines to achieve a collaborative activity (in this study, collaborative teaching). In the field of Intelligent Tutoring Systems (ITS), however, this consideration is more prevalent (Longo et al., 2017). Dellermann et al. (2019) stated that for effective human-machine collaboration, the machine should not deal with all of the roles in a team but should instead be built to complement human activity and intelligence (collaborative intelligence).

This idea has persisted in education, where Vuorikari et al. (2020) discuss three different approaches, namely, teacher-in-the-loop, teacher-over-the-loop and teacher-out-of-the-loop to deal with the distribution of responsibility between human teachers and an algorithm/machine in educational applications and services that rely on autonomous decision-making (e.g., AI). Han and Huang (2023) further articulated the idea that machines should empower human teachers and collaborate with them to better achieve a given educational goal. In many contexts, this augmentation role requires machines to be designed with human-like abilities, enabling them to act like humans (Nass & Moon, 2000). This vision has led to the development of the *computers are social actors theory* (or social response theory), which highlights that "humans mindlessly apply the same social rules used for human interactions to computers" (Nass & Moon, 2000, p. 669). This theory emphasizes anthropomorphism, attributing human characteristics to non-human actors (Qiu & Benbasat 2009; Watson 2019).

These characteristics could involve several aspects, including appearance, behavior, reasoning, etc. Following on from this theoretical perspective, the present study identifies two human-like machines (that can also be called technologies) that could collaborate with human tutors, namely *(1) physical robots* and *(2) avatars/agents*. The first type of machine (technology) allows collaborative teaching in physical spaces (e.g., classrooms), while the second type of machine (technology) allows collaborative teaching in cyberspaces. Specifically, Han et al. (2023) further point out the importance of providing realistic cyberspaces to enhance human-machine collaboration in education. To achieve this, the present study adopts *digital twins* as another important technology in human-machine collaboration in education.

Importantly, this study is also informed by the literature associated with human-centered design in the development of technology, as it provides guidance on principles for maintaining the preeminence of human agency within complex systems environments (Dart et al., 2019; Giacomin, 2014; ISO, 2019). Moreover, while human-machine collaboration is informed by this work, it also extends the scope. For example, due to advances in the Internet of Things (IoT), Cruickshank and Trivedi (2017) point out that in an IoT environment, a 'user' might be a toaster!

To summarize, to ensure human-machine synergy in collaborative teaching, this study proposes the iSTAR framework, which builds on the three identified enablers (technologies), namely (1) digital twins, (2) avatars/agents, and (3) physical robots, that can be intelligently tuned into synergistic relations with human input. The overall framework places a human at the centre of these three enablers, making explicit the relations with the human teacher. Details are presented in the next section.

iSTAR framework

Figure 1 shows the iSTAR framework, which depicts the various dimensions of human-machine collaborations (HMC) based on the three identified machines, namely digital twins, avatars/agents, and physical robots. Particularly, the iSTAR framework places the human teacher at the center. This means that the design of human-machine collaborative teaching should be human-centered, and machines should be used as enablers to augment human teachers (Dede et al., 2021; Dede et al., 2017) rather than replace them.

iSTAR dimensions

The three dimensions of iSTAR are described below.

Digital twins

A digital twin is the digital representation of a physical object, person, or process contextualized in a digital version of its environment. As one of the main technologies associated

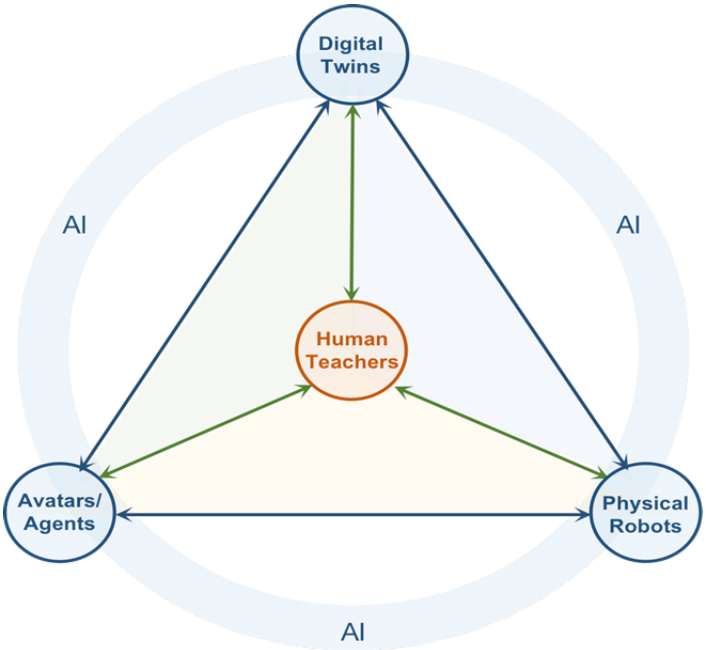


Figure 1. iSTAR framework.

with Industry 4.0, the terminology of 'digital twin' was first proposed by Michael Grieves, to mitigate issues leading to undesirable and unpredicted emergent behavior at the phases of creation and production and realized during the operational phase in complex systems (Sepasgozar, 2020). Prior research highlights that digital twin systems consist of two sub-systems, the physical system and a virtual system, which contain all essential information about the physical system (Liljaniemi & Paavilainen, 2020). Data flows between the physical-digital objects, which are fully integrated in both directions. This enables the virtual system to represent, monitor, and issue commands to the physical system while also understanding, evaluating, and predicting the state of the physical counterpart, generating insights and suggestions to optimize the system's performance throughout the lifecycle. Digital twin systems are becoming more and more common in the areas of manufacturing, finance, aerospace, etc. There are also attempts in education to investigate methods, benefits, and barriers to adopting digital twin technology (Tlili et al., 2023c).

Avatars and agents

Avatars and agents have demonstrated their effectiveness as valuable tools in education (Segaran, 2021). In past research, these two terms have been used interchangeably. Nevertheless, it is crucial to recognize that the two possess distinctive attributes and characteristics. Bailenson and Blascovich (2004) define avatars as "a perceptible digital representation whose behaviors reflect those executed, typically in real-time, by a specific human", while an agent is "a mathematical or computational formula designed to achieve a specific goal" (p. 65). In other words, the controller is one of the main differences between avatars and agents. Avatars are human-controlled representations of persons or other entities, whereas agents are computer-controlled representations. Agency is a broad term, however, and it should be noted that avatars could act as agents, with some control residing in the computer instead of a human. Such variations in control are seen in video games, where these avatars are referred to as Non-Player Characters (NPC). In addition, avatars and agents appear differently. Avatars are typically graphical representations, such as 3D/2D models or images, representing the user's visual persona (Blake & Moseley, 2010). Nevertheless, agents may not have a graphical representation as software programs or systems, such as chatbots, or they may have a graphical appearance to enhance social interaction (Baylor, 2011).

Physical robots

Robots can be used in intracurricular and extracurricular activities (Mubin et al., 2013). They can have different roles, including being used as learning tools/teaching aids (robotics education) or as co-learners, peers or companions, mentors, and tutors (Mubin et al., 2013). The robot's appearance has evidently affected students' responses and interactions with an education robot in different stages, from junior grade to undergraduate level. Junior-grade students prefer a toy-like robot with a cute design; middle-grade students care about the appearance of anthropomorphic robots; senior-grade students will keep interest in a robot if its responses are non-repeating; and undergraduate students will care about the functionalities of an education robot (Sun et al., 2018).

iSTAR scenarios

As shown in Table 1, various levels of human-machine collaboration can be conceived. Level 0 *depicts humans using machines simply as tools* (e.g., calculators) without any collaboration. Level 1 represents *Basic Human-Machine Collaboration (HMC)*, while Level 2 represents *Dual Human-Machine Collaboration (HM²C)*. The difference between these two levels is in terms of the established collaboration between the human teacher, the machine and the learning space.

Specifically, in basic Human-Machine Collaboration (HMC), collaborative teaching is established between the human teacher and only one machine type, which could be digital twins (HMC1), avatars/agents (HMC2), and physical robots (HMC3). Additionally, the learning space is either physical or cyber. With Dual Human-Machine Collaboration (HM²C), on the other hand, more complex collaborative teaching activities are enabled, where various types of machines could also collaborate, in addition to the human teacher, to achieve an educational objective. Therefore, M² represents two or more types of machines working together and amplifies the level of collaboration between machines. Additionally, it is seen that the learning space is becoming more complex, where a possible real-time collaboration in physical and cyber spaces could occur. Finally, *Complex Human-Machine Collaboration (HM^{n}C)* depicts the future development of this field, where human teachers could work with several machines in a balanced and safe ecosystem of humans and machines to achieve a specific educational goal.

Table 1. Classification of Human-Machine Collaboration.

Level 3	Complex Human-Machine Collaboration (HM^{n}C)	Humans collaborate with more than two types of machines
Level 2	Dual Human-Machine Collaboration (HM²C)	Digital Twins & Avatars/Agents
		Digital Twins & Physical Robots
		Avatars/Agents & Physical Robots
Level 1	Human-Machine Collaboration (HMC)	Digital Twins
		Avatars/Agents
		Physical Robots
Level 0	Human-Utilize machine (HUM)	Humans simply use machines as tools

Examples of each human-machine collaboration (Level 1 and Level 2) within the iSTAR framework are described below.

Level 1: Human-machine Collaboration (HMC)

Based on Figure 1, various educational scenarios are found in the literature related to the Level 1 of Human-Machine Collaboration, as follows:

HMC1 (human teachers-digital twins)

Kaarlela et al. (2022) introduce a novel robotics teleoperation platform supported by the emergence of Industry 5.0. The platform described by Kaarlela et al. (2022) is based on digital twins with bi-directional data transmission between the physical and digital counterparts. The proposed system allows teleoperation, remote programming, and near real-time monitoring of controlled robots, robot time scheduling, and social interaction between users. Teachers can use the platform as a teaching tool, cooperating with students to finish robotic programming.

HMC2 (human teachers-avatars/agents)

Mizrahi et al. (2022) presented a novel system for facilitating small group online talks using an avatar during video conferencing, where avatars act as agents to support teaching and learning. The avatar was pre-programmed, whereas the course instructors created the material for the activities. Students from the tenth grade interacted with the system in groups, where Mizrahi et al. (2022) compared avatar-facilitated activities to unfacilitated activities. The findings demonstrate that when compared to activities without avatar facilitation, students felt the activity with the avatar was much more efficient, more understandable, and encouraged more involvement. In addition, students were more likely to speak with avatar facilitation.

To provide inclusive education for deaf students, Brazil teachers collaborated with animated 3D avatars as the latter served as sign language translators. Using the avatar translator, Spanish speech can be translated to Spanish sign language through speech recognition technology. Therefore, it can translate what the teacher is saying automatically, allowing deaf and hard-of-hearing students to follow the instructor as easily as their hearing peers (De Martino et al., 2016).

HMC3 (human teachers-physical robots)

Following the pioneering work of Seymore Papert in the 1960s, there is already an extensive history of robots being successfully used in classrooms to teach programming concepts (Resnick, Ocko, & Papert, 1988). More recently, Kindergarten Social Assistive Robotics (KindSAR) is a practical example of using robots in contemporary education settings (Keren & Fridin, 2014). With the help of this cutting-edge technology, kindergarten teachers now have a creative means of fostering social learning. It has previously been shown that children in a preschool setting gain from using the KindSAR robot to play educational games (Keren & Fridin, 2014). An interactive robot worked as a teacher's aide, reading small groups of kids taped stories while combining songs and motor activities. For instance, KindSAR tracks children's development over time while giving children and the teaching staff detailed feedback on how well they performed in the game or assignment. Then, the kindergarten staff can use the visual and audio task performance data and feedback for further teaching design. The findings indicate that the kids respected the robot's authority and enjoyed engaging with it. This study reveals that implementing KindSAR in preschool education is feasible and will have the desired effects (Keren & Fridin, 2014).

Level 2: Dual Human-machine Collaboration (HM²C)

Based on Figure 1, various educational scenarios are found in the literature related to Level 2 of human-machine collaboration, as follows:

HM²C1 (human teachers-avatars/agents-digital twins)

Virtual Reality Learning Environments (VRLEs) describe a platform that uses digital twins to create a learning environment similar to the physical one, where students and teachers use their avatars to conduct various tasks within the designed virtual laboratories (Lugrin et al., 2016). In VRLEs, teachers and students can communicate and collaborate with peers and engage in educational tasks providing and receiving real-time feedback (ibid). One crucial aspect of the VR learning experience is the utilization of the multi-user VRLEs, as shared spaces or worlds, where students and educators have extensive control over individual 3D avatars. It was reported that the combination of multi-user VRLEs and avatar representation can enhance students' and teachers' engagement and performance in the teaching and learning process (O'Connor et al., 2018; Schild et al., 2018). For instance, in the TeachLivETM platform, teachers are immersed in a real-time 3D simulation of a classroom with a head-mounted display and headphones (ibid). Importantly, their body motion and facial expressions can be captured in real-time and projected onto a high-fidelity avatar. In addition, teachers can carry out the lectures and communicate with students in the virtual environment through the avatars.

HM²C2 (human teachers-avatars/agents-physical robots)

It is better for teachers to build a learning environment using digital reality in conjunction with robots due to the constraints of time and space as well as the restricted interaction capabilities of robots. For example, Al Hakim et al. (2022) developed an interactive situated learning approach to enhance students' learning performances. In their approach, students and robots role-play characters and immerse themselves in digital situated learning tasks and challenges. In addition, robots can provide real-time feedback to guide and assess how well students are applying their knowledge, based on pre-setted agents. The evaluation was conducted during interactions with the robot, virtual objects, and virtual characters based on textbook context and content. This approach encourages human-robot interaction while allowing students to study and engage in any situation relevant to the textbook subject that can be efficiently digitalized.

HM²C3 (human teachers-digital twins-physical robots)

The substantial expense associated with deploying robots and their utilization in widespread educational settings, particularly in underprivileged and distant regions, presents a formidable challenge. There is also a risk of causing personal injury when using robots in teaching. One way to solve this problem is to 'virtualize' robots. For instance, Shahab et al. (2021) developed Virtual Reality Robots (V2R) based on the social robot NAO to conduct music education for children with high-functioning autism. Virtual humanoid robots teach children with autism to play instruments in the virtual classroom. Human tutors act as an operator to control virtual robots and give assessments. For skills such as handwriting, painting, and driving, which require hands-on instruction, it is often difficult to achieve remote teaching. However, a haptic-based training system provides a solution in this scenario.

In the haptic-based training system, a network connecting two haptic devices plays the role of putting hands together, based on which the algorithms of haptic guidance and correction in real-time are developed for skill transmission between human experts and trainees (Liu et al., 2013). Solis et al. (2002) used haptic interfaces as cooperative systems to reproduce and simulate human actions, such as teaching people to write Japanese characters. The Reactive Robot systems provide the capability of interpreting the human teacher's actions and exert a more intelligent force feedback strategy.

Robots and digital twins can collaborate in effective teaching. In a classroom or at home, the robot sits next to the student. The educational lesson is displayed on a computer screen in front of the student and the robot. Through the online interface, a remote teacher can teach the student from anywhere in the world using a robot. A student who interacts with a robot in this way might behave similarly to kids who read to dogs in the Reading with Rover program (http://www.readingwithrover.org/) (Lee et al., 2008). This program has demonstrated that students do better when they read to a dog rather than a stranger adult because they are less anxious. It is plausible that a friendly robot character may elicit the same reaction. Additionally, because the robot is always under the control of the instructor, the instructor is prepared to respond to a student who is distracted from the course or poses a spontaneous question. The training system does not operate as a fully autonomous robot since current systems are still incapable of effectively managing the intricacies of human behavior.

A new form of teaching is a robot system for English classes, which utilizes a teleoperated robot controlled by a teacher from a remote site. By providing a unique operation interface that incorporates non-contact vision recognition technologies, a teacher can easily control the robot from a distance to provide lectures (Yun et al., 2013).

The development of human-robot interaction strategies, such as promoting trust between humans and machines in high-stakes situations like emergency response, will be greatly impacted by this digital twin (Pairet et al., 2019). Additionally, it will enable the evaluation of task planning algorithms for collaborative inspection and long-term autonomy, as well as human-guided supervision and management of the robotic assets from a remote-control station.

iSTAR considerations

This section introduces and discusses the considerations in relation to Design, Ethics and Learning, Teaching and Assessment (DELTA) aspects that need to be addressed when designing iSTAR implementations. Considering these aspects can support developing responsible and effective human-machine collaboration in education. Each of the DELTA considerations is discussed in the subsequent sections.

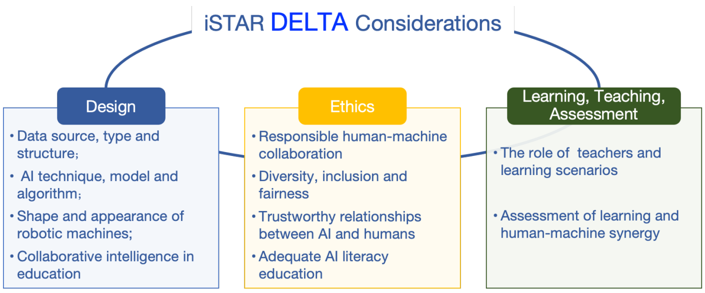


Figure 2. DELTA considerations for designing iSTAR.

Design

Data source, type and structure

The data produced through human-machine synergy is diverse and can support collaborative teaching in various ways. It can be collected from both the human and the machine to enhance the achievement of educational objectives. Human data could include their log data, behaviors, and facial expressions, among others. The machine data could include response accuracy, execution and feedback time, and educational support level, among others. For instance, feedback can be provided to the machine through various inputs such as gestures, behavior, touchscreens, keyboards, and voice commands. This feedback can be used to improve the machine's performance and adapt its behavior to better meet the needs of the users. For example, data from students' gestures or answers could be used to assess their understanding of the material and provide personalized feedback or instruction.

Robots can be equipped with various sensors to collect multimodal data such as audio, visual, and haptic feedback in teaching and learning environments. Cameras, for instance, can capture facial expressions, student gestures, and behavior; microphones can capture speech; eye-tracking devices can capture eye movements; and wearable sensors can obtain various physiological signals such as heart rate and brain waves. In this way, this data can provide valuable information about the interactions between the machine and the students or teachers by providing further data for analysis and improvement of the learning experience (Stracke & Skuballa, 2021).

Data types and structures in human-machine collaboration include multimodal data, which refers to data from different sensors or modes such as text, image, speech, video, notes, logs, gestures, sensors, behavior, and feedback. One of the challenges that come with multimodal data is standardization and interoperability (Yeo & Nielsen, 2020). An approach to aligning multimodal data, therefore, is to map the data from different modalities into a common representation space and then perform alignment in that space. This can be achieved by training a multimodal deep neural network (Summaira et al., 2021; Jabeen et al., 2022).

To ensure an effective human-machine collaboration in education, based on the foregoing discussion, it is vital to analyze what type of educational data from the human or the machine should be collected (e.g., learning navigation behavior, facial expression, etc.) depending on the key educational goals (prediction, personalization, etc.). The use of 'should' here is because there are ethical imperatives to consider when using human data. Additionally, it is important to study how the rich data could be standardized so it can be analyzed to reveal more insights about the educational process (Sampson et al., 2022).

AI technique, model and algorithm

AI technologies such as machine learning (ML), including reinforcement learning (RL), unsupervised learning (UL), and supervised learning (SL), demonstrate a huge potential in their usage and application in education covering many diverse scenarios (Bozkurt et al., 2023) and have been used in various ways to support human-robot collaboration (HRC) (Duan et al., 2019). Among them, unsupervised learning has been used to model human behavior and predict human intentions. Supervised reinforcement learning has been used to improve robot perception and recognition of human actions (Semeraro et al., 2021).

Reinforcement learning (RL) determines a policy for an agent to maximize a cumulative reward through learning by interaction with the environment. RL has been used to optimize task allocation between humans and robots. For example, the system may receive rewards or penalties when students succeed or fail to complete a learning task with robots. By continuously updating the allocation strategy based on this feedback, the system can learn to make better decisions over time and improve its performance. Unsupervised learning (UL) focuses on finding patterns in unlabeled data and can be used to analyze and cluster data to uncover hidden structures or relationships within the data. For example, UL models could be used to cluster students based on their learning preferences, allowing for more personalized instruction. Supervised learning (SL) involves training a model using labeled data. SL can be applied in teaching and learning through the integration in intelligent tutoring systems (ITSs). For example, ITSs use SL algorithms to identify areas where a student is struggling and provide targeted instruction or practice problems to help the student improve.

Deep learning (DL) is a collection of techniques and methods for using (artificial) neural networks to solve ML tasks, whether SL, UL, or RL. DL based on Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Transformers, and other frameworks is widely employed in HMC. Among them, Recurrent Neural Networks (RNNs) can effectively incorporate temporal dependencies, such as Long Short-Term Memory (LSTM) networks, while Transformers are based on an attention mechanism.

Deep neural networks have been successfully applied to computer vision, natural language processing, recommender systems, speech recognition, and other tasks. Computer vision includes object recognition and semantic segmentation. For example, a standing posture recognition system combined with vision-based techniques and deep learning can be utilized to detect the operator's posture and predict the operator's intended action in human–robot collaboration (Li et al., 2020). Applications of natural language processing (NLP) such as sentiment analysis, question answering, machine translation, and other tasks have been used to improve communication between humans and machines by enabling machines to understand and respond to natural language input from humans. Notably, ChatGPT is powered by a large language model (LLM) which can provide just-in-time feedback based on GPT (generative pre-trained transformer) in Human-AI Collaboration (HAC) (Sharma et al., 2022).

Shape and appearance of robotic machines

Robotic machines can be in different shapes, including humanoid, semi-humanoid, animal-like, etc. The shape and appearance of machines can largely affect their effect on users, as reported by numerous researchers. Differences in gender, skin color, size, and position of facial parts of machines influence the decision-making, impression and judgment by users, which makes the design of their shape and appearance especially important. One theoretical example is the Uncanny Valley Effect (UVE). Uncanny Valley was firstly proposed by Masahiro Mori to hypothesize about people's reactions to robots that looked and acted like humans. According to the theory, a person's response to a humanlike robot would abruptly shift from empathy to revulsion as it approached but failed to attain a lifelike appearance (i.e., robots should be almost just like humans, not only in terms of appearance but also in terms of touch, feelings, movements, etc.).

Given such predictions, the design of humanoid robots should certainly avoid the uncanny valley. Apart from that, the shape and appearance of a machine should be designed according to the educational field and level. For example, a robot-like agent that teaches robot history is more effective than a robot-like agent that teaches humanities (Matsui & Yamada, 2019). Another study (Ringwald et al., 2022) also found that preschool students are more interested in robots that look like animals, such as bees.

Therefore, to ensure an effective learning process, future research direction should investigate the shape and appearance of machines when working with humans, depending on the educational scenario, including educational field, educational level, and educational context.

Collaborative intelligence in education

The notion of collaborative intelligence (CI) can be traced back several decades to early AI theorizing (Selfridge, 1959). Likewise, the related construct of 'collective intelligence' has got the same history with similar semantics (Suran et al., 2020). CI implies that AI and human intelligence complement, or augment, each other in completing a given educational task. Such a conception builds on complementary strengths; the leadership, teamwork, creativity, and social skills of humans, and the speed, scalability, and quantitative capabilities of machines, hence co-evolving together. To realize such collaborative intelligence, the natural capabilities of humans and machines in teaching scenarios require detailed analysis and appreciation of distinctive capabilities, such as humans having a sense of immediate context, humor, and responsibilities, while machines have computing power and physical abilities that humans can't achieve. Note that one relevant practice is Collaborative Intelligent Tutoring Systems (CITS), which are learning systems that integrate AI into collaborative learning environments (Ubani & Nielsen, 2022).

Ethics

With the rapid development and widespread use of intelligent technology, especially AI, educational technologies have moved from the backwaters of academic research to the forefront of the public. UNESCO (2019) indicated that the real AI age must be based on multi-level and all-field human-machine collaboration. It has been suggested that human-machine collaboration has great service to education (Kaarlea et al., 2022; Mizrahi et al., 2022; Tlili et al., 2023a). However, the integration of AI in teaching also raises some fundamental ethical concerns, for example, human dignity, discrimination, inequality, and data privacy issues (UNESCO, 2021; European Commission, 2019; Holmes et al., 2023). The ethical issues should be considered as a prerequisite concern for effective human-machine collaboration. Moreover, it is arguable that the evolution of AI as a branch of computer science has reached a point where ethical and social responsibility imperatives deem it is now multi-disciplinary.

Responsible human-machine collaboration

As the integration of machines into education continues to advance, responsible human-machine collaboration has become increasingly important. In this context, responsible collaboration refers to the alignment of human values with machine capabilities to ensure that machines operate within acceptable ethical boundaries. In this context, three key pillars of responsible HMC should be considered, namely: (1) human dignity, (2) data privacy, and (3) technical robustness and safety.

Human dignity: According to the European Commission (2019), machines should support human autonomy and decision-making and be prescribed by the principle of respect for human autonomy. Furthermore, their usage and application should address and support human rights, democracy, and the rule of law (Holmes et al., 2023). Machines should not be designed to degrade or demean human beings in any way. Instead, machines should act as enablers of a democratic, flourishing, and equitable society by supporting human users' fundamental rights. For example, assistive technologies should be designed to improve the quality of life for learners with disabilities rather than simply replacing them with machines (Alnahdi, 2014). Additionally, machines should not be programmed to discriminate against certain individuals or groups based on factors such as race, gender, or religion.

Data privacy and governance: Intelligent machines, especially AI, usually require a large amount of data, which involves a large amount of confidential information of students and teachers; hence, ethical and security issues will arise when collecting, using, and disseminating (Chen et al., 2021). Therefore, data privacy is a critical aspect of responsible human-machine collaboration. To protect users' privacy, machines must ensure credible data protection and governance systems. This includes implementing strong encryption and access controls, as well as providing individuals with control over their own data. For example, learners should be able to decide what data they want to share with machines and who has access to that data.

Technical robustness and safety: A crucial component of achieving responsible HMC is the technical robustness of machines (European Commission, 2019). It is essential to guarantee that machines operate as intended and do not malfunction or fail unexpectedly. Machines must be designed with appropriate fail-safes and error correction mechanisms to prevent harm to humans or the environment (ibid). Additionally, machines must be thoroughly tested and evaluated to ensure that they meet performance and safety standards.

Diversity, inclusion, and fairness

Diversity, inclusion, and fairness are critical aspects of intelligent human-machine synergy in collaborative teaching. The European Commission (2019) emphasizes that ensuring diversity and inclusion in AI development and use is necessary to avoid bias and discrimination. This is because AI systems can perpetuate and even amplify existing societal biases if they are trained on biased data or developed without considering the diverse needs of users. Therefore, in the intelligent human-computer synergy in collaborative teaching, diversity, inclusiveness and fairness in the development and use of artificial intelligence should be considered.

To achieve diversity, inclusion, and fairness in intelligent human-machine synergy in collaborative teaching, it is essential to integrate these concepts into all levels of education and professional development (UNESCO, 2021). This integration should include technical aspects of AI, such as how to design and test AI systems for fairness and how to mitigate bias in data, as well as ethical and social aspects, such as the impact of AI on marginalized communities and the need for inclusive design. One approach to promoting diversity, inclusion, and fairness is through using diverse and representative data sets.

Buolamwini and Gebru (2018) noted that AI systems should be trained on diverse and representative data sets to avoid bias and discrimination. Therefore, it is essential to educate individuals on the importance of using diverse and representative data sets in AI development. Another approach is interdisciplinary collaboration. The European Commission (2019) states that interdisciplinary approaches to AI education can help individuals develop a holistic understanding of AI technologies and their potential implications. Such approaches can involve experts from computer science, ethics, law, social science, and humanities, working together to design and develop AI systems that are fair and inclusive.

Trustworthy relationships between AI and humans

For humans to effectively collaborate with machines, they must trust that the machine will behave ethically, accurately, and transparently. These are the foundations of trustworthiness and a focus of international standardization in the field of AI (ISO/IEC, 2020). Therefore, AI systems must be designed to be transparent, explainable, and accountable (Felzmann, 2020). AI systems must be transparent in their decision-making processes to ensure that humans can understand the rationale behind their decisions. This transparency is particularly important when AI systems are used to make decisions that directly impact human lives, such as in healthcare or legal contexts (Floridi et al., 2019; Tlili et al., 2023d). Additionally, AI must respect human values and not undermine human dignity (Coeckelbergh, 2017). Therefore, AI must be designed to enhance human capabilities rather than replacing them.

Moreover, accountability is another important ethical consideration in the social relationship between AI and humans. AI systems must be accountable for their actions and decisions, and mechanisms must be in place to hold them responsible for any harm that they cause (Scherer, 2016).

There are potential risks associated with the social relationship between AI and humans, which must be addressed ethically. For example, AI systems can reinforce existing biases and discrimination if not purposely designed with ethical considerations in mind (Mittelstadt et al., 2016). Additionally, AI can be used to invade learners' personal privacy, monitor and manipulate behavior, and promote unethical practices (Zuboff, 2019). Therefore, ethical considerations must be addressed in the design and implementation of AI systems to avoid these risks.

Adequate AI literacy education

Adequate AI literacy education in intelligent human-machine synergy in collaborative teaching is a critical aspect of preparing individuals for the increasing integration of AI into various aspects of society. As noted by the European Commission (2019), AI literacy education involves developing a comprehensive set of knowledge, skills, and attitudes necessary for individuals to interact with intelligent machines effectively and ethically in a collaborative context. Such AI literacy includes understanding AI technologies, their capabilities and limitations, and the ethical and social issues related to their use. Besides, Tlili et al. (2023a), based on different human-machine collaboration scenarios, specifically with ChatGPT, revealed that not only are ICT competences now required but also general skills, such as critical thinking and question-asking competences to get the best results of the machine. This 'old' need for generic horizontal competences in our digital era is becoming more demanding due to the introduction of AI (Stracke, 2011, 2014).

The need for adequate AI literacy education in intelligent human-machine synergy in collaborative teaching cannot be overstated. UNESCO (2021) emphasizes that individuals need to have the necessary skills to work effectively with intelligent machines. Without this education, individuals may be reluctant to adopt new AI technologies or may misuse them, resulting in unintended consequences. Moreover, the European Commission (2019) emphasizes that adequate AI literacy education is crucial for ensuring the ethical and socially responsible development and use of AI technologies.

To achieve adequate AI literacy education in intelligent human-machine synergy in collaborative teaching, it is essential to integrate AI literacy education into all levels of education, as well as in professional development and lifelong learning, as pointed out by UNESCO (2021) and the European Commission (2019). This integration should include technical aspects of AI, as well as ethical and social aspects, such as bias and discrimination, privacy and security, and the impact of AI on employment and social inequality. Additionally, to ensure effective AI literacy education, innovative and engaging education methods are necessary, as suggested by the European Commission (2019) and UNESCO (2021). Project-based learning that involves the development of AI applications, for instance, can help individuals develop a deeper understanding of AI technologies and their potential uses. At the same time, the European Commission (2019) stated that interdisciplinary approaches to AI education, which bring together experts in computer science, ethics, law, social science, and humanities, are also crucial to the use and evolution of AI as well as society in general. Such approaches can help individuals develop a holistic understanding of AI technologies and their potential implications.

In conclusion, adequate AI literacy education in intelligent human-machine synergy in collaborative teaching is critical for individuals to interact with intelligent machines effectively and ethically. Integrating AI literacy education into all levels of education, innovative and engaging education methods, and interdisciplinary approaches to AI education are necessary for achieving this goal.

Learning, teaching and assessment

Role of teachers and learning scenarios

In human-machine collaboration, the machine could take various roles and tasks to effectively complete a given educational objective with the teacher. Kaber (2018), in this context, mentioned that one of the core questions in human-machine collaboration is "Who does what?". In a study investigating the effects of humans collaborating with machines, Nass et al. (1996, p. 669) revealed that the "effects of being in a team with a computer are the same as the effects of being in a team with another human". In the same vein, de Vreede and Briggs (2019, p. 103) stated that, in the future, "artificial agents will become fully functional members of teams"; therefore, there is an urgent need to investigate which roles automated agents can fulfill and perform. This implies that machines should not replace humans and take over every role within a team. Rather, they should be designed to fulfill certain activities which are most fitting and effective and, thus, complement or augment the advantages of humans (Dellermann et al. 2019). In this context, Tlili et al. (2023b) also pointed out after conducting a meta-analysis on the effects of Intelligent Tutoring Systems (ITSs) on learning achievement that machines should not replace humans in education. They should, however, complement them to effectively achieve given educational objectives.

To investigate the different roles that machines could take when collaborating with a human teacher, Bittner et al. (2019), for instance, developed a taxonomy focusing on team composition and the role of machines within teams. They recognized various roles, such as facilitators (e.g., instructors), peers (e.g., teammates), and experts (e.g., analyst or evaluator). They further called for more investigation of the different roles within the taxonomy. Therefore, it is important to investigate how machines can work with human teachers to effectively achieve a given educational objective. Effective collaboration, in which the function of machines is no longer merely that of a tool but rather a team member, can only result from a perception of equal roles (Nass et al., 1996).

Assessment of learning and human-machine synergy

Assessment of learning and human-machine synergy, as well as human-machine collaborative teaching, is crucial for improving teaching and learning. The appropriate assessment results can drive students' learning and promote teachers' professional development. Assessment results can provide a clear picture of goal attainment. In addition, assessment results can shed light on how to improve teaching and learning to illuminate how to align instructional design and enactment.

To examine the effectiveness of human-machine synergy, many methods could be employed in practice. For example, automatic speech evaluation can be employed to investigate the effectiveness of human-machine synergy. In addition, a collaborative human-machine evaluation framework and tools can be developed to examine the effectiveness of human-machine synergy. Furthermore, whether learners have achieved personalized learning or not can be used to evaluate the effectiveness of human-machine synergy since human-machine synergy can empower personalized learning.

To examine the effectiveness of human-machine collaborative teaching, many methods can be employed in practice. For example, it is possible to use explicit methods, such as tests (pre- and post-tests). It is also possible to use implicit methods, for instance, by analyzing students' learning behaviors within the teaching practice (i.e., a human teacher and a machine teaching together), to draw conclusions accordingly. In addition, teachers or researchers can investigate learners' perceptions through questionnaires or semi-structured interviews to get an understanding of the effectiveness of human-machine collaborative teaching.

Quasi-experimental methods can be adopted to examine the effectiveness of collaborative teaching between the human teacher and the machine compared to the teaching practice without the machine (i.e., in the absence of human-machine collaboration). Furthermore, learners' learning engagement, cognitive and metacognitive skills, emotions, motivations, and behaviors can also be examined to measure the effectiveness of human-machine collaborative teaching. For example, Han et al. (2023) proposed a technology-enhanced Edu-Metaverse framework to promote learner engagement with human-machine interactions. Finally, existing standards and frameworks that are already used for technology-enhanced learning designs and their impact assessment can be applied such as the international ISO standard for digital learning ISO/IEC 40180 (2017, revision of the original standard ISO/IEC 19796-1 (2005)) and the Quality Reference Framework (QRF) for online learning developed and evaluated by more than 10,000 learners, designers and facilitators (Stracke et al., 2018).

Generally, it is most important to assess and evaluate the impact of human-machine synergy and human-machine collaborative teaching on all three educational levels: the micro, meso and macro level (Stracke, 2019). Table 2 provides an overview of the key leading questions and perspectives that must be addressed and differentiated for a complete impact assessment at all educational levels.

At the micro level, students and teachers make their own choices and the best ways to learn respectively to educate when involving and using machines. At the meso level, teams and organizations responsible for designing and providing courses and education need to reflect on how they can effectively use machines within their syllabus and learning opportunities. This also raises the question if future curricula should be revisited and redesigned to meet the new needs of this teaching practice (i.e., human-machine collaborative teaching). At the macro level, policy developers and politicians must think critically and decide how machines can be safely implemented in educational systems and curricula to achieve a positive societal impact for the commons. It is important that policymakers make clear guidelines and regulations to safely adopt certain technologies in education (e.g., the use of ChatGPT in schools and universities rather than simply banning it). Within all three levels, the learning processes (i.e., facilitated through human-machine synergy), as well as the learning designs (i.e., facilitated through human-machine collaborative teaching), should include objectives, realizations, and achievements in their impact assessment for a holistic evaluation (Stracke, 2017).

Table 2. Impact assessment on educational levels.

	Human-machine synergy in education	**Human-machine collaborative teaching**
Macro level	How can educational systems and institutions safely adopt and exploit machines?	How can society take advantage of machines for the common good?
Meso level	How can courses and curricula benefit from machines?	How can learning design integrate machines for effective and human-centered teaching practices?
Micro level	How can teachers and students interact or learn with machines?	How can teachers team-up with machines to achieve an educational objective?

All these relevant perspectives and examples demonstrate the potential applications and benefits of human-machine collaboration as well as the need for careful design. It also makes salient to the learning and teaching scenarios and processes involving AI and human-machine interactions. The relevant research has just started to reveal conditions and effects of effective and successful introduction of AI and human-machine collaboration in education.

Limitations

This study has a couple of limitations that should be acknowledged. For instance, this study is descriptive in nature and all the reported findings are mainly based on a review of the literature. Stated another way, no experimental studies were conducted to validate the components of the framework. In addition, as generative AI technologies evolve during the coming decade and other educational technologies emerge, the iSTAR framework will likely need further refinement and validation.

Conclusions and future directions

This study has discussed human-machine collaboration in education, and presented iSTAR as a reference framework. iSTAR presents a simple visual representation of the different types of entities that a human teacher can collaborate with to achieve a given educational objective, as well as the different levels of collaboration. It also highlights the different dimensions that should be considered for an effective and safe human-machine collaboration in education.

The study can contribute to the literature from different perspectives. From a theoretical perspective, this study contributes to the ongoing debate and progress of human-machine collaboration in the field of education, especially with the rapid development of AI technologies. From a practical perspective, this manuscript highlights different dimensions that various stakeholders (e.g., designers, developers, educators, policymakers, etc.) should pay attention to for a safe and effective learning experience; hence, helping ensure enhanced learning outcomes.

This study further suggests the significance of constructing an ethical framework to govern the domain of human-machine collaboration in the educational context. Notably, it suggests that future policies should encompass privacy protection, algorithmic transparency, accountability and human-machine teaming up together. In addition, fostering transparency, explainability, and user control assumes paramount importance in establishing trust and enabling fruitful collaborations. Therefore, future research could also focus on this line of research.

References

Adarkwah, M. A., Amponsah, S., van Wyk, M. M., Huang, R., Tlili, A., Shehata, B., Metwally, A. H. S., & Wang, H. (2023). Awareness and acceptance of ChatGPT as a generative conversational AI for transforming education by Ghanaian academics: A two-phase study. *Journal of Applied Learning and Teaching, 6*(2), 1-16. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.26

Al Hakim, V. G. A., Yang, S., Liyanawatta, M., Wang, J., & Chen, G. (2022). Robots in situated learning classrooms with immediate feedback mechanisms to improve students' learning performance. *Computers & Education, 182*, 104483. https://doi.org/10.1016/j.compedu.2022.104483

Alnahdi, G. H. (2014). Assistive technology in special education and the universal Design for learning. *Turkish Online Journal of Educational Technology, 13*(2), 18–23. http://files.eric.ed.gov/fulltext/EJ1022880.pdf

Asimov, I. (1942). Runaround. *Astounding Science Fiction, 29*(1), 94–103.

Asimov, I. (1950). *I, robot.* Roc.

Bailenson, J. N., & Blascovich, J. (2004). Avatars. *Encyclopedia of Human-Computer Interaction,* 64–68.

Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development, 59*(2), 291–300. https://doi.org/10.1007/s11423-011-9196-3

Blake, A. M., & Moseley, J. L. (2010). The emerging technology of avatars: Some educational considerations. *Educational Technology Archive, 50*(2), 13–20. https://dialnet.unirioja.es/servlet/articulo?codigo=4637640

Bolton, C., Machová, V., Kováčová, M., & Valášková, K. (2018). The power of human–machine collaboration: Artificial intelligence, business automation, and the smart economy. *Economics, Management, and Financial Markets, 13*(4), 51. https://doi.org/10.22381/emfm13420184

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., ... & Jandrić, P. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education, 18*(1), 53-130. https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/709

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Card, S. K., Newell, A., & Moran, T. P. (2018). The psychology of human-computer interaction. In *CRC Press eBooks.* https://doi.org/10.1201/9780203736166

Capek, K. (2004). *R.U.R. (Rossum's Universal Robots).* Penguin.

Chen, L., Huang, Y., Ouyang, S., & Xiong, W. (2021). *The data privacy paradox and digital demand (No. W28854).* National Bureau of Economic Research.

Coeckelbergh, M. (2016). Can machines create art? *Philosophy & Technology, 30*(3), 285–303. https://doi.org/10.1007/s13347-016-0231-5

Cruickshank, L., & Trivedi, N. (2017). Beyond human-centred design: Supporting a new materiality in the internet of things, or how to design when a toaster is one of your users. *Design Journal, 20*(5), 561-576. https://doi.org/10.1080/14606925.2017.1349381

Dart, K., Toohey, L., Moore, M., & Toohey, D. (2019). Meeting the access to civil justice challenge: Digital inclusion, algorithmic justice, and human-centred design. *Macquarie Law Journal, 19*, 133-156.

De Martino, J. M., Silva, I. R., Bolognini, C. Z., Costa, P. D. P., Kumada, K. M. O., Coradine, L. C., Brito, P. H. S., Amaral, W. M. D., Benetti, Â., Poeta, E. T., Angare, L. M., Ferreira, C. M., & De Conti, D. F. (2016). Signing avatars: Making education more inclusive. *Universal Access in the Information Society, 16*(3), 793–808. https://doi.org/10.1007/s10209-016-0504-x

Dear, B. (2017). *The friendly orange glow: The untold story of the rise of cyberculture.* Vintage.

Dede, C. Etemadi, A., & Forshaw, T. (2021). *Intelligence augmentation: Upskilling humans to complement AI.* The Next Level Lab at the Harvard Graduate School of Education. President and Fellows of Harvard College: Cambridge, MA.

Dede, C., Grotzer, T. A., Kamarainen, A., & Metcalf, S. (2017). EcoXPT: Designing for deeper learning through experimentation in an immersive virtual ecosystem. *Educational Technology & Society, 20*(4), 166-178.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, (pp. 1-7).

Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason.* MIT Press.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63-71.

European Commission. (2019). *European commission independent high-level expert group on artificial intelligence. Ethics guidelines for trustworthy AI.* https://ec.europa.eu/futurium/en/ai-alliance-consultation

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics, 26*(6), 3333-3361.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Rossi, F. (2018). *AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations, Atomium.* European Institute for Science, Media and Democracy Brussels, Belgium.

Gerber, A., Derckx, P., Döppner, D. A., & Schoder, D. (2020). Conceptualization of the human-machine symbiosis - a literature review. *Hawaii International Conference on System Sciences (HICSS),* Hawaii, USA.

Giacomin, J. (2014). What is human centred design?. *The Design Journal, 17*(4), 606-623.

Guo, X., Du, J., Pu, Y., Liu, Q., Wang, Y., & Li, J. (2022). 40-year development of man-machine-environment system engineering from scientific papers. In *Man-machine-environment system engineering: Proceedings of the 21st international conference on mmese: Commemorative conference for the 110th anniversary of xuesen qian's birth and the 40th anniversary of founding of man-machine-environment system engineering 21* (pp. 21-28). Springer Singapore.

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior, 119*, 106730.

Han, Z., Tu, Y., & Huang, C. (2023). A framework for Constructing a technology-enhanced education metaverse: Learner engagement with human–machine collaboration. *IEEE Transactions on Learning Technologies.*

Hoc, J. M. (2000). From human–machine interaction to human–machine cooperation. *Ergonomics, 43*(7), 833-843. https://doi.org/10.1080/001401300409044

Holmes, W., Stracke, C. M., Chounta, I.-A., Allen, D., Baten, D., Dimitrova, V., Havinga, B., Norrmen-Smith, J., & Wasson, B. (2023). AI and Education. A view through the lens of human rights, democracy and the rule of law. Legal and organizational requirements. In: N. Wang., G. Rebolledo-Mendez., V, Dimitrova., N, Matsuda., O. C. Santos (Eds.), *Artificial intelligence in education* (pp. 79-84). https://doi.org/10.1007/978-3-031-36336-8_12

ISO. (2020). *ISO/IEC TR 24028:2020. Information technology — artificial intelligence — overview of trustworthiness in artificial intelligence.* International Organization for Standardization (ISO).

ISO. (2019). *ISO 9241-210: 2019. International Organization for Standardization (ISO).* https://www.iso.org/standard/77520.html

ISO. (2017). *ISO/IEC 40180:2017. Information technology — quality for learning, education, and training — fundamentals and reference framework.* International Organization for Standardization (ISO).

Jabeen, S., Li, X., Amin, M. S., Bourahla, O. E., Li, S., & Jabbar, A. (2022). A review on methods and applications in multimodal deep learning. *ACM Transactions on Multimedia Computing, Communications and Applications, 19,* 1-41.

Jones, S. (2015, November 23). *PLATO*. Encyclopedia Britannica. https://www.britannica.com/topic/PLATO-education-system

Kaber, D. B. (2018). Issues in human–automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation. *Journal of Cognitive Engineering and Decision Making, 12*(1), 7-24.

Kaarlela, T., Arnarson, H., Pitkäaho, T., Shu, B., Solvang, B., & Pieskä, S. (2022). Common educational teleoperation platform for robotics utilizing digital twins. *Machines, 10,* 577.

Keren, G., & Fridin, M. (2014). Kindergarten Social Assistive Robot (KindSAR) for children's geometric thinking and metacognitive development in preschool education: A pilot study. *Computers in Human Behavior, 35*, 400-412.

Koedinger, K. R., & Aleven, V. (2016). An interview reflection on "intelligent tutoring goes to school in the big city". *International Journal of Artificial Intelligence in Education, 26*, 13-24.

Lee, J. K., Toscano, R. L., Stiehl, W. D., & Breazeal, C. (2008, August). The design of a semi-autonomous robot avatar for family communication and education. In *RO-MAN 2008-The 17th IEEE International symposium on robot and human interactive communication* (pp. 166-173). IEEE.

Liljaniemi, A. & Paavilainen, H. (2020). Using digital twin technology in engineering education – course concept to explore benefits and barriers. *Open Engineering, 10*(1), 377-385. https://doi.org/10.1515/eng-2020-0040

Longo, F., Nicoletti, L., & Padovano, A. (2017). Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & industrial engineering, 113*, 144-159. https://doi.org/10.1016/j.cie.2017.09.016

Lugrin, J. L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., & Grafe, S. (2016). Breaking bad behaviors: A new tool for learning classroom management using virtual reality. *Frontiers in ICT, 3*, 26.

Long, S., & Huang, Y. (2021). Man-machine-environment system engineering and its historical mission. In *Springer eBooks* (pp. 3–20). https://doi.org/10.1007/978-981-16-5963-8_1

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (2006). A proposal for Dartmouth summer research project on artificial intelligence. *AI Magazine, 27,* 1-12.

Mason, J. (2023). The search continues: Content used to be king, but context is now everything–and 'smart' can morph into 'stupid'. *Sociological Review Magazine.* http://dx.doi.org/10.51428/tsr.ngtq5087

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 2053951716679679.

Mizrahi, E., Danzig, N., & Gordon, G. (2022). vRobotator: A virtual robot facilitator of small group discussions for K-12. *Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2)*, 1-22.

Muin, S., & Mosalam, K. M. (2021). Human-machine collaboration framework for structural health monitoring and resiliency. *Engineering Structures, 235*, 112084.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates?. *International Journal of Human-Computer Studies, 45*(6), 669-678.

O'Connor, M., Deeks, H. M., Dawn, E., Metatla, O., Roudaut, A., Sutton, M., Thomas, L. M., Glowacki, B. R., Sage, R., & Tew, P. (2018). Sampling molecular conformations and dynamics in a multiuser virtual reality framework. *Science Advances, 4*(6), 2731.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 30*(3), 286-297.

Pairet, È., Ardón, P., Liu, X., Lopes, J., Hastie, H., & Lohan, K. S. (2019, March). A digital twin for human-robot interaction. In *2019 14th ACM/IEEE International conference on Human-Robot Interaction (HRI)* (pp. 372-372). IEEE

Qiu, L., & Benbasa, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems, 25*(4), 145-182.

Resnick, M., Ocko, S., & Papert, S. (1988). LEGO, logo, and design. *Children's Environments Quarterly,* 14-18.

Ringwald, M., Theben, P., Gerlinger, K., Hedrich, A., & Klein, B. (2023). How should your assistive robot look like? A scoping review on embodiment for assistive robots. *Journal of Intelligent & Robotic Systems, 107*(1), 12. https://doi.org/10.1007/s10846-022-01781-3

Ropohl, G. (1999). Philosophy of socio-technical systems. *Society for Philosophy and Technology Quarterly Electronic Journal, 4*(3), 186-194.

Sampson, D., Papamitsiou, Z., Ifenthaler, D., Giannakos, M., Mougiakou, S., & Vinatsella, D. (2022). *Educational data literacy.* Springer Nature.

Saurin, T. A., & Patriarca, R. (2020). A taxonomy of interactions in socio-technical systems: A functional perspective. *Applied Ergonomics, 82*, 102980. https://doi.org/10.1016/j.apergo.2019.102980

Schmidtler, J., Knott, V., Hölzel, C., & Bengler, K. (2015). Human centered assistance applications for the working environment of the future. *Occupational Ergonomics, 12*(3), 83-95.

Sepasgozar, S. M. (2020). Digital twin and web-based virtual gaming technologies for online education: A case of construction management and engineering. *Applied Sciences, 10*(13), 4678.

Scherer, R. (2016). Learning from the past–the need for empirical evidence on the transfer effects of computer programming skills. *Frontiers in Psychology, 7*, 1390.

Schild, J., Lerner, D., Misztal, S., & Luiz, T. (2018, May). EPICSAVE—Enhancing vocational training for paramedics with multi-user virtual reality. In *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)* (pp. 1-8). IEEE.

Segaran, K., Mohamad Ali, A. Z., & Hoe, T. W. (2021). Does avatar design in educational games promote a positive emotional experience among learners?. *E-Learning and Digital Media, 18*(5), 422-440.

Selwyn, N. (2023). Constructive criticism? Working with (rather than against) the AIED back-lash. *International Journal of Artificial Intelligence in Education,* 1-8. https://doi.org/10.1007/s40593-023-00344-3

Sharples, M. (2022). Automated essay writing: An AIED opinion. *International Journal of Artificial Intelligence in Education., 32*(4), 1119-1126. https://doi.org/10.1007/s40593-022-00300-7

Simmler, M., & Frischknecht, R. (2021). A taxonomy of human–machine collaboration: Capturing automation and technical autonomy. *AI & Society, 36*(1), 239-250. https://doi.org/10.1007/s00146-020-01004-z

Stracke, C. M. (2011). Competences and skills in the digital age: Competence development, modelling, and standards for human resources development. In: *Communications in computer and information science* (pp. 34-46). https://doi.org/10.1007/978-3-642-24731-6_4

Stracke, C. M. (2014). How innovations and competence development support quality in lifelong learning. *The International Journal for Innovation and Quality in Learning, 2*(3), 35-44. https://doi.org/10.5281/zenodo.3608669

Stracke, C. M. (2017). The quality of MOOCs: How to improve the design of open education and online courses for learners? In P. Zaphiris, & A. Ioannou (Eds.), *Learning and collaboration technologies. Novel learning ecosystems.* (pp. 285-293). https://doi.org/10.1007/978-3-319-58509-3_23

Stracke, C. M. (2019). Quality frameworks and learning design for open education. *The International Review of Research in Open and Distributed Learning, 20*(2), 180-203. https://doi.org/10.19173/irrodl.v20i2.4213

Stracke, C. M., & Skuballa, I. (2021). Using, analysing and evaluating eye-tracking at different educational levels: Towards an evaluation framework for impact assessment. *TEM JOURNAL - Technology, Education, Management, Informatics, 10*(2), 483-487. https://doi.org/10.18421/TEM102-01

Stracke, C. M., Tan, E., Texeira, A., Pinto, M., Vassiliadis, B., Kameas, A., Sgouropoulou, C., & Vidal, G. (2018). *Quality Reference Framework (QRF) for the Quality of Massive Open Online Courses (MOOCs).* https://doi.org/10.5281/zenodo.3894573

Summaira, J., Li, X., Shoib, A.M., Li, S., & Jabbar, A. (2021). *Recent advances and trends in multimodal deep learning: A review.* ArXiv, abs/2105.11087.

Semeraro, F., Griffiths, A., & Cangelosi, A. (2023). Human–robot collaboration and machine learning: A systematic review of recent research. *Robotics and Computer-Integrated Manufacturing, 79*, 102432.

Shahab, M., Taheri, A., Mokhtari, M., Shariati, A., Heidari, R., Meghdari, A., & Alemi, M. (2022). Utilizing social virtual reality robot (V2R) for music education to children with high-functioning autism. *Education and Information Technologies,* 1-25.

Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2022). Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence, 5*, 46 - 57.

Suran, S., Pattanaik, V., & Draheim, D. (2020). Frameworks for collective intelligence: A systematic literature review. *ACM Computing Surveys, 53*(1), 1-14. https://doi.org/10.1145/3368986

Tlili, A., Huang, R., & Kinshuk, X. (2023c). Metaverse for climbing the ladder toward 'industry 5.0'and 'society 5.0'?. *The Service Industries Journal, 43*(3-4), 260-287. https://doi.org/10.1080/02642069.2023.2178644

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023a). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Tlili, A., Salha, S., Wang, H., & Huang, R. (2023b). Intelligent tutoring systems examined in social experiments—is the magic gone? A meta-analysis. In *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)* (pp. 50-54).

Tlili, A., Huang, R., Mustafa, M. Y., Zhao, J., Bozkurt, A., Xu, L., ... & Burgos, D. (2023d) Speaking of transparency: Are all Artificial Intelligence (AI) literature reviews in education transparent?. *Journal of Applied Learning and Teaching, 6*(2), 1-12. https://doi.org/10.37074/jalt.2023.6.2.15

Tricco, A. C., Antony, J., Zarin, W., Strifler, L., Ghassemi, M., Ivory, J., ... & Straus, S. E. (2015). A scoping review of rapid review methods. *BMC Medicine, 13*(1), 1-15.

Turing, A. (1950). Computing machinery and intelligence. *Mind A Quarterly Review of Psychology and Philosophy, 236,* 433–460. https://doi.org/10.1093/mind/lix.236.433

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence.* Ethics of Artificial Intelligence | UNESCOunesco.orghttps://www.unesco.org › artificial-intelligence

UNESCO. (2019). *Beijing consensus on artificial intelligence and education.* Beijing Consensus on Artificial Intelligence and Educationunesco.orghttps://unesdoc.unesco.org › ark:

Vanderhaegen, F., Crevits, I., Debernard, S., & Millot, P. (1994). Human-machine cooperation: Toward an activity regulation assistance for different air traffic control levels. *International Journal of Human-Computer Interaction, 6*(1), 65-104. https://www.tandfonline.com/doi/abs/10.1080/10447319409526084

Vergara, D., Extremera, J., Rubio, M. P., & Dávila, L. P. (2019). Meaningful learning through virtual reality learning environments: A case study in materials engineering. *Applied Sciences, 9*(21), 4625.

Vuorikari, R., Punie, Y., & Cabrera Giraldez, M. (2020). *Emerging technologies and the teaching profession: Ethical and pedagogical considerations based on near-future scenarios.*

Watson, D. (2019). The rhetoric and reality of anthropomorphism in artificial intelligence. *Minds and Machines, 29*(3), 417-440.

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review, 96*(4), 114-123. https://hbr.org/2018/07/

collaborative-intelligence-humans-and-ai-are-joining-forces

Yun, S. S., Kim, M., & Choi, M. T. (2013). Easy interface and control of tele-education robots. *International Journal of Social Robotics, 5*, 335-343

Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. In *New Labor Forum, 28(*1), 10-29.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Speaking of transparency: Are all Artificial Intelligence (AI) literature reviews in education transparent?

Ahmed Tlili[A] — [A] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Ronghuai Huang[B] — [B] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Muhammad Yasir Mustafa[C] — [C] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Jialu Zhao[D] — [D] *Stanford University, California, USA*

Aras Bozkurt[E] — [E] *Distance Education Department, Open Education Faculty, Anadolu University, Eskişehir, Turkey*

Lin Xu[F] — [F] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Huanhuan Wang[G] — [G] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Soheil Salha[H] — [H] *Faculty of Educational Sciences and Teachers' Training, An-Najah National University, Nablus, Palestine*

Fahriye Altinay[I] — [I] *Near East University, Institute of Graduate Studies, Faculty of Education, Societal Research and Development Center, Nicosia, North Cyprus, Mersin, Turkey*

Saida Affouneh[J] — [J] *Faculty of Educational Sciences and Teachers' Training, An-Najah National University, Nablus, Palestine*

Daniel Burgos[K] — [K] *Research Institute for Innovation & Technology in Education (UNIR iTED), Universidad Internacional de La Rioja (UNIR). Logroño, La Rioja, Spain*

Keywords

Artificial intelligence;
education;
literature review;
transparency assessment.

Correspondence

ahmed.tlili23@yahoo.com [A]

Article Info

Received 17 July 2023
Received in revised form 26 July 2023
Accepted 26 July 2023
Available online 31 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.15

Abstract

Literature reviews are considered a core research approach in developing new theories and identifying trends and gaps in a given research topic. However, the transparency level of literature reviews might hinder the quality of the obtained findings, thus limiting their implications. As transparency is one of the core elements when implementing Artificial Intelligence (AI), this study assesses the transparency level of literature reviews on AI in education. Specifically, this study used a systematic review to collect and analyze information about reports of methodological decisions and research activities in 61 literature review papers. The obtained findings highlighted that 51.9% of the conducted reviews on AI in education are descriptive. Additionally, the transparency level of the conducted literature reviews was low; 40% of the reviews were in Q1 and 32% in Q2. Particularly, the quality assessment step had the lowest transparency level. The findings of this research can advance the educational technology field by underscoring the methodological gaps when conducting a literature review on AI in education and hence enhance the transparency and trustworthiness of the obtained findings.

Introduction

Since 2000, the term 'Technology-enhanced Learning' (TEL) has appeared more frequently in the educational landscape (Al-Ataby, 2020). A complex and intertwined relationship arose between education and technology, and the use of technology in education evoked pedagogical, social, political, and economic effects (Guilherme, 2017) In other words, the technology-intensive 21st century carries various educational implications. Specifically, the use of technology in learning and teaching can effectively facilitate the accomplishment of teaching tasks, improve learning outcomes, and increase classroom interaction and communication. The last 30 years, the advent of Artificial Intelligence (AI) technology penetrated the educational domain as well (Tahiru, 2021), and Artificial Intelligence in Education (AIED) emerged as a developing research field. In AIED, a machine is designed to mimic a human system to support human learning and teaching (Tahiru, 2021; Conati et al., 2018). Therefore, AIED has cognitive, adaptive, decision-making, problem-solving, modelling, and other capabilities to help perform different educational tasks more effectively, including reviewing and grading students' assignments, providing flexible and personalized learning experiences, and implicitly modelling students' profiles (Chen et al., 2020a; Essalmi et al., 2017; Pan et al., 2021; Tlili et al., 2022a).

To provide comprehensive insights into AI's use in education, several literature reviews have been conducted (e.g., Ouyang et al., 2022; Wang et al., 2023). Such studies are intended to provide a holistic perspective by analyzing the approaches and synthesizing the research findings across scholarly papers. Rowe (2014) stated that review papers can be grouped into four categories based on the type of contribution to theory, namely, describing, understanding, theory testing, and explaining a phenomenon (see Table 1).

Regardless of the category of a given review paper, a transparent, as well as a systematic process, among other factors, contributes to the formation of high-quality review papers and the production of new perspectives on the research field. Therefore, transparency is described as a meticulous and thorough reporting of methodological choices made during the review process (Templier & Pare, 2018). Explicit disclosure may strengthen the work and its conclusions' credibility. Further, it also helps to ensure the internal and external reliability of the review processes. Transparency creates methodological rigour and repeatability of studies. Its importance in scientific research is increasingly emphasized across disciplines in the social and natural sciences (Tuval-Mashiach, 2017; McIntosh et al., 2017). Paré et al. (2016) further highlighted two limitations of non-transparent literature reviews: (1) lack of clarity when discussing the methodology of the study and (2) structural constraints of the publishing environment on producing extensive information regarding systematicity in the process of research.

Motivated by this background, and since transparency has been also one of the key dimensions that should be considered when implementing AI in general or in education particularly (Tlili et al., 2021; Larsson & Heintz, 2020), this study answers the following research question: *What is the transparency level of the conducted literature reviews on AI in education?* Specifically, this study conducts a systematic review to identify literature reviews on AIED in the literature, and then assess their transparency level following the Rowe (2014) and Pare et al. (2016) classification and transparency assessment metrics in review articles. In other words, this present study analyzes how transparent the authors from the literature were when adopting a given methodology for their literature reviews on AIED. The findings of this study can contribute to the AIED field by highlighting the transparency gaps of the conducted AIED literature reviews, hence consider them in the future when conducting a literature review. This can ensure more reliable, reusable and trustworthy findings on AIED that can advance the field. Despite the importance of the topic, no previous research, to the best of our knowledge, has conducted a similar analysis.

Table 1. The classification of review papers and their contributions to the theory.

Overarching goal (Adopted from Rowe, 2014)	Types of literature reviews commonly accepted goals, and frequently researched questions (Adopted from Paré et al., 2016)
Describing	*Narrative review:* A narrative review summarizes what has already been reported/published on a specific topic and what do we know or have discovered about this topic? *Descriptive review:* A descriptive review analyses the trends or patterns in available theories, hypotheses, techniques, or conclusions.
Understanding	*Scoping review:* A scoping review provides a broad overview of scientific knowledge on a particular topic. This form of review also permits identifying the gaps in the literature as well as future potential research directions. *Critical review:* A critical review is a critical evaluation that identifies flaws, contradictions, disagreements, or inconsistencies on a certain topic.
Testing theory	*Meta-analysis:* A meta-analysis is a method of combining statistical data from main quantitative research to provide relevant findings on a given issue. *Qualitative systematic:* A qualitative systematic literature review compiles statistical/empirical data on a certain topic and displays it in a narrative style. *Umbrella review:* The umbrella review examines and combines the qualitative systematic reviews and meta-analyses to produce the highest level of proof.
Explaining	*Theoretical review:* A theoretical review helps to develop new logical structure and framework by extending current ideas. The study objectives are stated, although there are usually no formal research questions. *Realist review:* A realist review determin[es] what works for whom, under what conditions, in what areas, and also how.

Method

This study assesses the transparency of the conducted literature reviews on AIED. To identify these literature reviews, the recommended reporting items for systematic reviews and meta-analyses (PRISMA) criteria were followed (Page et al., 2021). A PRISMA technique is one of the standardized peer-reviewed methodologies that employ a guideline checklist to ensure the quality and reliability of the revision process.

Search strategy and selection criteria

An extensive search was undertaken in the following databases, namely: Web of Science, Scopus, Taylor & Francis, and Science Direct, IEEE Xplore, as they are very popular in the field of educational technology (Tlili et al., 2022b; Wang et al., 2023). Particularly, the following search string was used: (AI OR Artificial Intelligence OR machine learning OR deep learning or natural language processing) AND (education OR learning) AND (literature review OR systematic review OR meta-analysis or state-of-art). The search query was applied to titles and abstracts, and the search keywords were partially adopted from Zawacki-Richter et al. (2020). After searching the appropriate databases, two authors individually analyzed the extracted papers by titles, abstracts, and textual on the inclusion and exclusion criteria reported in Table 2. During this phase, to reach a final consensus, disagreements between the authors were resolved through discussion or arbitration by a third author who has experience in AI research.

Table 2. Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
Journal article	Conference proceedings, dissertations, novels, book series, and book chapters
Literature review	Not a literature review article or an article that does not explain how the literature review was conducted.
Papers focusing on AI in education	Papers not focusing on AI or discussing AI, not in education
Accessible online Papers in English	Not accessible online Papers not in English

The search yielded a total of 1,367 articles, where 1,330 articles remained after removing duplicates. The screening of titles and abstracts resulted in the removal of 1,009 articles. The remaining 321 papers were considered and assessed as a full text. 260 of these articles failed to meet the criteria for inclusion. As a result, 61 research articles were suitable to be included in this study (see Table 3).

Table 3. 61 included studies in this systematic literature review.

No.	Author	Title
1	Hannan & Liu, 2023	AI: new source of competitiveness in higher education
2	Su et al., 2023	Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities
3	Banihashem et al., 2022	A systematic review of the role of learning analytics in enhancing feedback practices in higher education
4	Bilgic et al., 2022	Exploring the roles of artificial intelligence in surgical education: A scoping review
5	Chu et al., 2022	Artificial intelligence-based robots in education: A systematic review of selected SSCI publications
6	Dai & Ke, 2022	Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review
7	Kirubarajan et al., 2022	Artificial Intelligence and Surgical Education: A Systematic Scoping Review of Interventions
8	Laupichler et al., 2022	Artificial intelligence literacy in higher and adult education: A scoping literature review
9	Maier & Klotz, 2022	Personalized feedback in digital learning environments: Classification framework and literature review
10	Murtazaet al., 2022	AI-Based Personalized E-Learning Systems: Issues, Challenges, and Solutions
11	Su & Yang, 2022	Artificial intelligence in early childhood education: A scoping review
12	Su et al., 2022	A meta-review of literature on educational approaches for teaching AI at the K-12 levels in the Asia-Pacific region
13	Tan et al., 2022	A systematic review of artificial intelligence techniques for collaborative learning over the past two decades
14	Xia et al., 2022	Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education
15	Xu & Ouyang, 2022	A systematic review of AI role in the educational system based on a proposed conceptual framework
16	Zafari et al., 2022	Artificial Intelligence Applications in K-12 Education: A Systematic Literature Review
17	Ahmad et al., 2021	Artificial intelligence and its role in education
18	Alamri & Alharbi, 2021	Explainable Student Performance Prediction Models: A Systematic Review
19	Aslam et al., 2021	Feature Evaluation of Emerging E-Learning Systems Using Machine Learning: An Extensive Survey
20	Balaji et al., 2021	Contributions of machine learning models towards student academic performance prediction: A systematic review
21	Bozkurt et al., 2021	Artificial intelligence and reflections from educational landscape: A review of AI studies in half a century
22	González-Calatayud et al., 2021	Artificial intelligence for student assessment: A systematic review
23	Hahn et al., 2021	A Systematic Review of the Effects of Automatic Scoring and Automatic Feedback in Educational Settings
24	Huang et al., 2021	A Review on Artificial Intelligence in Education
25	Hwang & Chang, 2021	A review of opportunities and challenges of chatbots in education
26	Hwang & Tu, 2021	Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review
27	Kabudi et al., 2021	AI-enabled adaptive learning systems: A systematic mapping of the literature
28	Kahn & Winters, 2021	Constructionism and AI: A history and possible futures
29	Kharbat et al., 2021	Identifying gaps in using artificial intelligence to support students with intellectual disabilities from education and health perspectives
30	Klyuchnikov et al., 2021	Some Aspects of Ai-Technologies in Education
31	Lee et al., 2021	Artificial Intelligence in Undergraduate Medical Education: A Scoping Review
32	Liang et al., 2021	Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach
33	Liu & Afzaal, 2021	Artificial Intelligence (AI) and Translation Teaching: A Critical Perspective on the Transformation of Education
34	Luan & Tsai, 2021	A Review of Using Machine Learning Approaches for Precision Education
35	Maghsudi et al., 2021	Personalized Education in the Artificial Intelligence Era: What to Expect Next
36	Ng et al., 2021	Conceptualizing AI literacy: An exploratory review
37	Nigam et al., 2021	A Systematic Review on AI-based Proctoring Systems: Past, Present and Future
38	Sekeroglu et al., 2021	Systematic literature review on machine learning and student performance prediction: Critical gaps and possible remedies
39	Sousa et al., 2021	The Potential of AI in Health Higher Education to Increase the Students' Learning Outcomes
40	Tahiru, 2021	AI in education: A systematic literature review
41	Tang et al., 2021	Trends in artificial intelligence-supported e-learning: a systematic review and co-citation network analysis (1998–2019)
42	Vázquez-Cano, 2021	Artificial intelligence and education: A pedagogical challenge for the 21st century
43	Zhai et al., 2021a	A Review of Artificial Intelligence (AI) in Education from 2010 to 2020

44	Zhai et al., 2021b	A Meta-Analysis of Machine Learning-Based Science Assessments: Factors Impacting Machine-Human Score Agreements
45	Zhang & Aslan, 2021	AI technologies for education: Recent research & future directions
46	Alfaro et al., 2020	A Review of Intelligent Tutorial Systems in Computer and Web based Education
47	Chen et al., 2020a	Artificial Intelligence in Education: A Review
48	Chen et al., 2020b	Application and theory gaps during the rise of Artificial Intelligence in Education
49	Du et al., 2020	Educational data mining: a systematic review of research and emerging trends
50	Guan et al., 2020	Artificial intelligence innovation in education: A twenty-year data-driven historical analysis
51	Sapci & Sapci, 2020	Artificial intelligence education and tools for medical and health informatics students: Systematic review
52	Voskoglou & Salem, 2020	Benefits and limitations of the artificial with respect to the traditional learning of mathematics
53	Yufeia et al., 2020	Review of the application of artificial intelligence in education
54	Zhai et al., 2020	Applying machine learning in science assessment: a systematic review
55	Ferro et al., 2019	Innovative trends in implant dentistry training and education: A narrative review
56	Soofi & Ahmed, 2019	A Systematic Review of Domains, Techniques, Delivery Modes and Validation Methods for Intelligent Tutoring Systems
57	Winkler-Schwartz et al., 2019	Artificial Intelligence in Medical Education: Best Practices Using Machine Learning to Assess Surgical Expertise in Virtual Reality Simulation
58	Zawacki-Richter et al., 2020	Systematic review of research on artificial intelligence applications in higher education - where are the educators?
59	Chassignol et al., 2018	Artificial Intelligence trends in education: a narrative overview
60	Pappas & Drigas, 2016	Incorporation of Artificial Intelligence Tutoring Techniques in Mathematics
61	Athanasios & Ioannidou, 2012	Artificial intelligence in special education: A decade review

Following the PRISMA guidelines, the study selection process is presented in Figure 1.

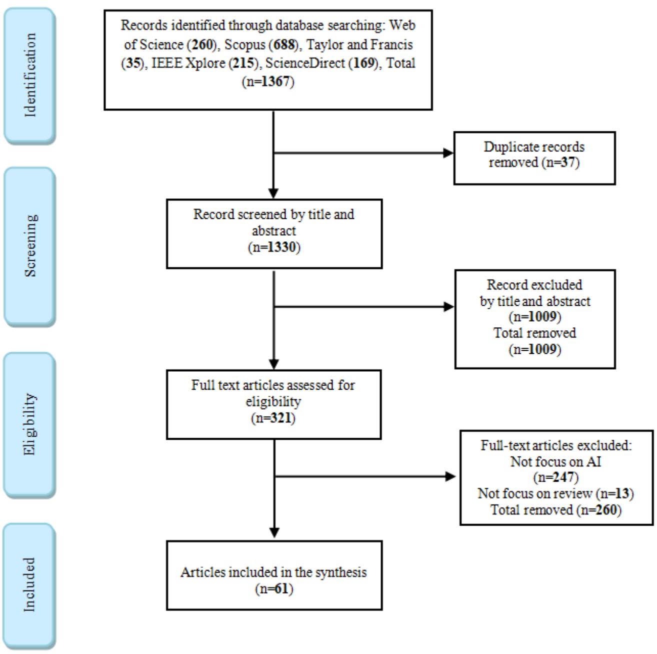


Figure 1. PRISMA Flowchart of the systematic review process.

Coding scheme

To assess the transparency of each identified literature review (among the 61 studies), this study uses the recommendation by Paré et al. (2016) on transparency and systematicity, which includes 17 questions split into six categories (see Table 4). Each paper was coded according to the information on the transparency characteristics to assess the level of its transparency. Specifically, for each question in Table 4, if the information exists, "Y" standing for "Yes" was assigned; otherwise, "N" standing for "No" was assigned. Particularly, all items with value = "Y" were counted and divided by the number of items in their group to calculate the transparency level in each group (e.g., for S01, we divided by 3, for S02, we divided by 4, for S04, we divided by 2, etc.). The researchers grouped the subtotal by the set of groups (six groups) throughout the assessment schema to calculate the overall assessment level. The data-gathering procedure was carried out throughout the article to reduce the risk of incomplete information, which is not mentioned in the methodology section. To reduce the opportunity for bias, an electronic data extraction form was designed (Tlili et al., 2022b), where two coders filled it in according to the coding scheme (see Table 5). To further ensure the reliability of the coding results, weekly meetings during the whole coding process were organized between the coders to discuss their coding progress, where disagreements were discussed and resolved by consensus.

Table 4. Transparency assessment in review articles (Paré et al., 2016).

Review steps to be assessed	Elements to be assessed in each step
S01 – Review planning	1. Are the objectives of review articles well argued? 2. Is the review type and procedures well described and justified in the study? (For reviews articles, use established frameworks or recommendations.) 3. Is the procedure or protocol described and publicized?
S02 – Search strategy	4. Is the search technique (for example, databases with coverage dates) well-defined? 5. Are the criteria for inclusion and exclusion explicitly mentioned? 6. Is there a complete electronic search technique for at minimum one database? (search terms & keywords) 7. Is there information about reference management tools and techniques, as well as other research processes?
S03 – Study selection	8. Is there a description of the screening and selection processes of the study mentioned? 9. Are there enough details about the included studies? 10. (if applicable) Is there a list of excluded studies, together with the reasons for the exclusion mentioned? 11. Is there a flow diagram depicting the study selection process?
S04 – Quality assessment	12. Are the results of each study's quality assessment presented? 13. Is there a description of the methods for integrating assessments into analyses?
S05 – Data extraction strategy	14. Are the methods and processes for data extraction mentioned? 15. Are the data extracted forms or items presented?

S06 – analysis and interpretation strategies	16. Are the main constructs or desired outcomes stated? 17. Are the analysis and synthesis methods explained and justified?

To assess the transparency level among the six criteria/ dimensions (see Table 4), the following information in Table 5 was coded. This information can help to provide comprehensive and deep insights related to the 17 items within the six steps (see Table 4).

Table 5. Coding scheme.

Code	Rationale, definitions, and examples
Article keywords	Author's article keywords
Electronic databases	The name of the database or the name of the electronic scientific index
Keywords and search terms	Authors' use of keywords and search terms
Years included	Years of studies included
Audience	Practitioners, researchers, policymakers, researchers, not stated, etc.
Type of primary sources	If the authors have mentioned: Conceptual/theoretical paper, Empirical – qualitative/quantitative, literature review, critical paper, bibliometric, mixed methods.
Number of studies	The total number of included reviewed articles.
General purpose	The author's purpose, intention, or aim for the article.
Research questions or Hypothesis	(If applicable) The author's research topics, hypotheses, or proposals
Review paper	"Y" if the authors clearly declared that their paper was a review paper
Author's review paper type	(if mentioned) the review type selected
Coders' review paper type	Review types assigned by coders
Scope of the research question	"N" means the scope is narrow; "B" means the scope is broad
Search strategy	Including representational, comprehensive, iterative, purposeful, and selective.
Articles explicitly mention the selection processes	"Y" is yes, "N" is no
Quality appraisal	"Y" means that the article is systematically screened for relevance and authenticity; "N" means that the article is not systematically screened
Nature of primary source	Contains meta-analysis, systematic analysis, qualitative, quantitative, and conceptual
Search method, which includes databases and timelines with a clear description	"Y" means that the article explicitly mentioned the search method; "N" means that the article did not
The disclosure of inclusion and exclusion criteria	"Y" if the article clearly listed the inclusion and exclusion criteria; "N" if the article did not
Tools and methods used to manage bibliographic materials	"Y" if the article clearly listed the methods used; "N" if the article did not
Revealing the list of the profile of included articles	"Y" if the article mentioned the list of included articles; "N" if the article did not
Revealing the list of the profile of excluded articles and the reasons for exclusion	"Y" if the article mentioned the excluded articles and the reasons of exclusion list; "N" if the article did not
Thorough information reporting on quality assessment results for specific sources	"Y" if the article detailed quality assessment results; "N" if the article did not
Information reporting on the quality assessment method	"Y" if the article assessed the quality of reviewed studies by collecting and analyzing data to demonstrate that a given study met the standards; "N" if the article did not
Data extraction techniques and methodologies reported	"Y" if the article reported on what methods and techniques were used to extract the data; "N" if the article did not
Disclosure of specific items or data extraction forms for structured data collection	"Y" if the article reported specific items or forms for data extraction; "N" if the article did not
Principal constructs or outcomes of interest mentioned	"Y" if the article detailed the principal constructs or outcomes; "N" if the article did not
Methods of analysis and synthesis described and justified	"Y" if the article detailed the analysis method with justifications; "N" if the article did not

Results and discussion

Descriptive summary of the AIED systematic reviews

Table 6 shows the goal of the included 61 literature reviews on AIED. These articles were published between 2012 and 2023. Additionally, based on the overarching goal of a literature review (Table 1), 57.36% of the AIED reviews had a primary purpose of describing a phenomenon with little or no addition to the theory (Rowe et al., 2012), as 54.09% of them were descriptive, while only 3.27% were narrative. Table 6 shows that the second highest type of review article is the critical review (31.14%) which comes under understanding. In contrast, review paper types other than descriptive and critical reviews are underrepresented, calling for more research in this context to cover those less applied types of literature review. It is noteworthy that the descriptive type's dominance probably stems from the relatively recent history of AI technologies. It is also important to note that the novelty effect of AI technologies expectedly urges researchers to investigate the phenomenon through descriptive review designs (Johnson et al., 2022). However, this tendency creates an imbalance and is a potential drawback to gaining a deeper understanding of AIED.

Table 6. The classification of systematic review articles (type and year) 2012–2023.

Overarching goal	Percentage	Literature review type	12	13	14	15	16	17	18	19	20	21	22	23	Total	%
Describing	57.36%	Narrative	0	0	0	0	0	0	1	1	0	0	0	0	2	3.27
		Descriptive	0	0	0	0	1	0	0	3	6	17	5	1	33	54.09
Understanding	31.14%	Scoping	0	0	0	0	0	0	0	0	0	3	2	1	6	9.83
		Critical	1	0	0	0	0	0	0	0	4	8	0	0	13	21.31
Testing theory	8.19%	Meta-analysis	0	0	0	0	0	0	0	0	0	2	3	0	5	8.19
		Qualitative systematic	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Explaining	3.27%	Theoretical	0	0	0	0	0	0	0	0	0	2	0	0	2	3.27
		Total	1	0	0	0	1	0	1	4	10	32	3	0	61	100%

Transparency assessment of the literature reviews

This section assesses the transparency criteria of each step (the six steps, see Table 4) of the reviewed articles. Each step is discussed in a subsequent section, while the overall transparency level is discussed in the final section.

Developing a review plan (S01)

This section assesses the transparency of the step of developing a review plan. The most important aspects of guaranteeing systematicity are planning the strategy, identifying the problem, proclaiming the purpose and research questions, and selecting and explaining the review type (Paré et al., 2016). Creating a review plan further improves the review process's systematicity and serves as the foundation for more extensive reporting of methodological decisions made during the study process (Templier & Pare, 2018). As demonstrated in Table 7, 81% (n = 50) of the review articles explicitly mentioned the targeted audience (researchers, practitioners, policy-makers, teachers, analysts, and students). Particularly, more than 70% of the reviewed articles centred on teachers as the audience when discussing AIED. Further, as shown in Table 6, 100% (n = 61) of the reviewed articles clearly stated their objectives or purposes by explicitly mentioning the research questions or hypotheses. Additionally, 81% (n = 50) of the studies clearly mentioned the scope of the research question (see Table 7).

We classified the literature reviews according to the type and breadth of their research questions based on their stated objectives and research questions (when available). All articles (n = 61) declared explicitly that their work was a review paper (i.e., the authors stated directly that their study was a review paper) and the authors' review paper type. However, only 50% (n = 31) of the articles mentioned the coders' review paper type (review types assigned by coders). Additionally, in the studies included, justifications for the review type selection were not identified. Furthermore, in the descriptive reviews, the authors stated their review objectives, types, and protocols at higher rates than in narrative, critical, scoping, or meta-analysis reviews. The use of explicit frameworks and guidelines for undertaking a literature review helps to explain some of these findings (Snyder, 2019).

Based on the above results, it is very important that authors put more attention on the rationale for choosing to conduct one type of literature review and not the other based on the research questions to be answered. For instance, if the authors want to measure the impact of a specific educational intervention, conducting a meta-analysis would be the most adequate type for this objective. Adding such information could increase the transparency of the conducted literature review, and help readers understand the ultimate goal of conducing a given literature review generally, and on AIED particularly.

Searching the literature (S02)

As shown in Table 7, 55% of the articles reported their search methods, which include databases and timelines with a clear description. Specifically, 75% of the articles reported the review period, and 81% of the articles reported the search queries. Additionally, all articles (100%) reported the electronic databases used to identify the research corpus. Particularly, when analyzing the mentioned electronic databases (in all 61 articles), Scopus (n = 25) was the most preferred electronic database to identify and review AIED papers, followed by Web of Science (n = 15), Science direct (n = 15) and IEEE Xplore (n = 6). However, only 45% of the articles explicitly reported the search queries used during the process of building a research corpus, and only 8% of the articles reported the tools and methods used to manage bibliographic materials.

Based on the above results, it is seen that authors should elaborate more in terms of the used search keywords to conduct their systematic reviews on AIED. This is crucial, especially for a complex topic like AIED, where several technologies (e.g., deep learning, machine learning, natural language processing) and techniques (e.g., learning analytics, prediction and modelling) can be used interchangeably with AIED. Besides, it is seen that less attention has been put by authors to elaborate on the used tools to manage their bibliographic materials, and this could be because these tools are more managerial and do not impact the research quality in any way.

Selecting studies (S03)

The selection process of articles provides important information on the sources used for interpretation and analysis, as well as the techniques used to pick these sources, by disclosing information about these aspects (Tricco et al., 2011). Consequently, it highlights the papers that are not relevant to the researchers' search interests. A

Table 7. Results of the coded information for transparency assessment.

Code	Total	Percentage
Article keywords	59	96%
Electronic databases	61	100%
Keywords and search terms	28	45%
Years included	46	75%
Audience	50	81%
Type of primary sources	58	95%
Number of studies	48	78%
General purpose	61	100%
Research questions or hypotheses	50	81%
Review paper	61	100%
Author's review paper type	58	95%
Coders' review paper type	31	50%
Scope of the research question	55	90%
Search strategy	46	75%
Articles explicitly mention the selection processes	46	75%
Quality appraisal	34	55%
Nature of primary source	45	73%
Search method, which includes databases and timelines with clear description	34	55%
The disclosure of inclusion and exclusion criteria	43	70%
Tools and methods used to manage bibliographic materials	5	8%
Revealing the list of the profile of included articles	31	50%
Revealing the list of the profile of excluded articles and the reasons for exclusion	21	34%
Thorough information reporting on quality assessment results for specific sources	5	8%
Information reporting on the quality assessment method	4	6%
Data extraction techniques and methodologies reported	10	16%
Disclosure of specific items or data extraction forms for structured data collection	36	59%
Principal constructs or outcomes of interest mentioned	37	60%
Methods of analysis and synthesis described and justified	36	59%

comprehensive and detailed screening approach decreases the risk of bias when it comes to including or excluding articles for further research. Researchers also give evidence on the usefulness of these studies for generating relevant results and addressing the study questions by giving thorough information about the included and excluded articles.

As presented in Table 7, 78% (n = 48) of the review articles mentioned the number of studies included in the review process. It is also critical for researchers and practitioners to have precise and organized information regarding exclusion methods since this allows them to assess the criteria's soundness and scientific rigour. Moreover, reporting inclusion and exclusion criteria provide insights that lead to generating research findings through reviewing the related literature and providing the base criteria for replicating the study and benchmarking the research process. Specifically, 70% of the articles mentioned their inclusion/exclusion criteria. 50% of the articles revealed the list of the profile of included articles, and just 34% of the articles provided information about the inclusion and exclusion criteria. Particularly, 50% (n = 31) of the publications presented the list of studies that were included and 45% (n = 28) of the studies graphically published data.

Inclusion/exclusion criteria are an important step to help readers understand how a given study might or might not be included within a given literature review. Based on the obtained results, it is seen that most studies did not elaborate on this step, as well as the final list of included studies. Consequently, this makes those conducted AIED literature reviews a black box, where it is not clear what was included and why. This also hinders understanding the obtained results, hence making full use of them to advance the AIED field, as the literature review input (i.e., included studies and how they are selected) is absent.

Assessing quality (S04)

This information can guarantee that only high-quality sources are obtained as a method of improving the quality of the findings and outcomes (Bandara et al., 2015). As shown in Table 7, only 8% of the articles thoroughly reported the quality assessment results for specific resources, and 6% of the articles reported information on quality assessment methods. On the other hand, 73% of the studies clearly mentioned the nature of the primary source. However, only 55% of the review studies mentioned the quality appraisal, where they compared the covered articles collectively. These findings are alarming, as assessing quality is related to the robustness of the research conducted.

Therefore, to increase the adoption and use of the obtained results given by some AIED literature reviews, the quality assessment should be highlighted. This is because researchers might always be hesitant to rely on some findings that they are not sure of their quality. This is even more pertinent in the AIED field as designing AI-based educational systems is very tricky and requires careful attention to not accidently harm (e.g., biased interventions) users (e.g., learners, educators) instead of supporting them.

Extracting data or key aspects from included studies (S05)

Whittemore et al. (2014) stated that the accurate reporting of individual research is vital to improving the quality of any knowledge synthesis technique. This aims to identify, organize, and carry out agreed-upon methods for collecting data from primary sources to mitigate the risks of omission, misclassification, or misrepresenting crucial information (Paré et al., 2016; Webster et al., 2002).

As shown in Table 7, only 16% of the articles reported their data extraction techniques and methods, whereas 59% of the articles disclosed the specific items or data extraction types for structured data collection. Furthermore, 95% of the reviewed articles included the items or information needed to extract data from their primary sources (e.g., descriptive, narrative, evidence-based qualitative, evidence-based quantitative, conceptual or theoretical paper, critical paper, bibliometric, mixed methods, or literature review). Some articles utilized a list of items to display this information (e.g., Al-Azawei et al., 2016), whereas others used structured approaches with more specific information (e.g., Arbaugh, 2014).

Based on the obtained results, authors should elaborate more on their extraction techniques, especially their coding scheme, to increase the replicability of their research (i.e., literature reviews) by others.

Synthesizing and interpreting data and formulating conclusions (S06)

As shown in Table 7, 60% of the studies indicated the major constructs or outcomes, while 59% described the analytical and synthesis methodologies. In both cases, the number of descriptive reviews outnumbered the remainder of the review articles.

This result might reveal that it is always easier to elaborate on some descriptive analysis, while it is not the case when the analysis is more advanced. For instance, when discussing meta-analysis review papers, there is a need to go beyond the simple description of the process and elaborate on the motivation of selecting a given technique, for instance, related to measuring effect size (Cohen's d and Hedges' g) or publication bias based on the different samples or research methods used in each included study within the review process.

The overall level transparency (S07)

Table 8 summarizes the overall level of transparency of the 61 review articles on AIED arranged by type and quartiles, with Q1 being the highest and Q4 the lowest. Specifically, a study belongs to Q1 if its overall transparency level is between 76% and 100%, Q2 if its overall transparency level is between 51% and 75%, Q3 if its overall transparency level is between 26% and 50%, and Q4 if its overall transparency level is between 0% and 25%. When describing research efforts, the findings reflect various levels of transparency; 40% of the articles were in Q1 and 32% in Q2. This implies that the transparency level of the conducted literature reviews on AIED is low. Therefore, future research should consider the transparency factor of the conducted literature reviews on AIED, as this may provide detailed insights about the field and positively impact the scientific community more broadly (Vom Brocke et al., 2018).

Additionally, descriptive review articles were in the top two quartiles (Q1 and Q2), while critical review articles (11 out of 13) were in the last quartile (Q4), implying that descriptive review articles have the highest transparency level while critical review articles have the lowest. Contrary to our findings, Castro-Gil and Correa (2021) found that the lowest transparency level was in descriptive literature reviews on blended learning in higher education.

Furthermore, Table 8 shows that step four "quality assessment" had the lowest transparency level. This implies that the reported reviews on AIED did not explicitly discuss the quality of the reviewed articles. Consequently, this might hinder the quality of the reported findings related to AIED. In this context, to ensure quality assessment when conducting review articles, several studies focused only on reviewing SSCI/SCIE or top journals in the field (e.g., Crompton & Burke, 2018; Hwang & Tsai, 2011).

Table 8. Studies fulfilling the transparency assessment items by type of review paper and quartile.

Review type and number of studies	Steps of the review process (S01–S06) and transparency assessment items (1–17)																	Quartiles				Total	%
	S01			S02				S03				S04		S05		S06		Q1	Q2	Q3	Q4		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17						
Descriptive	33	31	28	28	32	30	2	31	23	16	22	6	6	32	32	33	32	17	14	0	1	32	52.4
Narrative	1	1	0	0	1	0	0	1	0	0	0	0	0	1	1	1	1	0	1	1	1	3	4.9
Critical	3	2	3	2	2	2	0	1	1	0	1	0	0	1	1	1	1	0	1	1	11	13	21.3
Scoping	6	6	6	6	6	5	5	5	2	2	6	2	2	6	6	6	6	5	1	0	0	6	9.8
Meta-analysis	4	3	2	2	3	1	0	5	5	3	3	3	2	4	4	4	4	2	2	1	0	5	8.1
Theoretical	2	2	2	2	2	0	2	2	1	0	0	0	0	2	2	2	2	1	1	0	0	2	3.2
Total	49	45	41	40	46	38	9	44	39	21	32	11	10	46	46	4	46	40 %	32 %	4.9 %	22.95 %	61	100

Conclusions, implications, and limitations

This study conducted a transparency assessment of AIED review articles. The obtained findings showed that the transparency level is considerably low. Specifically, researchers should focus more on elaborating on the quality assessment of the reviewed articles, as well as the included and excluded articles.

The findings of this study can contribute to the educational technology field from different perspectives. From a theoretical perspective, this study can enrich the ongoing debate about the dimensions to consider for applying a transparent systematic review generally, and on AIED particularly. From a methodological perspective, this study presents how to conduct a transparency assessment of articles, as well as how to enhance the methodological part of a given literature review to obtain valid and reproducible research by others. From a practical perspective, this study can contribute to the AIED field by highlighting to researchers and practitioners the weaknesses of the conducted AIED literature reviews. Enhancing these parts can contribute to enhancing the quality of the obtained findings related to AIED, hence providing more insights to the working community on AIED, as well as providing evidence-based practices or making decision processes related to AIED.

Despite the solid ground of this study, it has some limitations that should be acknowledged. For instance, the covered literature review articles in this study might be limited due to the search queries and electronic databases used. Therefore, interested researchers can further complement the research presented in this study. Future research can focus on going beyond assessing the transparency level to analyze how the conducted literature reviews tackled AIED (e.g., from which perspective, the targeted stakeholders, education level and context, etc.). This might reveal the trends of AIED, as well as the gaps that researchers and practitioners should focus on in the future.

References

Ahmad, S. F., Rahmat, M. K., Mubarik, M. S., Alam, M. M., & Hyder, S. I. (2021). Artificial intelligence and its role in education. *Sustainability, 13*(22), 12902.

Alamri, R., & Alharbi, B. (2021). Explainable student performance prediction models: A systematic review. *IEEE Access, 9,* 33132-33143.

Al-Ataby, A. (2020). Technology-enhanced learning and teaching in COVID-19 era: Challenges and recommendations. *International Journal for Innovation Education and Research, 8*(10), 317–331. doi:10.31686/ijier.vol8.iss10.2684

Al-Azawei, A., Serenelli, F., & Lundqvist, K. (2016). Universal Design for Learning (UDL): A content analysis of peer-reviewed journals from 2012 to 2015. *Journal of the Scholarship of Teaching and Learning, 16*(3), 39–56. doi:10.14434/josotl.v16i3.19295

Alfaro, L., Rivera, C., Castaneda, E., Zuniga-Cueva, J., Rivera-Chavez, M., & Fialho, F. (2020). A review of intelligent tutorial systems in computer and web based education. *International Journal of Advanced Computer Science and Applications, 11*(2), 755-763.

Arbaugh, J. B. (2014). What might online delivery teach us about blended management education? Prior perspectives and future directions. *Journal of Management Education, 38*(6), 784–817. doi:10.1177/1052562914534244

Aslam, S. M., Jilani, A. K., Sultana, J., & Almutairi, L. (2021). Feature evaluation of emerging e-learning systems using machine learning: An extensive survey. *IEEE Access, 9,* 69573-69587.

Athanasios, S., & Ioannidou, R. (2012). Artificial Intelligence in Special Education: A decade review. *International Journal of Engineering Education, 28*(6), 1366-1372.

Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., & Beekhuyzen, J. (2015). Achieving rigor in literature reviews: Insights from qualitative data analysis and tool-support. *Communications of the Association for Information Systems, 37*, 8154-8204. doi:10.17705/1cais.03708

Balaji, P., Alelyani, S., Qahmash, A., & Mohana, M. (2021). Contributions of machine learning models towards student academic performance prediction: A systematic review. *Applied Sciences, 11*(21), 10007.

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review,* 37, 100489.

Bilgic, E., Gorgy, A., Yang, A., Cwintal, M., Ranjbar, H., Kahla, K., ... & Harley, J. M. (2022). Exploring the roles of artificial intelligence in surgical education: A scoping review. *The American Journal of Surgery, 224*(1), 205-216.

Bozkurt, A., Karadeniz, A., Baneres, D., Guerrero-Roldán, A. E., & Rodríguez, M. E. (2021). Artificial intelligence and reflections from educational landscape: A review of AI Studies in half a century. *Sustainability, 13*(2), 800.

Castro-Gil, R., & Correa, D. (2021). Transparency in previous literature reviews about blended learning in higher education. *Education and Information Technologies, 26*(3), 3399–3426. doi:10.1007/s10639-020-10406-x

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: A narrative overview. *Procedia Computer Science, 136*, 16-24.

Chen, L., Chen, P., & Lin, Z. (2020a). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264-75278.

Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020b). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1,* 100002.

Chu, S. T., Hwang, G. J., & Tu, Y. F. (2022). Artificial intelligence-based robots in education: A systematic review of selected SSCI publications. *Computers and education: Artificial intelligence,* 100091.

Conati, C., Porayska-Pomsta, K., & Mavrikis, M. (2018). AI in Education needs interpretable machine learning: Lessons from Open Learner Modelling. *ArXiv*, abs/1807.00154.

Crompton, H., & Burke, D. (2018). The use of mobile learning in Higher Education: A systematic review. *Computers & Education, 123*, 53–64. doi:10.1016/j.compedu.2018.04.007

Dai, C. P., & Ke, F. (2022). Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Computers and Education: Artificial Intelligence,* 100087.

Du, X., Yang, J., Hung, J. L., & Shelton, B. (2020). Educational data mining: a systematic review of research and emerging trends. *Information Discovery and Delivery, 48*(4), 225-236.

Essalmi, F., Tlili, A., Ayed, L. J. B., & Jemni, M. (2017). Toward modeling the learner's personality using educational games. *International Journal of Distance Education Technologies (IJDET), 15*(4), 21-38. doi:10.4018/IJDET.2017100102

Ferro, A. S., Nicholson, K., & Koka, S. (2019). Innovative trends in implant dentistry training and education: A narrative review. *Journal of Clinical Medicine, 8*(10), 1618.

Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies, 4*(4), 134-147.

Guilherme, A. (2017). AI and education: The importance of teacher and student relations. *AI & Society, 34*(1), 47–54. doi:10.1007/s00146-017-0693-8

González-Calatayud, V., Prendes, M. P., & Gutiérrez, J. M. (2022). The impact of educational data mining and learning analytics on the educational field: A bibliometric analysis. *Education Sciences, 12*(8), 178.

Hahn, M. G., Navarro, S. M. B., Valentín, L. D. L. F., & Burgos, D. (2021). A systematic review of the effects of automatic scoring and automatic feedback in educational settings. *IEEE Access, 9,* 108190-108198.

Hannan, E., & Liu, S. (2023). AI: new source of competitiveness in higher education. *Competitiveness Review: An International Business Journal, 33*(2), 265-279.

Huang, J., Saleh, S., & Liu, Y. (2021). A review on artificial intelligence in education. *Academic Journal of Interdisciplinary Studies, 10*(3), 206-206.

Hwang, G. J., & Chang, C. Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 1-14.

Hwang, G. J., & Tu, Y. F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics, 9*(6), 584.

Hwang, G.-J., & Tsai, C.-C. (2011). Research trends in Mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology, 42*(4). doi:10.1111/j.1467-8535.2011.01183.x

Johnson, P. C., Laurell, C., Ots, M., & Sandström, C. (2022). Digital Innovation and the effects of artificial intelligence on firms' research and development – Automation or augmentation, exploration or exploitation?, *Technological Forecasting and Social Change, 179*, 121636. doi:10.1016/j.techfore.2022.121636

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence, 2*, 100017.

Kahn, K., & Winters, N. (2021). Constructionism and AI: A history and possible futures. *British Journal of Educational Technology, 52*(3), 1130-1142.

Kharbat, F. F., Alshawabkeh, A., & Woolsey, M. L. (2021). Identifying gaps in using artificial intelligence to support students with intellectual disabilities from education and health perspectives. *Aslib Journal of Information Management, 73*(1), 101-128.

Kirubarajan, A., Young, D., Khan, S., Crasto, N., Sobel, M., & Sussman, D. (2022). Artificial intelligence and surgical education: A systematic scoping review of interventions. *Journal of Surgical Education, 79*(2), 500-515.

Klyuchnikov, D. A., Shurukhina, T. N., Gavrilova, T. A., Zhikharev, A. G., & Deeney, I. A. (2021). Some Aspects of Ai-Technologies in Education. *Revista San Gregorio, 1*(44).

Larsson, S., & Heinz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review, 9*(2), 1-16.

Laupichler, M. C., Aster, A., Schirch, J., & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence,* 100101.

Lee, J., Wu, A. S., Li, D., & Kulasegaram, K. M. (2021). Artificial intelligence in undergraduate medical education: a scoping review. *Academic Medicine, 96*(11S), S62-S70.

Liang, J. C., Hwang, G. J., Chen, M. R. A., & Darmawansah, D. (2021). Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments,* 1-27.

Liu, K., & Afzaal, M. (2021). Artificial intelligence (AI) and translation teaching: a critical perspective on the transformation of education. *International Journal of Educational Sciences, 33*(1), 64-73.

Luan, H., & Tsai, C. C. (2021). A review of using machine learning approaches for precision education. *Educational Technology & Society, 24*(1), 250-266.

Maier, U., & Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence, 3*, 100080.

Maghsudi, S., Lan, A., Xu, J., & van Der Schaar, M. (2021). Personalized education in the artificial intelligence era: What to expect next. *IEEE Signal Processing Magazine, 38*(3), 37-50.

McIntosh, L., Juehne, A., Hudson Vitale, C., Liu, X., Alcoser, R., Lukas, J. C., & Evanoff, B. (2017). Repeat: A framework to assess empirical reproducibility in biomedical research. *BMC Medical Research Methodology, 17*(143). doi:10.31219/osf.io/4np66

Murtaza, M., Ahmed, Y., Shamsi, J. A., Sherwani, F., & Usman, M. (2022). AI-based personalized e-learning systems: Issues, challenges, and solutions. *IEEE Access, 10,* 81323-81342.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041.

Nigam, A., Pasricha, R., Singh, T., & Churi, P. (2021). A systematic review on AI-based proctoring systems: Past, present and future. *Education and Information Technologies, 26*(5), 6421-6445.

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies, 27*(6), 7893-7925.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*(1), 1–11.

Pan, L., Tlili, A., Li, J., Jiang, F., Shi, G., Yu, H., & Yang, J. (2021). How to implement game-based learning in a smart classroom? A model based on a systematic literature review and Delphi method. *Frontiers in Psychology, 12*, 749837. https://doi.org/10.3389/fpsyg.2021.749837

Pappas, M., & Drigas, A. (2016). Incorporation of artificial intelligence tutoring techniques in mathematics. *International Journal of Engineering Pedagogy (iJEP), 6*(4), 12–16.

Paré, G., Tate, M., Johnstone, D., & Kitsiou, S. (2016). Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. *European Journal of Information Systems, 25*(6), 493–508. doi:10.1057/s41303-016-0020-3

Rowe, F. (2014). What literature review is not: Diversity, boundaries and recommendations. *European Journal of Information Systems, 23*(3), 241-255.

Rowe, M., Frantz, J., & Bozalek, V. (2012). The role of blended learning in the clinical education of healthcare students: A systematic review. *Medical Teacher, 34*(4), e216-e221.

Sapci, A. H., & Sapci, H. A. (2020). Artificial intelligence education and tools for medical and health informatics students: Systematic review. *JMIR Medical Education, 6*(1), e19285.

Sekeroglu, B., Abiyev, R., Ilhan, A., Arslan, M., & Idoko, J. B. (2021). Systematic literature review on machine learning and student performance prediction: Critical gaps and possible remedies. *Applied Sciences, 11*(22), 10907.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. doi:10.1016/j.jbusres.2019.07.039

Soofi, A. A., & Ahmed, M. U. (2019). A systematic review of domains, techniques, delivery modes and validation methods for intelligent tutoring systems. *International Journal of Advanced Computer Science and Applications, 10*(3), 99-107.

Sousa, M., Dal Mas, F., Pesqueira, A., Lemos, C., Verde, J. M., & Cobianchi, L. (2021). The potential of AI in health higher education to increase the students' learning outcomes. *TEM Journal,* (2), 488-497.

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence, 4*, 100124.

Su, J., & Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. *Computers and Education: Artificial Intelligence, 3*, 100049.

Su, J., Zhong, Y., & Ng, D. T. K. (2022). A meta-review of literature on educational approaches for teaching AI at the K-12 levels in the Asia-Pacific region. *Computers and Education: Artificial Intelligence, 3*, 100065.

Tahiru, F. (2021). AI in education: A systematic literature review. *Journal of Cases on Information Technology (JCIT), 23*(1), 1-20.

Tan, S. C., Lee, A. V. Y., & Lee, M. (2022). A systematic review of artificial intelligence techniques for collaborative learning over the past two decades. *Computers and Education: Artificial Intelligence*, 100097.

Tang, K. Y., Chang, C. Y., & Hwang, G. J. (2021). Trends in artificial intelligence-supported e-learning: A systematic review and co-citation network analysis (1998–2019). *Interactive Learning Environments,* 1-19.

Templier, M., & Pare, G. (2018). Transparency in literature reviews: An assessment of reporting practices across review types and genres in top IS journals. *European Journal of Information Systems, 27*(5), 503-550.

Tlili, A., & Burgos, D. (2022a). Unleashing the power of Open Educational Practices (OEP) through Artificial Intelligence (AI): Where to begin?. *Interactive Learning Environments,* 1-8. https://doi.org/10.1080/10494820.2022.2101595

Tlili, A., Padilla-Zea, N., Garzón, J., Wang, Y., Kinshuk, K., & Burgos, D. (2022b). The changing landscape of mobile learning pedagogy: A systematic literature review. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2022.2039948

Tlili, A., Zhang, J., Papamitsiou, Z., Manske, S., Huang, R., Kinshuk, & Hoppe, H. U. (2021). Towards utilising emerging technologies to address the challenges of using open educational resources: A vision of the future. *Educational Technology Research and Development, 69*, 515-532. https://doi.org/10.1007/s11423-021-09993-4

Tricco, A. C., Tetzlaff, J., & Moher, D. (2011). The art and science of knowledge synthesis. *Journal of Clinical Epidemiology, 64*(1), 11-20.

Tuval-Mashiach, R. (2017). Raising the curtain: The importance of transparency in qualitative research. *Qualitative Psychology, 4*(2), 126–138. doi:10.1037/qup0000062

Vázquez-Cano, E. (2021). Artificial intelligence and education: A pedagogical challenge for the 21st century. *Educational Process: International Journal (EDUPIJ), 10*(3), 7-12.

Voskoglou, M. G., & Salem, A. B. M. (2020). Benefits and limitations of the artificial with respect to the traditional learning of mathematics. *Mathematics, 8*(4), 611.

Vom Brocke, J., Maaß, W., Buxmann, P., Maedche, A., Leimeister, J. M., & Pecht, G. (2018). Future work and enterprise systems. *Business & Information Systems Engineering, 60*(4), 357–366. doi:10.1007/s12599-018-0544-2

Wang, H., Tlili, A., Huang, R., Cai, Z., Li, M., Cheng, Z., ... & Fei, C. (2023). Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective. *Education and Information Technologies*, 1-36.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii–xxiii. doi:http://www.jstor.org/stable/4132319

Whittemore, R., Chao, A., Jang, M., Minges, K. E., & Park, C. (2014). Methods for knowledge synthesis: An overview. *Heart & Lung, 43*(5), 453–461. doi:10.1016/j.hrtlng.2014.05.014

Winkler-Schwartz, A., Bissonnette, V., Mirchi, N., Ponnudurai, N., Yilmaz, R., Ledwos, N., ... & Del Maestro, R. F. (2019). Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation. *Journal of surgical education, 76*(6), 1681-1690.

Xia, Q., Chiu, T. K., Zhou, X., Chai, C. S., & Cheng, M. (2022). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence,* 100118.

Xu, W., & Ouyang, F. (2022). A systematic review of AI role in the educational system based on a proposed conceptual framework. *Education and Information Technologies,* 1-29.

Yufeia, L., Salehb, S., Jiahuic, H., & Syed, S. M. (2020). Review of the application of artificial intelligence in education. *Integration (Amsterdam), 12*(8), 1-15.

Zafari, M., Bazargani, J. S., Sadeghi-Niaraki, A., & Choi, S. M. (2022). Artificial intelligence applications in K-12 education: A systematic literature review. *IEEE Access, 10,* 61905-61921.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2020). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education, 17*(1), 39.

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021a). A review of Artificial Intelligence (AI) in education from 2010 to 2020. *Complexity, 2021*, 1-18.

Zhai, X., Shi, L., & Nehm, R. H. (2021b). A meta-analysis of machine learning-based science assessments: Factors impacting machine-human score agreements. *Journal of Science Education and Technology, 30*, 361-379.

Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C., & Shi, L. (2020). Applying machine learning in science assessment: a systematic review. *Studies in Science Education, 56*(1), 111-151.

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence, 2*, 100025.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : **http://journals.sfu.ca/jalt/index.php/jalt/index**

Artificial intelligence in higher education. A protocol paper for a systematic literature review

Fadhil Ismail[A] — A *Lecturer, Kaplan Singapore*

Eunice Tan[B] — B *Head of Programme, Academic Division, Singapore Institute of Management, Singapore*

Jürgen Rudolph[C] — C *Director of Research, Kaplan Singapore*

Joseph Crawford[E] — D *Senior Lecturer in Management, University of Tasmania, Australia*

Shannon Tan[D] — E *Research Executive, Kaplan Singapore*

Keywords

Artificial Intelligence (AI);
assessments;
chatbots;
ChatGPT;
conversational agents;
curriculum design;
generative AI;
GPT (Generative Pretrained Transformer);
higher education;
large language models (LLMs);
systematic literature review.

Correspondence

Fadhil.md@gmail.com [A]

Article Info

Received 18 September 2023
Received in revised form 23 October 2023
Accepted 23 October 2023
Available online 23 October 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.34

Abstract

Higher education continues to be confronted with significant learning and teaching challenges. Still reeling from the effects of the pandemic, the sector has grappled for the past year with the advent and impact of generative artificial intelligence (AI). Since the introduction of ChatGPT by OpenAI in November 2022, a growing number of studies have discussed AI models and their impacts and influence on higher education. However, the novelty of what we aim to do in a future paper, outlined in the current one, lies in the systematicity of our approach. There is yet to be a study in which a systematic search strategy is developed to critically review extant research longitudinally across all available generative AI chatbot models within higher education. This protocol paper identifies a prospective systematic approach to reviewing the emergent literature. In addition, this protocol paper documents the structural approach to facilitate a systematic literature review. We seek to offer a systematic approach to create an open-access resource to support future learning and teaching scholars to gain timely access to pre-examined literature on different forms of generative AI and their impact on higher education. This protocol paper, as such, offers an approach that can be used to initiate closer scrutiny of the metadata of articles published on AI models in higher education since its initiation in November 2022. We also suggest that the protocol presented in this paper be considered a relevant and rigorous approach for conducting systematic literature reviews in other domains.

Introduction

ChatGPT-3.5 became publicly available in November 2022 (Haleem et al., 2022). This iteration became known for its quick and comprehensive responses to queries in various domains. However, upon further review, it turned out that these responses, despite their refined articulations and structured presentations, were, in many parts, inaccurate and unsubstantiated (Kleesiek et al., 2023; Sallam, 2023). Subsequently, there have been accelerated developments, discourses and predictions offered about generative AI chatbots from various scholars and organisations within the teaching and learning spaces. The rise of ChatGPT and other competing generative AI models is expected to transform teaching and learning journeys in academia (Rudolph et al., 2023a, 2023b). AI chatbots have evolved from being topics in intellectual discussions to challenging realities confronting every higher education stakeholder, including universities, institutional policymakers, lecturers, curriculum and assessment developers, and students.

The rapid and transformational effect the COVID-19 pandemic had on higher education learning and teaching left scholars expected to innovate and respond to the emergent challenges of continuity of learning. Journals with an average CiteScore of 7.3 saw hundreds of article retractions for poor-quality research rushed to print (Taros et al., 2023). The challenges are still being uncovered today (e.g., Glushenkova & Zagato, 2023). While the human cost of artificial intelligence is likely lower than that of the COVID-19 pandemic, there is a need for clear and quality approaches to consider the rapid and transformative effects of AI, particularly generative AI, on higher education learning and teaching.

This protocol paper documents the method for creating a systematic literature review to facilitate an analysis of studies on generative AI and higher education. The main inspiration for this effort is to provide an open-access resource to support and facilitate academics and stakeholders in higher education to gain timely access to the research literature on generative AI and higher education. Using a rigorous approach, this protocol paper proposes a systematic approach to craft out and analyse the metadata of articles published on specific types of generative AI and higher education one year after their release to consider the impact it has on shaping the future of higher education. By providing an open-access database, we aim to facilitate future research in AI chatbots and their global impact on the higher education space.

The introduction of ChatGPT-3.5 in late 2022 instigated an array of research studies in relation to ChatGPT, generative AI and higher education being published in various higher education learning and teaching journals and databases (e.g., Adarkwah et al., 2023; Chaka, 2023; Crawford et al., 2023; Eager & Brunton, 2023; Firat, 2023; Gamage et al., 2023; Hassoulas et al., 2023; Ifelebuegu, 2023; Ifelebuegu et al., 2023; Kelly et al., 2023; Khademi, 2023; Limna et al., 2023; Mills et al., 2023; Mohammadkarimi, 2023; O'Dea & O'Dea, 2023; Perkins, 2023; Popenici, 2023; Popenici et al., 2023; Rasul et al., 2023; Rudolph et al., 2023a, 2023b; Calonge et al., 2023; Sullivan et al., 2023; Xames & Shefa, 2023). These publications, while numerous, only paint a partial picture of an area of rapidly growing knowledge.

Given the exponential growth of publications, it is valuable to take stock of the findings in these papers and their respective quality. More importantly, this will enable the synthesis of the findings to understand the functions and implications of AI applications (Rudolph et al., 2023b). Specifically, it will allow us to determine the opportunities and threats to learning and teaching in higher education. Rasul et al. (2023) highlighted that ChatGPT and other generative AI chatbots can potentially enhance and augment learning outcomes and experiences in higher education. However, there is a need to investigate its potential benefits and challenges to ensure its ethical, effective, and responsible use. With the increase of publications doing this, it becomes critical to not only synthesise the information through comprehensive literature reviews but to also conduct meta-analyses to understand the implications of this increase in academic literature in different contexts. The novelty of what we aim to do in a future paper, as outlined by our recommended approach, lies in the systematicity of our approach.

Since the introduction of ChatGPT-3.5 by OpenAI in November 2022, a growing number of studies have discussed AI models and their impacts and influence on higher education. Although there is an excellent paper by Tlili et al. (2023) on transparent AI literature reviews, there is yet to be a study in which a systematic search strategy is developed to review extant research longitudinally across all available generative AI chatbot models within higher education. Moreover, current publications on AI applications in relation to higher education still tend to be in their infancy. Efforts to establish coherence among these publications are rather disjointed and, more often than not, conducted at a granular level. The dearth of systematic and macro-level research inspired our research team (based in Australia and Singapore) to create a rigorous research protocol to examine research on AI applications and higher education. A rigorous systematic review is sound when designed and administered effectively and aligned to core research thematic dimensions within a structured methodology (Crawford & Cifuentes-Faura, 2022). Toward this end, we crafted the following research objective:

> To design a rigorous research protocol to curate and conduct a systematic literature review on the first year of published literature on AI applications (e.g., Bard, Bing Chat, ChatGPT, and Ernie) to support policymakers, educators, and researchers in higher education.

This protocol paper, as such, offers an approach that can be used to initiate closer scrutiny of the metadata and findings of articles published on AI applications in the higher education space, such as ChatGPT. Moreover, we suggest that the protocol presented in this paper be considered a relevant and rigorous approach for conducting systematic literature reviews in other domains as well.

Method

Systematic reviews aim to collate and synthesise the extant state of knowledge on a particular area of research via a systematic, structured analysis of aggregated findings from research outputs based on prespecified criteria (Higgins et al., 2011; Motyka, 2018). Research metrics are useful instruments to assess the quality and impact of research outputs (Moed & Halevi, 2015). However, it is important to note that each metric only measures a particular aspect and has limitations (Nestor et al., 2020). Therefore, it is critical not to consider any particular metric in isolation but instead to consider a series of metric measurements to evaluate the quality of the database or journal. Consequently, the databases chosen for this systematic review were selected based on known metrics such as Journal Impact Factor, h-index, g-index, Eigenfactor score, and Altmetrics (alternative metrics).

Search strategy

This protocol paper suggests a systematic approach for article selection guided by PRISMA – the Preferred Reporting Items for Systematic Reviews and Meta-analyses (Moher et al., 2009; Page et al., 2021). Specifically, it outlines the reporting recommendations for systematic reviews suggested in the PRISMA 2020 guidelines to reflect recent developments and protocol suggestions in systematic review methodologies (Page et al., 2021). This approach to a systematic review is commonly used in educational and sustainability research and has been previously described by Butler-Henderson et al. (2020a, 2021a) and explored by Bearman et al. (2012). Following PRISMA search guidelines, the proposed systematic review will conduct a database search of all published journal articles (including those published online first) and preprints that relate to the topic of generative AI and teaching and learning in higher education. Special consideration will be paid to preprint articles for quality and those articles that are yet to undergo peer review. All research outputs published between 30 November 2022 and 31 December 2023 in the following sources will be considered: (1) Academic Search Ultimate, IEEE Xplore, Informit Online, Ovid, Proquest, ScienceDirect, Scopus, and Web of Science; (2) Google Scholar (the first ten pages for each search string will be reviewed). A snowball reference analysis will also be conducted based on the articles extracted first.

A comprehensive and rigorous search strategy requires clearly aligning the search phrases (search terms, keywords and Boolean Operators) to the thematic dimensions relevant to the study's research objectives. Within the context of this study, Boolean Operators and key search terms were generated based on definitions and thematic dimensions derived from the preliminary literature search, which are congruent to the study's focus on generative AI in higher education. The search strings suggested to be searched in the title, abstract, or keywords in the proposed protocol include the following: For each search, the first core strings (higher education, artificial intelligence, and the focal artificial intelligence) will be paired with one of the other strings to complete five strings. Where possible, we relied on existing reviews that included one of these framed, noting some required adaption to this context. For example, Spelt et al. (2009) use interdisciplinary in all searches (e.g., interdisciplinary curriculum), and this review splits concepts to refer to them separately. For the focal AI, it could include reviews on diverse generative AI chatbots (e.g., ChatGPT, GPT-4, Bard, Bing Chat, Claude, or Ernie) and generative non-chatbot AI (e.g. DALL-E, GitHub Copilot, GPT-4 plugins, Midjourney, Runway, or Synthesia).

Concept	Search string	Review that guided this frame.
1. **Higher education**	"Higher education" OR university* OR college OR tertiary OR undergrad* OR graduate OR postgrad*	Butler-Henderson et al. (2022); Zawacki-Richter et al. (2019)
2. **Artificial intelligence**	"artificial intelligence" OR "machine intelligence" OR "intelligent support" OR "intelligent virtual reality" OR "chat bot*" OR "machine learning" OR "automated tutor" OR "personal tutor*" OR "intelligent agent*" OR "expert system" OR "neural network" OR "natural language processing"	Zawacki-Richter et al. (2019)
3. Focal artificial intelligence	ChatGPT* OR "Chat Generative Pre-trained Transformer"	*Use specific tool related text.*
4. **Learning Setting**	Curricul* OR learn* OR student*	Zawacki-Richter et al. (2019)
5. **Education policies**	Polic*	Aikens et al. (2016)
6. **Assessment**	Assess*	Struyven et al. (2005)
7. **Teachers and lecturers**	Teach* OR Lectur*	Spelt et al. (2009)
8. **Pedagogical Approaches**	Pegagog*	Spelt et al. (2009)

Eligibility criteria, selection procedure, and quality assessment

In the search, only English-language academic journals and pre-prints are planned to be included, with a time-based limit of 12 months following release. Given ChatGPT was released towards the end of 2022, the first planned review will include papers up until 31 December 2023. The eligibility criteria for inclusion are as follows: articles related to teaching, curriculum, education, and students, including assessments, teaching practice, and course design in higher education, and relate to the specific artificial intelligence tool of reference. Excluded articles included those concerning university administrative processes unrelated to teaching and learning. In the instance where articles were about students but not related to teaching or learning, they were also excluded. For example, if the article discusses an issue that does not include any connection to higher education, that study will be excluded.

A double screening procedure will be adopted in the planned systematic review during the verification process across initial title and abstract screening and full-text screening to determine the final selection of sources of evidence for analysis. An appropriate reliability check (e.g., Cohen's Kappa) will be conducted with at least fair agreement between all pairs required prior to progression. The quality of the evidence gathered in the planned systematic review will be evaluated using the Cochrane Collaboration's tool for risk of bias assessment (Higgins et al., 2011; Page et al., 2021; Zeng et al., 2020) to minimise bias. The flow of information through this systematic review and aggregated findings based on the prespecified criteria will be subsequently reported through a PRISMA Statement (Figure 1). The quality assessment tool and PRISMA Checklist to appraise the study validity are discussed in the next section.

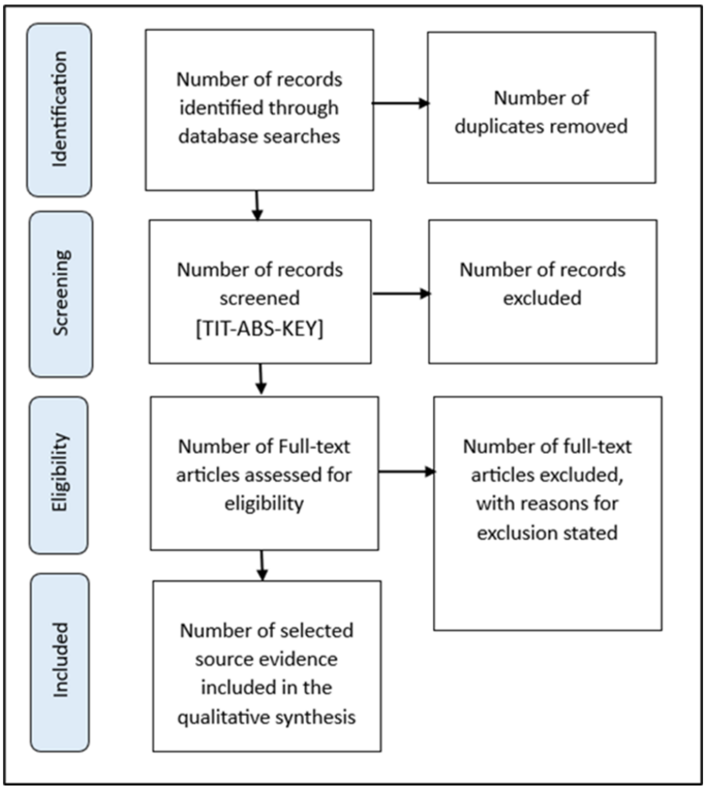


Figure 1: Proposed PRISMA statement.

Study validity assessment

We will use the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist and critical appraisal tools suited to the methods of the included studies to appraise and critically assess the validity of studies. The PRISMA checklist is a document that guides reporting systematic reviews and meta-analyses clearly and transparently. It ensures that the systematic review is written comprehensively and transparently so that readers can assess the quality and validity of the evaluation (PRISMA, 2020).

Critical Appraisal Tool

A critical appraisal tool assesses the dependability, significance, and practical relevance of evidence (Centre for Evidence-Based Medicine, 2023). In cases where any of these inquiries yields a negative response, it may be judicious to abstain from further engagement with the material (see Crowe & Sheppard, 2011; Zeng et al., 2015). An appropriate critical appraisal tool (or suite) will be selected and used for manuscripts included in the review based on the methods of included manuscripts.

Data coding and extraction strategy

Our data coding and extraction strategy includes the production of a detailed spreadsheet that will be available as an open-access database for scholarly reuse (similar to Butler-Henderson et al., 2021a). In constructing the database, we will incorporate certain theoretical assumptions detailed in Table 1. These are shared to present our reflexivity as researchers and to help others understand the adaptability of the data for their respective contexts. Although many data elements are clear and can be readily used in future research (like DOI, journal metadata, and country of origin), others, like the quality assessment score, study type, and participant type, necessitate further explanation.

The discipline and sub-discipline categories require some elaboration. The discipline category is grouped in four ways: health science, humanities and social science, STEM (science, technology, engineering and mathematics), and ‘others’. In addition, we categorise the subdisciplines in the same way the researchers did -- for example, ‘chemistry’ or ‘journalism’ (see Butler-Henderson et al., 2020a).

The type of study is defined as theoretical, quantitative, qualitative, or mixed methods. Should the method remain unclear, the field will be blank. For participants, possible categories are academic, professional or management, undergraduate student, postgraduate student, and doctoral student, mixed staff (including two or more categories of staff), mixed students (including two or more categories of students), and mixed staff and students (for samples comprising both students and staff) (see Butler-Henderson et al., 2020a).

Table 1: Description of data elements.

No.	Data element	Field type	Description
1	Year of publication	Numeric	the year of the manuscript publication
2	Month of publication	Numeric	the month of the manuscript publication
3	First author’s last name	Alphabetic	last name of first author of the study
4	Generative chatbots written about	Alphabetic	the names of the generative chatbots that the study focuses on (e.g. ChatGPT-3.5, GPT-4, Bard, Bing Chat, Claude, Ernie)
5	Non-chatbot generative AI	Alphabetic	the names of the generative non-chatbot AI that the study focuses on (e.g. DALL-E, GitHub Copilot, GPT-4 plugins, Midjourney, Runway, Synthesia)
6	Quality assessment score	Numeric	score derived from the quality tool (QATTL) used for the study: poor, low, medium, or high
7	Country of origin	Alphabetic	the country or countries of origin of the authors
8	Article type	Alphabetic	type of article (e.g. case study, commentary, research study, review article, opinion piece, etc.)
9	Type of study	Alphabetic	type of study (qualitative, quantitative, mixed methods, or theoretical)
10	Participant type	Alphabetic	type of participants (academic, professional or management, undergraduate student, postgraduate student, and doctoral student, mixed staff (including two or more categories of staff), mixed students (including two or more categories of students), and mixed staff and students (for samples comprising both students and staff)

11	Discipline	Alphabetic	broad higher education discipline groupings (e.g. health science, humanities, STEM)
12	Sub-discipline	Alphabetic	the higher education sub-discipline if available (e.g. chemistry, dentistry, teacher education)
13	Authors (listing all authors)	Alphabetic	names of all authors of study (APA7 style)
14	Title	Alphabetic	study title (APA7 style)
15	Journal	Alphabetic	name of journal where study is published
16	Volume	Numeric	volume number of journal where study is published
17	Issue	Numeric	issue number of journal where study is published
18	Pages	Numeric	page numbers in journal where study is published
19	DOI	URL	digital object identifier of the study

To test the repeatability of our process, the description of the above data elements will be executed with different researchers. The outcomes from each repetition will be recorded and compared for consistency using the metrics described in Table 1. In instances of missing or unclear data, authors will be contacted via email. They will be given a window of 14 calendar days to provide the necessary clarifications. Once received, the clarified data will be cross-referenced with the original submission to ensure consistency.

To ensure intercoder reliability, all coders will undergo standardised training using Table 1 as a shared coding manual. Their outputs will be periodically cross-checked against one another to assess consistency. Reliability will be statistically measured using Cohen's Kappa, with a threshold set at 0.80 (Warrens, 2015). Should reliability metrics fall below this, coders will undergo refresher training sessions, and the problematic data will be recoded.

Data synthesis and presentation

The process of writing a narrative synthesis can be particularly challenging, especially if the review includes a large number of different types of studies. Approaches like Braun and Clarke's (2006) thematic analysis can be helpful in developing initial themes for presentation. Complexities can also arise from examining a variety of complex interventions and outcomes. However, it is important to note that adopting a systematic approach to synthesis is key to making sense of the results in these different studies. The research question that underpins the review will determine the type of approach chosen to synthesise and present the findings of the review. The process of synthesising the data must also be rigorous and transparent, completely aligned with the methods specified in the protocol. These methods should be justified and followed systematically.

A narrative synthesis can be a useful first step in analysing and organising the data extracted from the review systematically and presenting the data in a coherent structure that can inform readers (Popay et al., 2006). Synthesising the results of different studies in a review in a narrative form is not simply describing or summarising the main features of each study, although doing this can be a useful start when writing systematic literature reviews. It would be practical, for example, to describe and comment on the methodological quality of each study that may provide significant insights for readers to become acquainted with the data presented in them.

One way to approach a narrative synthesis is by combining and evaluating data from several different studies. This step is taken to draw insights and conclusions about outcomes, effects, limitations of the studies and the applicability of findings in these studies. A narrative synthesis includes an examination of the similarities and differences between the findings of different studies, as well as an exploration of patterns in the data presented in these studies. For example, the similarities and differences in study design, populations, interventions or other aspects of the study can be examined and presented. This could include examining related factors and associations between research study designs and the findings. Some examples of synthesising the findings from different studies could be comparing the different research designs (e.g., RCT or mixed-methods approach) or with possible explanations to account for the pattern of results. Another way of organising narrative synthesis could also be to look at the different interventions or implementation strategies in the studies. This might involve examining associations between their research purpose, the manner in which the findings will be applied, and any other factors influencing the design and conduct of the research study. Studies with incomplete or missing information or an ambiguous description of the data will not be included in this synthesis. In addition, quantitative information in the form of tables, graphs, and figures will be summarised and presented in table form in the narrative synthesis to compare the different findings of the literature examined.

The next steps

This protocol has designed and outlined a rigorous systematic review method to ensure the maximal utility of the information and metadata in the databases mentioned in this paper. This includes an approach that can be applied to conduct an extensive search across the literature, databases, and online sources to ensure coverage of publications for the curation of a database. This approach resource will be critical in supporting researchers, educators in higher education, curriculum designers, assessors and policymakers in learning and teaching and providing them with a guide to navigate the AI space. We aim to update this approach longitudinally, with additional time periods to refine and revise coding rules and include other relevant databases over the coming years to make this resource robust and relevant. This will provide us with an impetus to mitigate and manage the impacts of AI and other EdTech technologies. It will facilitate the shift for the global higher education community towards new insights in learning and teaching as we navigate an evolving realm challenged and transformed by AI applications.

References

Adarkwah, M. A., Amponsah, S., van Wyk, M. M., Huang, R., Tlili, A., Shehata, B., Metwally, A. H. S., & Wang, H. (2023). Awareness and acceptance of ChatGPT as a generative conversational AI for transforming education by Ghanaian academics: A two-phase study. *Journal of Applied Learning and Teaching, 6*(2), 1-16. https://doi.org/10.37074/jalt.2023.6.2.26

Aikens, K., McKenzie, M., & Vaughter, P. (2016). Environmental and sustainability education policy research: A systematic review of methodological and thematic trends. *Environmental Education Research, 22*(3), 333-359. https://doi.org/10.1080/13504622.2015.1135418

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development, 31*(5), 625-640. https://doi.org/10.1080/07294360.2012.702735

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Butler-Henderson, K., Crawford, J., Rudolph, J., Lalani, K., K. M. Sabu. (2020a). COVID-19 in Higher Education Literature Database (CHELD V1): An open access systematic literature review database with coding rules. *Journal of Applied Learning & Teaching, 3*(2), 11-16. https://doi.org/10.37074/jalt.2020.3.2.11.

Butler-Henderson, K., Crawford, J., Rudolph, J., Lalani, K. & Sahu, K. M, (2020b). COVID-19 in Higher Education Literature Database (CHELD V1). *Institute of Research Innovation.* https://doi.org/10.37074/jalt.2020.3.2.11d

Butler-Henderson, K., Tan, S., Lalani, K., Mandapam, S. K., Kemp, T., Rudolph, J., & Crawford, J. (2021a). Update of the COVID-19 Higher Education literature database (CHELD v2). *Journal of Applied Learning & Teaching, 4*(1), 134-137. https://doi.org/10.37074/jalt.2021.4.1.22

Butler-Henderson, K., Tan, S., Lalani, K., Karakka Mandapam, S., Kemp, T., Rudolph, J., Crawford, J. (2021b). COVID-19 in Higher Education Literature Database (CHELD). Version 2. *Institute of Research Innovation*. https://doi.org/10.37074/jalt.2021.4.1.22d

Calonge, D. S., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching, 6*(2), 1-15. https://doi.org/10.37074/jalt.2023.6.2.20

Centre for Evidence-Based Medicine. (CEBM). (2023). *Critical appraisal tools.* University of Oxford. https://www.cebm.ox.ac.uk/resources/ebm-tools/critical-appraisal-tools

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching, 6*(2), 1-11. https://doi.org/10.37074/jalt.2023.6.2.12

Crawford, J., & Cifuentes-Faura, J. (2022). Sustainability in higher education during the COVID-19 pandemic: A systematic review. *Sustainability, 14*(3), 1879. https://doi.org/10.3390/su14031879

Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice, 20*(3), 02. https://doi.org/10.53761/1.20.3.02

Crowe, M., & Sheppard, L. (2011). A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. *Journal of Clinical Epidemiology, 64*(1), 79-89. https://doi.org/10.1016/j.jclinepi.2010.02.008

Eager, B., & Brunton, R. (2023). Prompting higher education towards AI-augmented teaching and learning practice. *Journal of University Teaching & Learning Practice, 20*(5), 02. https://doi.org/10.53761/1.20.5.02

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching, 6*(1), 57-63. https://doi.org/10.37074/jalt.2023.6.1.22

Gamage, K. A., Dehideniya, S. C., Xu, Z., & Tang, X. (2023). ChatGPT and higher education assessments: More opportunities than concerns?. *Journal of Applied Learning and Teaching, 6*(2), 1-12. https://doi.org/10.37074/jalt.2023.6.2.32

Glushenkova, M., & Zagato, M. (2023). Effect of COVID-19 on digitalisation of higher education. A tale of one business school. *Journal of University Teaching & Learning Practice, 20*(6). https://doi.org/10.53761/1.20.6.10

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil transactions on Benchmarks, Standards and Evaluations, 2*(4), 100089. https://doi.org/10.1016/j.tbench.2023.100089

Hassoulas, A., Powell, N., Roberts, L., Umla-Runge, K., Gray, L., & Coffey, M. (2023). Investigating marker accuracy in differentiating between university scripts written by students and those produced using ChatGPT. *Journal of Applied Learning & Teaching, 6*(2), 1-7. https://doi.org/10.37074/jalt.2023.6.2.13

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... & Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343*, 1-9. https://doi.org/10.1136/bmj.d5928

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2), 1-8. https://doi.org/10.37074/jalt.2023.6.2.2

Ifelebuegu, A. O., Kulume, P., & Cherukut, P. (2023). Chatbots and AI in Education (AIEd) tools: The good, the bad, and the ugly. *Journal of Applied Learning and Teaching, 6*(2), 1-14. https://doi.org/10.37074/jalt.2023.6.2.29

Kelly, A., Sullivan, M., & Strampel, K. (2023). Generative artificial intelligence: University student awareness, experience, and confidence in use across disciplines. *Journal of University Teaching & Learning Practice, 20*(6), 12. https://doi.org/10.53761/1.20.6.12

Khademi, A. (2023). Can ChatGPT and bard generate aligned assessment items? A reliability analysis against human performance. *Journal of Applied Learning & Teaching, 6*(1), 75-80. https://doi.org/10.37074/jalt.2023.6.1.28

Kleesiek, J., Wu, Y., Stiglic, G., Egger, J., & Bian, J. (2023). An opinion on ChatGPT in health care—written by humans only. *Journal of Nuclear Medicine, 64*(5), 701-703. https://doi.org/10.2967/jnumed.123.265687

Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching, 6*(1), 64-74. https://doi.org/10.37074/jalt.2023.6.1.32

Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1), 16-30. https://doi.org/10.37074/jalt.2023.6.1.34

Moed, H. F., & Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology, 66*(10), 1988-2002. https://doi.org/10.1002/asi.23314

Motyka, B. (2018). Employee engagement and performance: A systematic literature review. *International Journal of Management and Economics, 54*(3), 227-244. https://doi.org/10.2478/ijme-2018-0018

Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *Journal of Applied Learning and Teaching, 6*(2), 1-9. https://doi.org/10.37074/jalt.2023.6.2.10

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264-269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Nestor, M. S., Fischer, D. L., Arnold, D., Berman, B., & Del Rosso, J. Q. (2020). Rethinking the journal impact factor and publishing in the digital age. *The Journal of Clinical and Aesthetic Dermatology, 13*(1), 12. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7028381/pdf/jcad_13_1_12.pdf

O'Dea, X. C., & O'Dea, M. (2023). Is Artificial Intelligence really the next big thing in learning and teaching in higher education? A conceptual paper. *Journal of University Teaching and Learning Practice, 20*(5), 05. https://doi.org/10.53761/1.20.5.05

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ, 372*. http://dx.doi.org/10.1136/bmj.n160

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice, 20*(2), 07. https://doi.org/10.53761/1.20.02.07

Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching, 6*(2), 1-7. https://doi.org/10.37074/jalt.2023.6.2.4

Popenici, S., Rudolph, J., Tan, S., & Tan, S. (2023). A critical perspective on generative AI and learning futures. An interview with Stefan Popenici. *Journal of Applied Learning and Teaching, 6*(2), 1-21. https://doi.org/10.37074/jalt.2023.6.2.5

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version, 1*(1), b92. https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf

PRISMA. (2020). *PRISMA checklist*. http://prisma-statement.org/prismastatement/checklist.aspx?AspxAutoDetectCookieSupport=1

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Sallam, M. (2023). ChatGPT utility in healthcare, education, research and practice. Systematic review on the promising perspectives and valid concerns. *Healthcare, 11*(1), 887. https://doi.org/10.3390/healthcare11060887

Spelt, E. J., Biemans, H. J., Tobi, H., Luning, P. A., & Mulder, M. (2009). Teaching and learning in interdisciplinary higher education: A systematic review. *Educational Psychology Review, 21*, 365-378. https://doi.org/10.1007/s10648-009-9113-z

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education, 30*(4), 325-341. https://doi.org/10.1080/02602930500099102

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching, 6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Taros, T., Zoppo, C., Yee, N., Hanna, J., & MacGinnis, C. (2023). Retracted Covid-19 articles: significantly more cited than other articles within their journal of origin. *Scientometrics, 128*(5), 2935-2943. https://doi.org/10.1007/s11192-023-04707-4

Tlili, A., Huang, R., Mustafa, M. Y., Zhao, J., Bozkurt, A., Xu, L., ... & Burgos, D. Speaking of transparency: Are all Artificial Intelligence (AI) literature reviews in education transparent? *Journal of Applied Learning and Teaching, 6*(2), 1-12. https://doi.org/10.37074/jalt.2023.6.2.15

Warrens, M. J. (2015). Five ways to look at Cohen's kappa. *Journal of Psychology & Psychotherapy, 5*. https://doi.org/10.4172/2161-0487.1000197

Xames, M. D., & Shefa, J. (2023). ChatGPT for research and publication: Opportunities and challenges. *Journal of Applied Learning and Teaching, 6*(1), 390-395. https://doi.org/10.37074/jalt.2023.6.1.20

Zawacki Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International Journal of Educational Technology in Higher Education, 16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zeng, X., Zhang, Y., Kwong, J. S., Zhang, C., Li, S., Sun, F., ... & Du, L. (2015). The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: A systematic review. *Journal of Evidence-based Medicine, 8*(1), 2-10. https://doi.org/10.1111/jebm.12141

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Construction and standardisation of an instrument measuring lecturers' persistence to publish in Scopus-indexed journals

Christian M. Stracke[A] — A University of Bonn, Germany

Irene-Angelica Chounta[B] — B University of Duisburg-Essen, Germany

Wayne Holmes[C] — C University College of London, UK

Ahmed Tlili[D] — D Beijing Normal University, China

Aras Bozkurt[E] — E Anadolu University, Turkey

Keywords

AIED;
AI literacy;
AI&ED;
Artificial Intelligence;
education;
ethical;
learning;
PRISMA;
systematic review;
trustworthy.

Correspondence

stracke@uni-bonn.de [A]

Article Info

Received 31 October 2023
Received in revised form 25 November 2023
Accepted 26 November 2023
Available online 27 November 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.38

Abstract

By using standardised approaches, systematic reviews of the educational, scientific literature can inform educational research and influence educational policies and practices. However, the various systematic reviews of the scientific literature in the field of Artificial Intelligence (AI) and education all adopt individual approaches, making it challenging to systematically compare their conclusions. Accordingly, this paper presents a standardised protocol for conducting systematic reviews of the scientific literature on AI and education (AI&ED), including both literature on teaching and learning with AI (AIED) and literature on teaching and learning about AI (AI literacy). Our protocol applies the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines and is presented here for the purpose of replication and validation. We exemplify our protocol by means of a systematic review of the scientific literature on *trustworthy and ethical AI&ED*, which was undertaken iteratively in symbiosis with the development of the protocol, informing each other throughout. In the future, we intend to apply our novel protocol for other search terms of relevance to AI&ED, as well as for the same search terms over a longer time period, in order to allow comparisons and the exploration of trends.

Introduction

Artificial Intelligence (AI) has been controversial since it was first introduced (Aiken & Epstein, 2000; Chaka, 2023; Huang et al., 2023; McCarthy et al., 1955). Nonetheless, it has now infiltrated almost all academic disciplines and most aspects of life outside academia (Borenstein & Howard, 2021). In particular, teaching and learning with AI (AIED) has been researched for around 50 years (Dillenbourg, 2016; Holmes et al., 2019; Holmes & Tuomi, 2022; Ifelebuegu et al., 2023; Mills et al., 2023; Pinkwart, 2016), both in K-12 education (Hrastinski et al., 2019) and in Higher Education (Crompton et al., 2020; Rasul et al., 2023). However, the implementation of AIED in classrooms, although growing rapidly, is still in its early stages. The same is true of teaching and learning about AI (AI literacy) (Holmes et al., 2022a).

Systematic reviews of the educational scientific literature using standardised approaches can inform educational research and can influence policies and practices in education. In fact, several systematic reviews of AIED have been published, including reviews about AIED in K-12 (Crompton et al., 2022; Sanusi et al., 2022), AIED in Higher Education (Zawacki-Richter et al., 2019), and AIED for specific educational purposes (Kurdi et al., 2020; Sottilare et al., 2018).

Meanwhile, global sustainability is emerging as an ambitious objective of AI developments. This is particularly true in the context of education (Chen et al., 2020; Chounta et al., 2022; European Parliament, 2021; Miao et al., 2021). The argument is that, to achieve a sustainable society, one thing that is necessary is to improve education about technology's (especially AI's) impact on humans, society and the environment (AI literacy) (Holmes et al., 2022a; Holmes & Tuomi, 2022; Holmes, 2023). In other words, to help ensure a sustainable society, we need students and citizens to have competences in both, the human and technological dimensions of AI, alongside other digital competences (Holmes et al., 2022b; Stracke et al., 2022a, 2022b, Vuorikari & Holmes, 2022). However, to date there has been limited research on the teaching and learning about these human and technological dimensions of AI literacy.

While, as noted, several systematic reviews have analysed the state-of-the-art of AIED (e.g., Chen et al., 2020; Crompton et al., 2022; Sanusi et al., 2022; Zawacki-Richter et al., 2019), they have mostly adopted individual approaches, making it challenging to systematically compare their outcomes and conclusions. Meanwhile, Tlili et al. (2023) have shown that the transparency level of literature reviews of AIED has been low. This low transparency level and the lack of comparability, due to there being no agreed or common approach, together highlight the need for a standardised protocol that might be used by researchers to enhance the transparency, comparability and quality of AIED (and, by extension, AI literacy) literature reviews. Such a protocol might better advance AI&ED (following Holmes et al., 2022a, we use AI&ED as shorthand for the combination of AIED and AI Literacy). However, prior to the reported study, no such common or standardised protocol or approach appeared to exist.

We contend that a key issue meriting systematic review within AI&ED is the intertwined conceptual pair of trustworthiness and ethics. Indeed, trustworthiness and ethics in AIED have been discussed in various recent publications (e.g., Bozkurt et al., 2023; European Commission, 2022; HLEG on AI, 2019; Holmes, 2023; Holmes et al., 2022a; Kazim & Koshiyama, 2021; Miao & Holmes, 2023; UNESCO, 2021). However, to the best of our knowledge, no study has systematically reviewed the relationship between education and trustworthy and/or ethical AIED, or indeed between education and trustworthy and/or ethical AI literacy.

To address these research gaps, the lack of an agreed protocol for systematic reviews of AI&ED and the lack of research into trustworthy and ethical AI&ED, this present study developed a standardised Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol for systematic reviews of AI&ED and, in symbiosis (both to illustrate and to inform the protocol), we undertook an example systematic review of trustworthy and ethical AI&ED. The protocol was designed to both, inform future research and to be used as a framework to help differentiate and classify theoretical concepts and practical approaches.

Our proposed standardised protocol for systematic reviews into AI&ED builds upon the PRISMA model in two ways. First, we identify and recommend particular starting search terms for the PRISMA phase 1, to which further terms can be added to narrow the search to the particular AI&ED issue of interest. Second, we identify and recommend particular inclusion and exclusion criteria for the PRISMA phases 2 and 3. In this way, our novel standardised protocol might help the comparability of future AI&ED systematic reviews, enhance the quality of AI&ED research findings, and increase the replicability of the methods adopted by AIED researchers and applications (in this sense, our protocol complements Ismail et al., 2023, which proposes a future systematic literature review on AIED in higher education also based on the PRISMA guidelines).

In summary, the present paper reports a standardised PRISMA-based protocol that researchers might adopt to conduct systematic reviews of scientific literature of AI&ED, to enhance the robustness of results and to enable systematic comparisons of results, that was developed in symbiosis with (i.e., each informing the other) an example project on trustworthy and ethical AI&ED. In the following sections, the full procedure of our example systematic review is presented in italic text. The outcomes of that example systematic review, which is not the core focus of this paper, will be presented in a separate paper; in this paper, we focus on the protocol. The pre-stage of the protocol, the selection of appropriate search terms, is described in the following section.

Pre-selection of search terms

The pre-selection of the search terms used for a systematic review is critical to ensure its feasibility. We propose that a fundamental requirement is that the records identified by the selected search terms can be handled by the researchers (*in our case, the co-authors of this paper*). Therefore, we adopt the principle that the number of identified records should be higher than 50, to allow a meaningful analysis but lower than 1,000 to avoid an impractical workload. Naturally, a higher number of reviewers would be able to handle a higher number of papers.

For our standardised protocol for future systematic reviews, we propose the use of the electronic database Web of Science (WoS, www.webofscience.com), because it offers rigorous indexing services available for scientific and peer-reviewed publications. In WoS, putting "TS" ("Topic") in the Advanced Search Query Builder search string causes the search to be undertaken in the following fields within each record: Title, Abstract, and Author Keywords. Comparing the results of various search strings for our example systematic review, for our standardised protocol, we decided upon the search term "TS = ((Artificial Intelligence) AND (education*) AND ([ISSUE OF INTEREST]))". We also tried including alternative terms for "artificial intelligence", but it became clear that presumably because other terms and synonyms are only used in combination with "artificial intelligence", this did not reveal a noticeably greater number of records.

In our example systematic review, we submitted several "TS" search strings to the Advanced Search Query Builder on 21st of November 2022 (see Table 1). In order to identify papers that considered trustworthy and ethical AI&ED, we replaced "([ISSUE OF INTEREST])" with "((trust) OR (ethic*))".*

Table 1: Search strings used in the example systematic review's pre-selection, and the numbers of identified records.

#	Search string	Identified records
1	TS = (Artificial Intelligence)	79,429
2	TS = (education*)	1,203,303
3	TS = ((Artificial Intelligence) AND (education*))	3,694
4	*TS = (trust*)*	*164,738*
5	*TS = (ethic*)*	*265,794*
6	*TS = ((trust*) OR (ethic*))*	*423,626*
7	*TS = ((Artificial Intelligence) AND ((trust*) OR (ethic*)))*	*4,068*
8	*TS = ((Artificial Intelligence) AND (education*) AND ((trust*) OR (ethic*)))*	*324*
9	*TS = ((Artificial Intelligence) AND (education*) AND ((trust*) OR (ethic*)) AND (higher education*))*	*92*

For our example systematic review, we selected search term #8 (i.e., "TS = ((Artificial Intelligence) AND (education) AND ((trust*) OR (ethic*)))"), because it included our issue of interest ("trustworthy and ethical") and identified a number of records that could be properly analysed by a small team (as is typical of most research labs).*

The four selection phases of the proposed systematic review protocol, *illustrated by our example systematic review*, are detailed in the following section.

Standardised protocol methodology

Our standardised protocol defines the methodology for the systematic review of AI&ED. It strictly follows the PRISMA statement and its procedures while adapting and specifying the original four PRISMA phases for the specific selection of AI&ED articles (Moher et al., 2009; Page et al., 2021). The protocol was developed in symbiosis with our example *systematic review of trustworthy and ethical AI&ED* (both informing each other). The four PRISMA-based phases for the selection of articles are:

1. Identification,
2. Screening,
3. Eligibility, and
4. Included.

To ensure the reliability of the process, the four phases of the PRISMA process should be undertaken by at least two reviewers, each of whom having research experience in AI&ED. The rationale is that two knowledgeable researchers working independently and then sharing and discussing their results until they reach a consensus will reduce personal biases sufficiently.

In our example systematic review, the pre-selection of the search strings and the identification of the articles using the search terms were undertaken by two reviewers, supported by three further reviewers as required, all of whom have extensive research experience in AI&ED.

Phase 1

In the first PRISMA-based phase, *identification*, the selected database(s) is searched using the pre-defined search string. This phase is concluded by the elimination of duplicates.

The full procedure for the first phase of our example systematic review was as follows. The Advanced Search Query Builder of WoS was used with the search string ("TS = (("Artificial Intelligence") AND (education) AND ((trust) OR (ethic*)))") (see Section "Pre-selection of search terms"). As noted, this phase identified 324 records. The elimination of duplicates was not necessary here as only one data source was used.*

Phase 2

In the second PRISMA-based phase, *screening*, the title and abstract of the records identified in the first phase are reviewed, to identify and remove all records that do not meet the search aims. Table 2 gives an overview of the inclusion and exclusion criteria that we defined for the second phase of our standardised protocol.

As noted above, at least two reviewers should review in parallel the titles and abstracts of a randomly selected subset of all records. This subset should contain a minimum 5% of all records identified during the first phase because the reviewers should compare a substantial number of records after their independent reviews. Adopting a lower margin of 5% is a common sense decision designed to ensure that the

Table 2: Inclusion and exclusion criteria for screening the identified records.

Inclusion criteria	Exclusion criteria
The full text can be downloaded.	The full text of the record is not available.
The record is written in English.	The record is not written in English.
The record is published in a peer-reviewed journal.	The record is not published in a peer-reviewed journal (including conference proceedings).
The record is a scientific article.	The record is not a scientific article (instead it is an editorial, interview, research protocol, etc.).
The record is substantial (>2 pages).	The record is not substantial (<3 pages).

selection is representative while minimising the number of false positives. In their independent reviews, the reviewers should apply the inclusion and exclusion criteria outlined in Table 2. In cases of uncertainty, the related record should be kept. Afterwards, reviewers should compare their independent results for the random subset and discuss them in detail aiming to reach consensus on all decisions.

We propose that there are two possible outcomes of the independent reviews of the random subset during the second phase depending on the results of their independent reviews. The threshold of the criterion is set to zero because the reviewers should achieve common understanding and complete agreement about the inclusion or exclusion of records.

> Outcome 1: If the independent review results show one contradictory case or more, another random subset of records should be identified and independently reviewed in parallel. The subsequent independent review results should then again be compared and discussed as explained earlier.
>
> Outcome 2: If the independent review results are exactly the same, the remaining records should be shared among the reviewers to complete the second review phase. During that review, the reviewers should note all decisions about which they are not certain, for later discussion until a consensus is reached.

The full procedure for the second phase of our example systematic review was as follows. Two reviewers reviewed in parallel the titles and abstracts of a random subset of 5% of all the records identified in the first phase, using the inclusion and exclusion criteria outlined in Table 2. The researchers first worked independently, and then discussed their results until they reached a consensus (any records about which a researcher was uncertain or about which the researchers disagreed were discussed in depth in order to reach the consensus). In this way, personal biases were reduced. In our case, we took a first random subset of 24 records, leaving exactly 300 records that could be easily shared among the researchers to complete the second review phase.

Phase 3

In the third PRISMA-based phase, eligibility, the full texts of the remaining records are reviewed to finalise the selection of texts, ensuring that they all fulfil content-related requirements. Table 3 gives an overview of the inclusion and exclusion criteria that were defined for the third phase of our standardised protocol.

At least two reviewers should review in parallel the full text of a random subset of 50 records generated by the second phase, using the inclusion and exclusion criteria outlined in Table 3. The amount of 50 records is selected because the threshold can be set to two contradictory cases while still keeping the 95% margin of the normal distribution (2 cases out of 50 records are 4%).

Table 3: Inclusion and exclusion criteria for content-related screening of collected records.

Inclusion criteria	Exclusion criteria
The record addresses trustworthy or ethical AI.	The record does not address trustworthy or ethical AI.
The record addresses education.	The record does not address education.

As before, the researchers first should work independently, and then discuss their results until they reach a consensus (any records about which a researcher is uncertain or about which the researchers disagree should be discussed in depth in order to reach the consensus). Afterwards, reviewers should compare their independent results and discuss them in detail to reach consensus on all decisions. In cases of uncertainty, the records should be kept.

We propose that there are two possible outcomes of the third phase independent reviews of the random subset of 50 records depending on the results of their independent reviews. The threshold of the criterion is set to two because this limit keeps the 95% margin of the normal distribution (2 cases out of 50 records are 4%) and the reviewers need to discuss any contradictory case in details to achieve consensus in the end.

> Outcome 1: If the independent review results show more than two contradictory cases, then another random subset of records should be generated and independently reviewed in parallel. The independent review results should be compared and discussed in the same way as explained before.
>
> Outcome 2: If the independent review results are the same or differ in only one or two cases, the remaining records will then be shared among the reviewers to complete the third review phase. During that review, the reviewers should note all decisions about which they are not certain, for later discussion until a consensus is reached.

The full procedure for the third phase of our example systematic review closely followed the steps just outlined.

Phase four

In the fourth and final PRISMA-based phase, *included*, the remaining records are used for the actual systematic review, involving an in-depth analysis and discussion with respect of the research question(s). To begin with, the reviewers should propose an initial categorisation for the articles selected in the previous phases because such a categorisation is required for a systematic assignment and analysis of the articles. This categorisation should be discussed until a consensus about the terms and their categorical structure is reached. The discussion should include various dimensions of the topic in question, comments on general trends, and limitations of the systematic review and its analysis because all reviewers should be explicit about their analysis perspectives. The included papers should then be categorised according to this nominal taxonomy, using an iterative process. The systematic review will conclude with an outlook on future research needs and potential research questions.

The full procedure for the fourth phase of our example systematic review will closely follow the steps just outlined. As this example systematic review is not the core focus of this paper, its outcomes will be presented in a separate paper.

Results

The results we report here are for the standardised protocol (*the results of the example systematic review will be presented in a separate paper*). Figure 1 presents a template (*illustrated with numbers from our example systematic review*) that may be used to report the results of the four PRISMA-based phases determining the final selection of articles for in-depth analysis. It should be used, adapted with the appropriate numbers, for the four protocol phases of any future systematic review and its results.

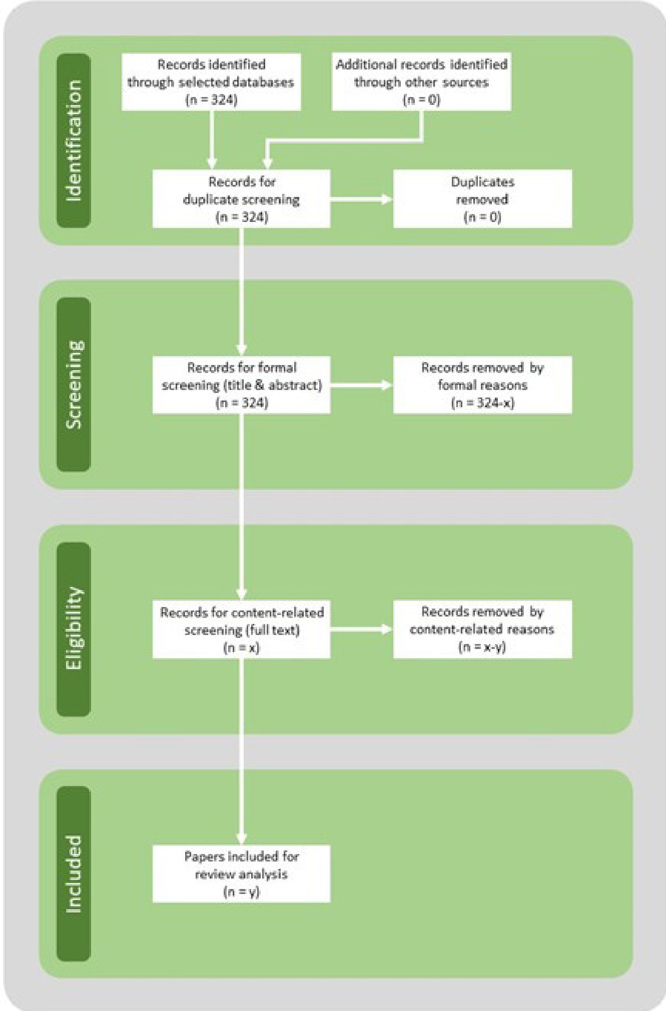


Figure 1: Template for the summary of the four phases of the standardised protocol.

Conclusions

This paper presented a standardised protocol which could serve as the basis of systematic literature reviews of AI&ED. It was developed in symbiosis with an example systematic review of trustworthy and ethical AI&ED research (the results will be published after finalising the analysis), informing each other throughout. The standardised protocol and the example systematic review were mutually informed by means of sense-testing and evaluation throughout.

To the best of our knowledge, no study in the literature has provided such a protocol on this increasingly important topic area before. The protocol identifies a suitable database (WoS), offers pre-defined search terms with the opportunity to fine-tune them to the issue of interest, and provides a structure that can be used in the systematic review of any aspect of AI&ED. By means of this standardised protocol, personal bias can be reduced, and the quality of the reported findings can be enhanced. It will be easier to systematically compare the results with those of other studies using the same protocol. In turn, this should advance AI&ED research, development and application.

In particular, our standardised protocol offers a template for undertaking future systematic reviews of AI&ED. The precise steps outlined above build a standardised protocol that anyone can easily repeat, and its repeated usage will lead to its validation and its continuous improvement. We envision that this protocol can contribute towards the standardisation of systematic reviews in the field of AI&ED, support the comparison of findings, enable the mapping of research trends over time, inform policymakers and educators, and influence policies and practices in AI&ED. Our current research focuses on applying our standardised protocol to a larger and up-to-date dataset of trustworthy and ethical AI&ED.

Acknowledgements

This paper was supported by Anadolu University as part of a project supervised by Aras Bozkurt (Grant number: SBA-2023-1852).

References

Aiken, R. M., & Epstein, R. G. (2000). Ethical guidelines for AI in education: Starting a conversation. *International Journal of Artificial Intelligence in Education, 11*, 163–176.

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics, 1*(1), 61–65. https://doi.org/10.1007/s43681-020-00002-7

Bozkurt, A., Karadeniz, K., Baneres, D., Guerrero-Roldán, A. E., & Rodríguez, M. E. (2021). Artificial intelligence and reflections from educational landscape: A review of AI studies in half a century. *Sustainability, 13*(2), 1-16. https://doi.org/10.3390/su13020800

Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., ..., & Jandrić, P. (2023). Speculative futures on ChatGPT and generative Artificial Intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education, 18*(1), 53-130. https://doi.org/10.5281/zenodo.7636568

Chaka, C. (2023). Fourth industrial revolution—a review of applications, prospects, and challenges for Artificial Intelligence, robotics and blockchain in higher education. *Research and Practice in Technology Enhanced Learning, 18*(2), 002. https://doi.org/10.58459/rptel.2023.18002

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chounta, I.-A., Bardone, E., Raudsep, A., & Pedaste, M. (2022). Exploring teachers' perceptions of Artificial Intelligence as a tool to support their practice in Estonian K-12 education. *International Journal of Artificial Intelligence in Education, 32*(3), 725-755. https://www.doi.org/10.1007/s40593-021-00243-5

Crompton, H., Bernacki, M. L., & Greene, J. (2020). Psychological foundations of emerging technologies for teaching and learning in higher education. *Current Opinion in Psychology, 36*, 101–105. https://doi.org/10.1016/j.copsyc.2020.04.011

Crompton, H., Jones, M. V., & Burke, D. (2022). Affordances and challenges of Artificial Intelligence in K-12 education: A systematic review. *Journal of Research on Technology in Education.* https://doi.org/10.1080/15391523.2022.2121344

Dillenbourg, P. (2016). The evolution of research on digital education. *International Journal of Artificial Intelligence in Education, 26*, 544-60. https://doi.org/10.1007/s40593-016-0106-z

European Commission. (2022). *Ethical guidelines on the use of Artificial Intelligence (AI) and data in teaching and learning for educators.* https://data.europa.eu/doi/10.2766/153756

European Parliament (2021). *Report on Artificial Intelligence in education, culture and the audiovisual sector (2020/2017(INI))*. Committee on Culture and Education. https://www.europarl.europa.eu/doceo/document/A-9-2021-0127_EN.html

High-Level Expert Group (HLEG) on AI. (2019). *Ethics guidelines for trustworthy AI.* https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Holmes, W. (2023). *The unintended consequences of artificial intelligence and education.* Education International Research. https://www.ei-ie.org/en/item/28115:the-unintended-consequences-of-artificial-intelligence-and-education

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in education. Promises and implications for teaching and learning.* Center for Curriculum Redesign. Boston, MA.

Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova., V. (2022a). *Artificial Intelligence and Education. A critical view through the lens of human rights, democracy and the rule of law.* https://rm.coe.int/artificial-intelligence-and-education-a-critical-view-through-the-lens/1680a886bd

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., ..., & Koedinger, K. R. (2022b). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education 32*(3), 504-526. https://www.doi.org/10.1007/s40593-021-00239-1

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education, 57*(4), 542–570. https://doi.org/10.1111/ejed.12533

Hrastinski, S., Olofsson, A. D., Arkenback, C., Ekström, S., Ericsson, E., Fransson, G., ..., & Utterberg, M. (2019). Critical imaginaries and reflections on Artificial Intelligence and robots in post digital K-12 education. *Postdigital Science and Education, 1*(2), 427–445. https://doi.org/10.1007/s42438-019-00046-x

Huang, R., Tlili, A., Xu, L., Chen, Y., Zheng, L., Saleh Metwally, A. H., ..., & Bonk, C. J. (2023). Educational futures of intelligent synergies between humans, digital twins, avatars, and robots - the iSTAR framework. *Journal of Applied Learning & Teaching, 6*(2), 1-16. https://doi.org/10.37074/jalt.2023.6.2.33

Ifelebuegu, A. O., Kulume, P., & Cherekut, P. (2023). Chatbots and AI in Education (AIEd) tools: The good, the bad, and the ugly. *Journal of Applied Learning and Teaching, 6*(2), 1-14. https://doi.org/10.37074/jalt.2023.6.2.29

Ismail, F., Tan, E., Rudolph, J., Crawford, J., & Tan, S. (2023). Artificial intelligence in higher education. A protocol paper for a systematic literature review. *Journal of Applied Learning & Teaching 6*(2), 1-8. https://doi.org/10.37074/jalt.2023.6.2.34, last accessed 2023/10/26.

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns, 2*(9), 100314. https://doi.org/10.1016/j.patter.2021.100314

Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education, 30*, 121-204. https://doi.org/10.1007/s40593-019-00186-y

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). *A proposal for Dartmouth summer research project on Artificial Intelligence*. https://www-formal.stanford.edu/jmc/history/dartmouth.pdf

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: Guidance for policy-makers*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000376709

Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1), 16-30. https://doi.org/10.37074/jalt.2023.6.1.34

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ..., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews, 10*, 89. https://doi.org/10.1186/s13643-021-01626-4

Pinkwart, N. (2016). Another 25 years of AIED? Challenges and opportunities for intelligent educational technologies of the future. *International Journal of Artificial Intelligence in Education, 26*, 771-83. https://doi.org/10.1007/s40593-016-0099-7

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ..., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J., & Tukiainen, M. (2022). A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies*. https://doi.org/10.1007/s10639-022-11416-7

Sottilare, R. A., Burke, S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2018). Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence in Education, 28*, 225-64. https://doi.org/10.1007/s40593-017-0146-z

Stracke, C. M., Burgos, D., Santos-Hermosa, G., Bozkurt, A., Sharma, R. C., Swiatek, C., ..., & Truong, V. (2022a). Responding to the initial challenge of COVID-19 pandemic: Analysis of international responses and impact in school and higher education. *Sustainability, 14*(3), 1876. https://doi.org/10.3390/su14031876

Stracke, C. M., Sharma, R. C., Bozkurt, A., Burgos, D., Swiatek, C., Inamorato dos Santos, A., ..., & Truong, V. (2022b). Impact of COVID-19 on formal education: An international review on practices and potentials of Open Education at a distance. *The International Review of Research in Open and Distributed Learning, 23*(4), 1-18. https://doi.org/10.19173/irrodl.v23i4.6120

Tlili, A., Huang, R., Mustafa, M. Y., Zhao, J., Bozkurt, A., Xu, L., ..., & Burgos, D. (2023). Speaking of transparency: Are all Artificial Intelligence (AI) literature reviews in education transparent? *Journal of Applied Learning and Teaching, 6*(2), 1-12. https://doi.org/10.37074/jalt.2023.6.2.15

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2021). *Recommendation on the Ethics of Artificial Intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Vuorikari, R., & Holmes, W. (2022). DigComp 2.2. Annex 2. Citizens Interacting with AI Systems. In R. Vuorikari, S. Kluzer, & Y. Punie (Eds.), *DigComp 2.2, The digital competence framework for citizens: With new examples of knowledge, skills and attitudes* (pp. 72–82). Publications Office of the European Union. https://data.europa.eu/doi/10.2760/115376

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on Artificial Intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : **2591-801X**

Content Available at : **http://journals.sfu.ca/jalt/index.php/jalt/index**

Investigating marker accuracy in differentiating between university scripts written by students and those produced using ChatGPT

Athanasios Hassoulas[A] — *[A] Cardiff University, Cardiff, United Kingdom*

Ned Powell[B] — *[B] Cardiff University, Cardiff, United Kingdom*

Lindsay Roberts[C] — *[C] Cardiff University, Cardiff, United Kingdom*

Katja Umla-Runge[D] — *[D] Cardiff University, Cardiff, United Kingdom*

Laurence Gray[E] — *[E] Cardiff University, Cardiff, United Kingdom*

Marcus J Coffey[F] — *[F] Professor, Cardiff University, Cardiff, United Kingdom*

Keywords

Assessment;
ChatGPT;
digital education;
generative AI;
higher education;
medical education.

Correspondence

HassoulasA2@cardiff.ac.uk [A]

Article Info

Received 16 July 2023
Received in revised form 21 July 2023
Accepted 22 July 2023
Available online 25 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.13

Abstract

The introduction of OpenAI's ChatGPT has widely been considered a turning point for assessment in higher education. Whilst we find ourselves on the precipice of a profoundly disruptive technology, generative artificial intelligence (AI) is here to stay. At present, institutions around the world are considering how best to respond to such new and emerging tools, ranging from outright bans to re-evaluating assessment strategies. In evaluating the extent of the problem that these tools pose to the marking of assessments, a study was designed to investigate marker accuracy in differentiating between scripts prepared by students and those produced using generative AI. A survey containing undergraduate reflective writing scripts and postgraduate extended essays was administered to markers at a medical school in Wales, UK. The markers were asked to assess the scripts on writing style and content, and to indicate whether they believed the scripts to have been produced by students or ChatGPT. Of the 34 markers recruited, only 23% and 19% were able to correctly identify the ChatGPT undergraduate and postgraduate scripts, respectively. A significant effect of suspected script authorship was found for script content, $X^2(4, n=34) = 10.41$, $p<0.05$, suggesting that written content holds clues as to how markers assign authorship. We recommend consideration be given to how generative AI can be responsibly integrated into assessment strategies and expanding our definition of what constitutes academic misconduct in light of this new technology.

Introduction

The use of technology in enhancing coursework submissions is by no means a new trend. From Microsoft Word's spell check and autocorrect to the more recent use of products such as Grammarly, the use of such tools has significantly improved our ability to produce well-structured written documents with the aid of inbuilt spelling and grammar assistants (Behrens et al., 2019). Artificial intelligence (AI) technology dates back decades to platforms such as ELIZA, which utilised early language models to engage in conversation, but more sophisticated generative AI technology is now capable of producing written scripts that pose a problem for higher education assessments (Rudolph et al., 2023a).

The introduction of OpenAI's ChatGPT (Generative Pre-Trained Transformer), in particular, has been viewed as a watershed moment in higher education due to the ability of the tool, through the large language model (LLM) it employs, to learn rapidly and develop sophisticated responses to a range of instructions. Objectively, ChatGPT is, therefore, the first such LLM that has captivated a global mainstream audience (Hosseini et al., 2023). The various applications of this technology for educators, researchers and students have been demonstrated impressively through a published journal article written by the chatbot on what its existence means for higher education (Bishop, 2023).

The consensus within global higher education is that the technology is here to stay and will have profound consequences for assessment strategies across all programmes of study. Immediate discussions and challenges will pertain to updating our definition, or perhaps redefining, terms such as plagiarism and academic integrity in light of this revolutionary technology (e.g., Debby et al., 2023). The advantages that this new technology also presents, however, cannot be ignored. Not only do disruptive tools such as ChatGPT provide an ideal opportunity to modernise certain outdated assessment practices, but they may, when used appropriately, significantly enhance students' learning experiences and productivity (Fauzi et al., 2023). Indeed, the technology may revolutionise the manner in which students learn and work academically.

Conversely, in the context of academic integrity, others assert that this new technology may not be as disruptive as is currently anticipated (Cotton et al., 2023), and some have suggested that this potential issue could be addressed by replacing some assessments with formats that require evidence of reflective practice by students. However, even without further evolution, it appears likely that even the current commonly available generative AI tools may be capable of deceiving coursework markers reviewing reflective student scripts as well as essay-type assessments.

To our knowledge, there has been no published study to date comparing marker accuracy in differentiating between human-written coursework submissions and AI chatbot-generated scripts in both essay-type scripts and reflective writing tasks. On this basis, we designed a study that included original student submissions and scripts generated by ChatGPT-3.5 and then investigated the performance of experienced coursework markers in terms of how they graded the assessments, as well as determining whether they could accurately differentiate between the student submissions and ChatGPT scripts.

Materials and methods

Participants

A total of 34 experienced academic and clinical academic coursework markers from a medical school in Wales were recruited to participate in this study. Participants were presented with undergraduate reflective writing submissions and postgraduate extended essays. Participants had the option to review just the reflective pieces, just the essays, or both, and were asked to review the submission formats they routinely marked. 23 participants marked the undergraduate reflective writing submissions, and 22 participants marked the postgraduate essay scripts (11 participants marked both undergraduate and postgraduate scripts). Participant confidentiality and response anonymity were assured. Consent was provided by all participants, as well as by the students whose scripts were anonymously included as examples of undergraduate reflective writing and postgraduate essays, with all identifiable information removed before being included in the survey.

Materials and procedure

The survey was designed in and disseminated using the digital survey platform Online Surveys (formerly Bristol Online Surveys). Three undergraduate reflective writing scripts were presented, along with three postgraduate extended essay scripts. Each undergraduate reflec-tive writing script was approximately 1,500 words in length, whilst the postgraduate extended essays were each approximately 3,000 words in length. Of the three reflective writing scripts, two were student submissions, and one was generated using ChatGPT-3.5. Equally, two of the postgraduate essays were student submissions, whilst one essay was generated using ChatGPT-3.5. For the undergraduate reflective writing task, the wording of the instructions provided to students was identical to the ChatGPT prompt, but the latter included additional information on a specific clinic experience to base the ChatGPT-generated reflective script on (since the undergraduate students based their reflections on actual patients that they encountered whilst on clinical placement). For the postgraduate extended essay, the wording of the ChatGPT prompt was identical to the instructions that students assigned this specific essay topic received.

Consent was captured before participants were permitted to proceed to the next part of the survey, where they considered the various scripts provided. After participants read each script, they were asked three questions. Initially, they were asked to grade each script on the basis of writing style as well as in terms of content. Four grading options were provided: Excellent, Good, Adequate, and Poor. Participants were then asked whether they suspected the script was written by a student, generated using ChatGPT-3.5, or whether they didn't know either way. An open-ended, free text item was also included asking participants to provide a brief rationale as to why they may have felt the script was authored by a student or generated using ChatGPT-3.5. A debrief was provided at the end of the survey. Ethical approval was sought at provided by the School of Medicine Research Ethics Committee (SMREC 23/38).

Data analysis

As quantitative and qualitative data were collated using the online survey, a mixed methods cross-sectional study design was deemed appropriate. A Chi-square test was run in IBM SPSS (version 27), given the non-parametric, categorical nature of the quantitative data collated. The open-ended qualitative data collated from the free text items were analysed using content analysis; all written responses provided by participants were carefully reviewed. Content analysis has been identified as being well-suited to research in qualitative healthcare education (e.g., Downe-Wamboldt, 1992; Hassoulas et al., 2023).

Results

Analysis of quantitative data

Participants rated each script on the basis of writing style and script content on a four-point scale from Excellent to Poor. They were also asked to identify the author of each script as either human, a chatbot or to declare whether they were uncertain as to script authorship. Overall, for the undergraduate reflective writing scripts, 50% of participants correctly identified the two student submissions, whilst only 23% correctly identified the ChatGPT script. In addition, 59% of participants incorrectly attributed authorship of the student submissions to ChatGPT. This suggests that a larger proportion of markers attributed authorship of the student scripts to the generative AI tool. This further highlights the difficulty that even experienced markers may experience in differentiating between scripts that are authored by students and those prepared using such generative AI tools (see Table 1).

Table 1: Undergraduate marker responses in differentiating student reflective submissions from ChatGPT-3.5 scripts.

Actual author	Marker assessment of likely author		
	Student	ChatGPT	Uncertain
Student	50	22	28
ChatGPT	59	23	18

A similar picture emerged for the postgraduate extended essay scripts, with 50% of markers correctly identifying the two student submissions once again but with only 19% correctly identifying the ChatGPT script. The degree of uncertainty in identifying authorship, however, was higher for the postgraduate markers than those who marked the undergraduate scripts. Specifically, 37% of participants who considered the postgraduate scripts were uncertain as to whether the ChatGPT script was written by a human or by the chatbot, as compared to just 18% of participants who considered the undergraduate scripts being uncertain as to the authorship of the ChatGPT script (see Table 2).

Table 2: Postgraduate marker responses in differentiating student extended essay submissions from ChatGPT-3.5 scripts.

Actual author	Marker assessment of likely author		
	Student	ChatGPT	Uncertain
Student	50	31	19
ChatGPT	44	19	37

Categorical data collated on participants' assessment of script writing style and content were analysed using chi-square tests. There was a significant effect of author identification for content specifically, $X^2(4, n=34) = 10.41$, $p<0.05$, but not for writing style ($p>0.05$). Interesting-ly, participants graded the undergraduate reflective writing submissions slightly lower on content than they did the ChatGPT script, whilst postgraduate extended essay student content was marked higher in comparison to the content generated by ChatGPT.

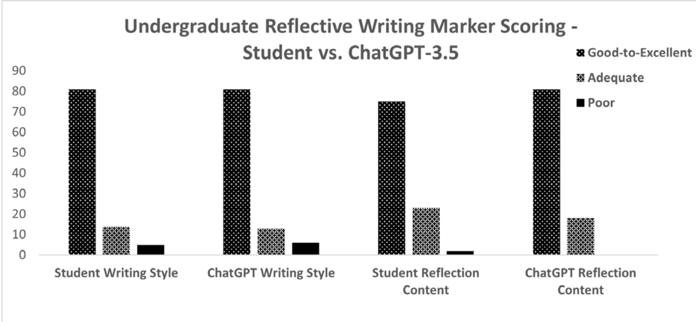


Figure 1. Marker assessment of undergraduate student submissions and the ChatGPT scripts.

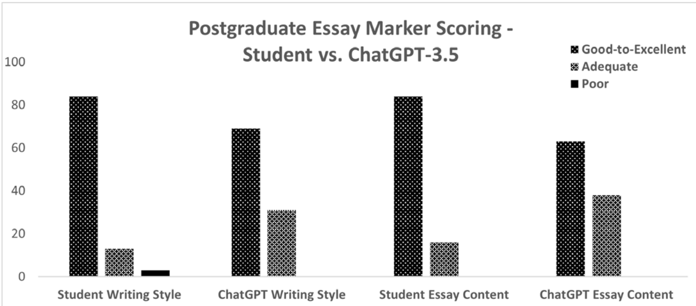


Figure 2. Marker assessment of postgraduate student submissions and the ChatGPT script.

These findings suggest that whilst writing style was statistically indistinguishable between human scripts and ChatGPT texts, script content does appear to hold certain clues as to how generative AI performed on this specific domain and whether experienced markers are able to identify clues to authorship in coursework content (see Figures 1 and 2).

Content analysis of qualitative data

Free-text responses by participants to the open-ended items included in the survey were considered in relation to the re-occurrence of key terms. As such, content analysis was performed to gain granular insight into what markers identified as key features that influenced their responses. Four key themes emerged, with the use of language (including grammar, phrases and expressions, and syntax) accounting for more than half of all free-text responses (see Table 3).

Table 3. Content analysis frequency table of key themes identified by markers for undergraduate reflective writing scripts prepared by students.

	Use of Language	Personal & reflective	Structure and writing style	Referencing
Total Frequency (%)	56	24	15	5
Student Identified	29	12	7	2.5
ChatGPT Identified	17	7	2.5	0
Authorship Uncertain	10	5	5	2.5

In relation to the ChatGPT-3.5 constructed script, markers once again identified the use of language as a key factor influencing their decision as to whether the script was written by a human or the chatbot. The same proportion of markers who alluded to the use of language in their responses had identified the author of this script as being human, too, suggesting an inability to accurately and confidently distinguish between student-specific language and proficiency versus the language being produced by the chatbot in response to the instructions provided. Furthermore, the ChatGPT-3.5 script, in particular, revealed that fewer markers suspected the use of language within the script to be suggestive of generative AI. A larger proportion of markers, however, emphasised that they found it difficult to identify the author as being human or a chatbot based solely on inconsistencies detected in the use of language (see Table 4).

Table 4. Content analysis frequency table of key themes identified by markers for undergraduate reflective writing scripts prepared by ChatGPT-3.5.

	Use of Language	Personal vs. formulaic	Structure and writing style	Referencing
Total Frequency (%)	46	32	18	4
Student Identified	28	9	18	0
ChatGPT Identified	4	14	0	0
Authorship Uncertain	14	9	0	4

Structure and writing style were a theme identified by markers in the context of the ChatGPT-generated script as well, although all those who made reference to structure and style of writing incorrectly identified the author as being a student. This suggests that the structure of the script and style of writing deceived markers regarding the identity of the author, with a large degree of certainty, as being human. As such, generative AI may be beneficial to students as a tool to improve the structure and style of academic writing.

Markers once again identified personalised writing as a key theme. However, more reference was also made to the reflection appearing more "formulaic" in the ChatGPT script. The largest proportion of markers referring to clues identified in terms of the personal and reflective nature of the writing considered the script to have been chatbot-generated. This suggests that in the context of reflective writing, generative AI tools are yet to master the ability to deceive markers specifically in relation to the depth of reflective practice demonstrated.

Inconsistency with regard to referencing, and sources cited for which no actual reference could be located, were identified as a key theme influencing markers' suspicions as to the authorship of the respective script. For the reflective writing scripts written by students, none of the markers who alluded to referencing identified the author of the scripts as being the chat-bot, whilst for the ChatGPT-generated script, markers responded with greater uncertainty regarding authorship but did not confidently identify the script as being authored by a student. This suggests that currently, citations and referencing may hold clues as to the authorship of scripts.

Regarding the postgraduate extended essays, markers once again identified the use of lan-guage as being a key factor in considering authorship, particularly in the context of the ChatGPT-generated script (see Table 4) as opposed to the student scripts (see Table 3), where the use of language was the second most common theme. Whilst almost half (47%) of post-graduate markers referred to the use of language in the context of the ChatGPT-generated script, no marker suspected the language used as being suggestive of generative AI use, with 33% suspecting the author of being a student whilst 14% reported that they were uncertain as to the authorship of the script. This is in contrast to the student-written scripts, where those who made reference to the use of language mostly identified the scripts as being written by students, with a lower degree of uncertainty regarding authorship overall.

The structure and layout of the extended postgraduate essays were identified as the most frequent theme referred to by markers when considering the student-written script, with the majority also correctly identifying the authors of the scripts as being human. This particular theme was only the third most frequently referred to by the same group of markers in considering the ChatGPT-generated script, with no markers, however, correctly identifying the author of that particular script as being the chatbot. Whilst themes such as the use of language as well as structure and layout were commonly referred to by markers, inaccuracy in differentiating between human and ChatGPT scripts remained an issue.

Knowledge and appraisal of the literature was a key theme identified in postgraduate markers. However, as with the structure and layout theme, inaccuracy in differentiating between student scripts and AI-generated scripts was problematic on this basis too. It was in relation to citations and referencing once again (as with the undergraduate reflective writing scripts) where differentiating between student-written and ChatGPT-generated scripts did appear to yield more promising results in accurately identifying authorship. Whilst only the fourth most frequently considered theme, no markers who alluded to referencing in relation to the ChatGPT script suspected student involvement. Equally, for the student-written essays, the majority who alluded to referencing suspected that the scripts had been written by students (see Table 5).

Table 5. Content analysis frequency table of key themes identified by markers for postgraduate extended essay scripts prepared by students.

	Structure and Layout	Use of Language	Knowledge & Appraisal	Referencing	Construction and style
Total Frequency (%)	30	19	19	19	14
Student Identified	19	11	11	11	11
ChatGPT Identified	5.5	3	8	3	3
Authorship Uncertain	5.5	5	0	5	0

An additional theme identified by markers in the context of the student-written scripts was that of construction and style, with the majority of markers considering this particular theme correctly identifying the author of the scripts as being human (see Table 3). The same cohort of markers did not refer to construction and style, however, in the context of the ChatGPT-generated script (see Table 6).

Table 6. Content analysis frequency table of key themes identified by markers for postgraduate extended essay scripts prepared by ChatGPT-3.5.

	Use of Language	Knowledge & Appraisal	Structure and Layout	Referencing
Total Frequency (%)	47	27	13	13
Student Identified	33	20	13	0
ChatGPT Identified	0	0	0	6.5
Authorship Uncertain	14	7	0	6.5

Discussion and conclusion

Our results suggest that experienced markers are currently unable to consistently differentiate between student-written scripts and text generated by natural language processing tools, such as ChatGPT. This appears to be the case for both undergraduate reflective writing tasks as well as postgraduate extended essays that form respective key components in undergraduate and postgraduate medical curricula. Whilst a significant effect of content on suspected authorship of the scripts was revealed, further analysis of the free-text qualitative data collated revealed that marker uncertainty, and even inaccuracy, in terms of which script was AI generated highlights the key difficulty that universities will face.

Whilst the application of this technology appears to be incredibly far-reaching, even in the medical sphere, from optimising clinical decision making (Liu et al., 2023) to scientific writing (Salvagno et al., 2023) as well as healthcare education and training in general (Hosseini et al., 2023), there is currently no study to our knowledge investigating human marker accuracy in differentiating between student-written scripts and generative AI produced text. Tools such as DetectGPT claim to detect the use of generative AI (on the basis of five open-source LLMs) with a 95% accuracy (Mitchell et al., 2023). These, however, remain under development and review and, as such, provide little current technological support for markers of modular coursework submissions. In an academic world rife with appeals, it is unlikely that less than 100% accuracy will be acceptable to universities, but given the stochastic nature of LLMs, this is likely to remain unachievable.

An assessment of ChatGPT's ability to accurately generate responses to complex medical que-ries has been reported by Johnson et al. (2023). There have been limitations reported, though, in regard to the robustness and reliability of using such tools in their present form in a clinical setting. Once again, whilst the outputs produced by ChatGPT may seem impressive, it is important to keep in mind that the tool currently makes use of a sophisticated model in responding to instructions and learning from its own prior responses. It is, therefore, important to healthcare professionals, students, and patients alike to continue to consult reliable sources in confirming information generated by such LLM tools. Such tools may be beguiling but also carry risks in terms of questionable source data and the perpetuation of dominant stereotypes (Bender et al., 2021). Even so, it appears likely that such tools may, over time, enhance the way in which we work, study and share information but should not be seen as accurate or reliable substitutes for human appraisal and reasoning influenced by evidence-based practice. Our findings confirm that, despite markers suspecting the use of tools such as ChatGPT at times, their suspicions were not proven to be valid on most occasions. The exception appears to be in relation to some aspects of content creation and particularly in terms of referencing, where markers were most accurate at differentiating between student-written and chatbot-generated scripts on this basis. Subsequent versions of ChatGPT as well as other LLMs such as Google's Bard, which will serve as the powerful search engine's direct interface, will undoubtedly aim to address key flaws identified in earlier versions of open-source generative AI tools (Rudolph et al., 2023b).

Given the limited accuracy demonstrated by experienced markers in differentiating between student-written scripts and those prepared by LLM tools such as ChatGPT, it would appear to be imperative that higher education assessment strategies be reconsidered to adapt to the increasing presence of such tools (Ifelebuegu, 2023). As we embark on an era where generative AI will be interwoven with, and embedded into, the student learning experience and possibly teaching provision, it is crucial that faculty work with students as partners in negotiating the responsible use of such new innovations. Knee-jerk reactions, such as the outright banning of ChatGPT that we have seen by some universities, will achieve little and will likely prove unrealistic, given the reach and implications of this technology. Furthermore, students will likely be engaging with these new technologies in the workplace. Our duty as educators has always been to ensure that students are equipped with the necessary skills to join the workforce. This now extends to the responsible use of new and emerging generative AI technologies.

Establishing trust between students and faculty, and re-evaluating what constitutes academic misconduct in light of the revolutionary shift in information creation and dissemination, should form the cornerstone of any initial response to this technology (Mills et al., 2023). Providing clear guidance to students as to what constitutes academic misconduct in relation to the misuse of generative AI is key. Such guidance will need to align with teaching information literacy, incorporating generative AI and the appropriate use thereof. Specifically, students should demonstrate an awareness of how such LLMs generate outputs, what the advantages are of using such platforms, as well as what the limitations are of this technology (Rasul et al., 2023). Support on how to critically appraise responses generated by generative AI platforms forms a crucial part of such training, ensuring the responsible integration of these technologies in our ever-expanding toolbox of resources at our students' disposal. As such, students should be encouraged to embrace new and emerging technologies but receive the necessary training on how to appropriately apply outputs of prompts to their scholarly practices without demonstrating an overreliance on this single source of information or passing responses off as their own.

Proactive management of expectations (both student and staff) is recommended. As opposed to such generative AI tools being simply viewed as a threat, it would be preferable to instead consider how such tools can be embraced appropriately. Where transgressions of professional boundaries do occur, however, academic misconduct procedures should be updated to reflect what is considered appropriate and what is an inappropriate use of this emerging technology (Mohammadkarimi, 2023). It is no easier to ban the use of generative AI at this stage than it would have been to stop the internet from going mainstream three decades ago. Negotiating our relationships with these new tools and how they can enhance various aspects of our lives is key, without abuse of this new technology limiting our own personal and professional development.

References

Behrens, S. J., Gomez, A., & Krimgold, J. (2019). Spelling counts: The educational gatekeeping role of grading rubrics and spell check programs. *Research and Teaching in Developmental Education*, 7-20.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on fairness, accountability, and transparency (FAccT '21),* (pp. 610–623). https://doi.org/10.1145/3442188.3445922

Bishop, L. (2023). *A computer wrote this paper: What ChatGPT means for education, research, and writing.* SSRN, 4338981, 10.2139/ssrn.4338981

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 60,* 1-12. doi: 10.1080/14703297.2023.2190148

Downe-Wamboldt, B. (1992). Content analysis: Method, applications, and issues. *Health Care for Women International, 13*(3), 313-321. doi: 10.1080/07399339209516006

Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. & Hatta, H. (2023). Analysing the role of ChatGPT in improving student productivity in higher education. *Journal on Education 5*(4), 14886-14891. doi: 10.31004/joe.v5i4.2563

Hassoulas, A., de Almeida, A., West, H., Abdelrazek, M., & Coffey, M. J. (2023). Developing a personalised, evidence-based and inclusive learning (PEBIL) model of blended learning: A cross-sectional survey. *Education and Information Technologies*,1-18. doi: 10.1007/s10639-023-11770-0

Hosseini, M., Gao, C. A., Liebovitz, D. M., Carvalho, A. M., Ahmad, F. S., Luo, Y., MacDon-ald, N., Holmes, K. L., & Kho, A. (2023). An exploratory survey about using ChatGPT in edu-cation, healthcare, and research. *medRxiv*: 2023.03.31.23287979. https://doi.org/10.1101/2023.03.31.23287979

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://journals.sfu.ca/jalt/index.php/jalt/article/view/875

Johnson, D., Goodman, R., Patrinely, J., Stone, C., Zimmerman, E., Donald, R., Chang, S., Berkowitz, S., Finn, A., Jahangir, E., Scoville, E., Reese, T., Friedman, D., Bastarache, J., van der Heijden, Y., Wright, J., Carter, N., Alexander, M., Choe, J., Chastain, C., … Wheless, L. (2023). Assessing the accuracy and reliability of AI-generated medical responses: An evaluation of the ChatGPT model. *Research Square,* rs.3.rs-2566942. https://doi.org/10.21203/rs.3.rs-2566942/v1

Liu, S., Wright, A. P., Patterson, B. L., Wanderer, J. P., Turer, R. W., Nelson, S. D., McCoy, A. B., Sittig, D. F., & Wright, A. (2023). Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *Journal of the American Medical Informatics Association: JAMIA, 30*(7), 1237–1245. https://doi.org/10.1093/jamia/ocad072

Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1), 16-30. https://journals.sfu.ca/jalt/index.php/jalt/article/view/843

Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writing in the era of artificial intelligence. *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://doi.org/10.37074/jalt.2023.6.2.10

OpenAI. (2023). *ChatGPT.* https://chat.openai.com/chat

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://journals.sfu.ca/jalt/index.php/jalt/article/view/787

Rudolph, J., Tan., S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342-363. doi: 10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Critical Care, 27*(1), doi: 10.1186/s13054-023-04380-2

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : **http://journals.sfu.ca/jalt/index.php/jalt/index**

Awareness and acceptance of ChatGPT as a generative conversational AI for transforming education by Ghanaian academics: A two-phase study

Michael Agyemang Adarkwah[A] — [A] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Samuel Amponsah[B] — [B] *University of Ghana, Ghana*

Micheal M van Wyk[C] — [C] *University of South Africa, South Africa*

Ronghuai Huang[D] — [D] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Ahmed Tlili[E] — [E] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Boulus Shehata[F] — [F] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Ahmed Hosny Saleh Metwally[G] — [G] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Huanhuan Wang[H] — [H] *Smart Learning Institute of Beijing Normal University, Beijing, China*

Keywords

Artificial intelligence (AI);
chatbots;
ChatGPT;
conversational agents;
diffusion of innovation;
technology awareness.

Correspondence

adarkwahmichael1@gmail.com [A]

Article Info

Received 15 August 2023
Received in revised form 8 September 2023
Accepted 8 September 2023
Available online 11 September 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.26

Abstract

By increasing technology trend awareness, individuals can leverage novel and ground-breaking technologies to complete mundane activities and buy time to focus on other projects. This article presents an overview of why there is a slow pace of digital transformation in education in Ghana using ChatGPT (an advanced chatbot) as a case scenario. In this two-phase study, which used a triangulation approach (an exploratory sequential design), we found that most of the authors of publications about ChatGPT were not from the African continent or were affiliated with international institutions. A thematic analysis of interview data involving 34 academics in Ghana about ChatGPT revealed that most academics had limited knowledge about ChatGPT and artificial intelligence-powered chatbots. The main themes generated comprised the purpose of ChatGPT and chatbots, their usability and accuracy, and ChatGPT and artificial intelligence (AI) enthusiasm. The quantitative phase of the study surveyed the views of 50 academics who confirmed the minimal awareness of ChatGPT by Ghanaian academics. There were mixed views about the relevance and usefulness of ChatGPT in work-related tasks. Following the findings, we provide ways to create technology trend awareness for academics from African countries like Ghana to transition from being "laggards" to "early adopters", as explained by Rogers' diffusion of innovation theory. The findings call for policymakers and educators to promote technological awareness.

Introduction

One of the tragedies in life is people's ignorance of available tools and devices that have the potential to transform their lives. This situation becomes more compelling when the resources are available, but people's ignorance prevents them from accessing or maximizing their use – this situation resonates with Richard Auty's (1994) resource curse theory, also termed the paradox of plenty (Auty, 2007). This unfortunate situation befell Ghana before the arrival of the colonial masters when gold was so abundant that the citizenry could not recognise its value at the time. That part of history is gone; gold and other precious minerals are scarce and of high value. However, one wonders if Ghanaians and people from other developing economies value the 'modern gold' (technology) as they should.

Undoubtedly, disruptive technologies are shaping education like never before in this intelligent era (Adarkwah & Huang, 2023). Mundane administrative, teaching and learning activities are easily catered for and facilitated through various innovative technologies, such as AI, robotics, learning analytics, blockchain, etc. With all these technological developments, many developing countries, especially those in the African region, such as Ghana, are still lagging behind the rapid utilization of advanced technologies in education (Adarkwah & Huang, 2023). Ghana launched its ICT (information and communication technology) for Accelerated Development (ICT4AD) in 2003 with the objective of transforming Ghana into an information and technology-driven high-income economy through education (Adarkwah, 2021; Adarkwah & Huang, 2023). Although the policy is almost at the end of its lifespan, Ghana is still taking baby steps towards digital transformation in education. Education is supposed to work smarter by co-evolving with technology, but there has been slow progress towards the digital transformation of education in Ghana (Adarkwah, 2021; Agyei & Voogt, 2012). One of the curtailing factors to the digital transformation of education in the country is the late recognition of new technologies and their impact on everyday life, as well as the non-readiness to use such technologies for educational purposes (International Finance Corporation, 2019).

Given the preceding, we are interested in the technology awareness of Ghanaian academics about a newly launched technology called ChatGPT (Generative Pre-trained Transformer), which is currently making waves in advanced countries in terms of research and educational practice. The emergence of ChatGPT has sparked many discussions on how it can augment or transform education. For instance, Santandreu et al. (2023) have tested the ability of chatbots to provide personalized support, a real-time interactive platform and immediate constructive feedback. Hassoulas et al. (2023) describe AI's enormous benefits to higher education as a watershed moment. On the other hand, some authors have identified the drawbacks of ChatGPT, such as limited context understanding, its inability to incorporate visuals, potential errors in the solutions it provides, its ability to enable large scale cheating in assessments by students, and inaccessibility due to cost implications (Adarkwah, 2021, Santandreu et al., 2023).

By way of context, ChatGPT is a conversational agent based on natural language processing (NLP) which engages users in a human-like conversation (Tlili et al., 2023a). According to OpenAI (2023), ChatGPT "answers follow-up questions, admits its mistakes, challenges incorrect premises, and rejects inappropriate requests." The GPT model has been evolving to reach GPT-4, which is the latest update, released in March 2023 (Rudolph et al., 2023b). The update in the GPT models seeks to enhance capabilities, and provide more fine-tuning, a larger dataset, and more human-like text generations. It can create new things and allow for more realistic natural dialogues (Santandreu et al., 2023).

Although research about ChatGPT is in its infancy, it is interesting to note that Rudolph et al. (2023b) trace the evolution of chatbots to the past 57 years. Besides, several blogs and news media continue to share the perception of various stakeholders (academics, administrators, policymakers, etc.) towards ChatGPT, especially from the early adopters (e.g. Mogavi et al., 2023). However, most of these blogs and news media are either Western-based or developed countries and regions. The paucity of conversation on ChatGPT from the African region in terms of research and news media motivated us to investigate the level of awareness of ChatGPT among academics in the African region nearly three months after its release. Pandey et al. (2021, p. 2) state that technology trend awareness is "the skill of an individual to be aware and mindful of new and popular technology that has been gaining widespread acceptance across concerned industries or markets". In this light, there is a need for educational institutions in the African region, such as those in Ghana, to be aware of the ground-breaking technological innovations like ChatGPT to revolutionize education in terms of policy and best practices in the use of the conversational agent and other emerging technologies.

It is worth adding that the awareness of new technologies has been underscored by several scholars to foster positive attitudes among users towards the rapid adoption of technology (Dinev & Hu, 2007; Pandey et al., 2021). Carpenter et al. (2022) argued that it is necessary for academics to be aware of innovative technologies to be able to implement innovative pedagogies in the classroom. In this twenty-first century, technologies, such as ChatGPT, which has transformative potential in education (Tlili et al., 2023a), need to be leveraged to equip learners with transferable skills needed in the labor market (UNESCO, 2021). Although scholars have debated the positive and negative disruptive potential of ChatGPT, scholars assert that this signals a paradigm shift in the educational landscape and other aspects of life (e.g. Mills et al., 2023; Popenici, 2023; Tlili et al., 2023a). In this light, this study seeks to answer the following research questions:

RQ1. What are academics' awareness and understanding of ChatGPT as an AI conversational pedagogical tool?

RQ2. What are the experiences of academics who have used ChatGPT as an AI conversational pedagogical tool in their teaching or research?

RQ3. What is the level of acceptance of ChatGPT as an AI pedagogical tool among academics?

Ghana is selected as the study context to investigate how the delay in the digital transformation might impact academics' awareness of valuable technologies in education, hence missing opportunities for revolutionizing education in Africa. The findings of the study can further provide insights to various stakeholders about the importance of raising awareness among African academics related to technological innovation to ensure quality education, which dovetails with the Sustainable Development Goals (SDG 4) of the United Nations.

Theoretical underpinning

We based our argument within the confines of the diffusion of innovation theory and ChatGPT as an artificial intelligence tool to foreground our study within the Ghanaian institutions of higher learning. In this study, we propose that Rogers' diffusion of innovation theory (DIT) (2003) is a commonly used change model for implementing technological innovation. Specifically, we will apply this theory to the introduction of ChatGPT, an innovative AI language transformer application. DIT refers to the processes that occur as people adopt a new idea, product, philosophy, or practice (Dearing & Cox, 2018). The model explains the likelihood of an individual/people adopting new technology, as is the case of hype around ChatGPT. According to Rogers (2003), there are five groups of adopters when a new technology emerges: innovators, early adopters, early majority, late majority, and laggards. There are also five stages at the individual level where the diffusion of innovation occurs: (1) knowledge awareness stage (an individual is exposed to innovation but lacks complete information about it); (2) persuasion or interest stage (an individual becomes interested in the new idea and makes additional inquiries about it); (3) decision or evaluation stage (an individual mentally applies innovation to his present and anticipated future situation and makes a decision to try it or not to try it); (4) implementation or trial stage (an individual makes full use of innovation) stage; and (5) confirmation or adoption stage (an individual decides to continue the full use of innovation) (Rogers, 2003). The focus of this study is on the first stage, knowledge awareness. We put forward that in the advent of an innovation/new technology, such as ChatGPT, most educators/academics in African countries like Ghana lack the awareness of the innovation or might have heard of it but lack complete information about its use (Adarkwah, 2021; Van Wyk et al., 2023). As a result, they are often not early adopters, as witnessed in developed countries where users (academics or learners) interact with advanced or immersive technologies. Most academics in Ghana and countries with similar contexts are often perceived as 'laggards'.

Related works on ChatGPT

Since November 2022, the hype around the phenomenon of ChatGPT has generated and exponentially accelerated the use of and adoption in education (Santandreu et al., 2023). Tlili et al. (2023a) conducted a three-stage sentiment analysis of the concerns about using ChatGPT in education. After an analysis of 2,330 tweets from 1,530 Twitter users from December 23, 2022, to January 6, 2023, it was revealed that the public generally has a positive perception of ChatGPT and its use in educational settings (Tlili et al., 2023a). Moreover, qualitative interviews involving early adopters indicated that a cautious approach has to be taken in adopting ChatGPT in education due to issues relating to how it will transform education, response quality, perceived usefulness, personality and emotions, and ethical issues. Moreover, an analysis of user experience also highlighted issues, such as academic integrity/cheating, the accuracy of prompts, fairness in the provision of contents, privacy issues, and concerns about the manipulation of the output of information to users (Cotton et al., 2023; Tlili et al., 2023a; Vaishya et al., 2023).

Furthermore, scholars have argued that ChatGPT significantly supports learning and teaching across different levels of education (Alshurafat, 2023; Baidoo-Anu & Owusu Ansah, 2023; Khan et al., 2023; Tlili et al., 2023a; Zhai, 2022). In particular, students with disabilities could use this tool to increase their reading, writing, problem-solving, communicative skills, and language skills (Kasneci et al., 2023). These benefits also supported the professional development of teachers, academics and managers in writing reports, managing projects and supporting continuous professional initiatives (Amponsah & Bekele). Based on the latter, Rospigliosi (2023) questioned the use of ChatGPT as a tool for teaching and postulated that it could significantly transform the way we teach. Rospigliosi (2023) made particular reference to lesson planning, the creation of personalized learning experiences, assessment, and professional development.

Kasneci et al. (2023) also wrote a position paper about the educational opportunities and challenges of large language models such as ChatGPT for education. Their study revealed that ChatGPT provides educational resources for different types and levels of learners, promotes group and remote learning, empowers learners with disabilities, assists in professional training programs, and also presents opportunities for teaching (i.e. personalized learning, lesson planning, assessment and evaluation, etc.). On the same note, Mogavi et al. (2023) conducted a qualitative content analysis of four major social media platforms (Twitter, Reddit, YouTube, and LinkedIn) to identify the user experience (UX) and perspectives of early adopters toward ChatGPT. Their study found that ChatGPT is mostly used in the contexts of higher education, K-12 education, and practical-skills learning (Mogavi et al., 2023). Also, some early adopters tend to consider ChatGPT as a revolutionary technology to facilitate students' self-efficacy and motivation to learn. In terms of challenges, Mogavi et al. (2023) and Tlili et al. (2023a) proposed that there are certain competencies and literacies teachers and learners need to develop in the use of large language models like ChatGPT while users need to be aware of the limitations and drawbacks of using such technology (Rudolph et al., 2023a).

Moreover, Zhai (2022) conducted a pilot test to gauge the efficacy of ChatGPT in writing a research paper titled "Artificial Intelligence for Education". The findings revealed that ChatGPT can aid scholars in constructing their research papers in a coherent, informative, systematic, and accurate manner. According to Zhai (2022), users do not need to have expert knowledge about the subject and the chatbot was able to provide an extremely efficient write-up within two to three hours. Zhai (2022) recommended that educators should search for ways to use ChatGPT and other AI tools to foster creativity and critical thinking in students rather than focusing on general skills. Also, because of the possibility of relying on ChatGPT for assessment tasks, new formats of assessment might be needed (Hassoulas et al., 2023; Zhai, 2022).

In a similar vein, Susnjak (2022) questioned whether ChatGPT could mark the end of academic integrity. After an evaluation to examine the ability of ChatGPT to perform high-level cognitive tasks in a human-like manner (Susnjak, 2022). Similar studies also observed that ChatGPT is able to exhibit critical thinking skills and generate highly realistic prompts with minimal input, which poses a threat to academic integrity, especially online exams in tertiary institutions (Van Wyk et al., 2023; Santandreu et al., 2023). To mitigate high levels of academic dishonesty, Chaka (2023) recommends traditional forms of assessment, such as oral examinations and the use of AI detectors.

Similarly, studies by Cotton et al. (2023) and Tsigaris and Teixeira da Silva (2023) raised the issues of academic integrity and educational opportunities offered by ChatGPT and found that it can result in plagiarism among students. It was also revealed that ChatGPT offers an undue advantage to students who can access the advanced and paid versions over those who cannot. Besides, it presents difficulty to academic staff to determine whether texts are AI-generated or human-generated (Hassoulas et al., 2023; Hosseini et al., 2023; Hu, 2023; Tang, 2023; Alshurafat, 2023). At the same time, it has been observed that ChatGPT can increase student engagement through asynchronous communication, help teachers customize exam questions or evaluations, help create an interactive and game-based assessment, and can also be used for grading students' assignments or providing feedback (Baidoo-Anu & Owusu Ansah, 2023). However, the researchers found issues with privacy, accuracy, and the possibility of biases in data training as some of the limitations found with the chatbot (Kooli, 2023; Rasul et al., 2023).

Furthermore, Gilson et al. (2022) evaluated the performance of ChatGPT using the United States Medical Licensing Examination Step 1 and Step 2 exams and further analyzed user interpretability based on the responses from ChatGPT. In comparison to earlier NLP models, ChatGPT was found to be more advanced. Its performance exceeded the 60% threshold on the National Board of Medical Examiners (NBME), which implies a pass rate for a third-year medical student (Gilson et al., 2022). The answers ChatGPT provided were found to be logical across multiple answers. Also, Kung et al. (2023) assessed the performance of ChatGPT in a United States Medical Licensing Exam (USMLE) and observed that it has the potential to assist with medical education and clinical decision-making as it performed at or near the passing score for all three exams in USMLE. It is imperative to add that GPT-4 produces even more advanced functions which are semantically richer, takes into account contextual factors, and generates more realistic human-like dialogue (Santandreu et al., 2023; Rudolph et al., 2023b).

Frieder et al. (2023) assessed the mathematical capabilities of ChatGPT by testing it on large datasets and how it can assist professional mathematicians with routine tasks that come with their work. They observed that ChatGPT's ability to solve mathematical problems is significantly below that of an average mathematics student. According to Frieder et al. (2023), ChatGPT, in most cases, is able to understand mathematical questions but often fails to produce the right responses. This was recognized as a serious setback to ChatGPT in performing an educational task. However, the latest version (GPT-4) has plugins with improved capabilities to perform mathematical functions. It can now solve statistical, arithmetic and mathematical problems with precision and speed (Abramski et al., 2023; Santandreu et al., 2023).

Various concerns, on the other hand, were reported about ChatGPT in education, including tracking academic dishonesty and cheating. For instance, Aydın and Karaarslan (2022) expressed concerns about using ChatGPT to generate literature reviews or abstracts on a given topic. For instance, the Ithenticate software was used to check the plagiarism of the output by ChatGPTand it was found that ChatGPT is capable of helping in the academic publishing process with minimal human effort. This raises concerns about the scientific integrity of the written research outcomes. At the same time, the authors observed that ChatGPT is unable to produce original texts after paraphrasing (the Ithenticate software showed a 40% similarity index when texts created by authors and ChatGPT were assessed together). Besides, a commentary paper by Saliba and Boittsios (2023) reported that this language transformer tool could be the "death knell" in academic publishing because it could create "cheating and academic fraud" on a massive scale, which ultimately impacts scholarly creativity, innovative writing and intellectual property rights. It, therefore, comes with no surprise that out of 34 expert markers recruited by Cardiff University in the UK, 23% could not distinguish between essays generated by undergraduate students and ChatGPT, while 19% could not do the same for graduate-level papers (Hassoulas, 2023).

Currently, there are a few published studies on ChatGPT from the African perspective and in higher education. These include studies by Chaka (2023) and Ifelebuegu (2023) from South African and Uganda, respectively. This depicts the relative dearth of research on this novel phenomenon from an African perspective. It warrants the need to contribute to the discussion on ChatGPT using the Ghanaian context and how it can transform educational practices. Furthermore, the fast evolution of the GPT models urges in-depth research on the awareness and acceptance of these models for transforming education. For instance, the evolution from ChatGPT (which is based on GPT-3.5) to GPT-4 has already provided better capabilities, more fine-tuning, and enhanced human-like text generations. It is able to allow for more realistic natural dialogues. This is why many researchers recommend investigating ChatGPT and its new models, such as GPT-4, in terms of the awareness and acceptance of these models (Rudolph et al., 2023b; Gimpel et al., 2023).

Method

To address the research questions, we employed the sequential exploratory mixed-method design, which involves collecting qualitative data and analyzing them first, followed by quantitative data (Hanson et al., 2005). In this design, quantitative data are used primarily to augment the qualitative data. The application of ChatGPT in education is an emerging phenomenon, and research about it has gained momentum across diverse fields. The sequential exploratory design was appropriate because it helped in exploring a phenomenon and in generalizing qualitative findings to a specific population (Hanson et al., 2005). At the same time, it enabled the researchers to build on the results of the qualitative phase with quantitative data and analysis (Creswell et al., 2006). For example, in order to understand human and AI chatbot relationships, Pentina et al. (2023) used an exploratory qualitative in-depth interview analysis and followed it with a survey-based confirmatory hypotheses-testing. In the current study, we sought to explore the academics' awareness of ChatGPT as an AI-conversational agent in education and investigate their acceptance and use of ChatGPT using a modified version of the technology acceptance model (TAM) survey. Hence, the study is organized into two phases, a qualitative and a quantitative one. In Study 1, the qualitative phase explores academics' exposure and a broad understanding of ChatGPT within three months of its launch (January 2023). Afterwards, the quantitative phase in Study 2 presents academics' use and acceptance of ChatGPT as an AI tool in education. Study 2 was conducted two months after the qualitative phase (March 2023) after many academics had interacted with ChatGPT.

Study 1 (qualitative phase)

Design

The narrative inquiry approach is used when detailed accounts or experiences from individuals are collected and chronologically ordered to ascertain the meanings of those experiences. The "narrative might be the term assigned to any text or discourse, or it might be text used within the context of a mode of inquiry in qualitative research" (Creswell et al., 2007, p. 240) and "narrative is understood as a spoken or written text giving an account of an event/action or series of events/actions, chronologically connected" (Czarniawska, 2004, p. 17). In the context of this study, the narrative refers to the experiences of the academics with their exposure to ChatGPT.

Participants

Academics included in the study were conveniently sampled because they were accessible during the time of the study and voluntarily opted to participate in the study. They included professors (n = 3), associate professors (n = 11), senior lecturers (n = 7) and lecturers (n = 13). The 34 academics whose views were solicited for this study were drawn from three higher education institutions in Ghana. For ethical reasons, no names were used. The selection of the institutions and study participants is purely based on convenience sampling.

Ethical considerations

Before we started the study, ethical considerations were obtained, and participants gave consent. Though data collection for this study was via semi-structured interviews, ethical considerations were put in place. The academics were informed about the use of the conversations for research purposes. In view of that, the data collected were used for no other purpose other than for the publication of this study. Also, participants' names were replaced with their designation and number during the discussion of themes (e.g., a professor who is the third participant in the study is represented as P3). All efforts were put in place to ensure that the quotations used for this study cannot be traced to the participants. Lastly, the academics were contacted close to the publication of this study to confirm the use of the information they had shared. This was also for the purposes of member checking as the participants agreed that the transcripts represented the views they shared during the interviews.

Data collection and instrumentation

A semi-structured interview guide was developed based on extant literature. For example, one key study that guided the construction of the interview guide was the study on techno trend awareness by Pandey et al. (2021). The interview guide was assessed by the authors for modification until a consensus was reached. All the semi-structured interviews for the purpose of data collection took place between January 1st and January 31st, 2023. The conversations were all in the English language. In eliciting data from the study participants, phone interviews with the participants using WhatsApp audio calls were conducted. Some of the key questions used to elicit data during the informal virtual conversations included: 'Tell me about your general knowledge of chatbots.' 'Have you heard about ChatGPT?' 'What has been your experience with ChatGPT?' (see Appendix 1). The duration for each interview was approximately 35 minutes.

When the data collection had ended, we performed a thematic analysis of the interview data to detect recurring words or themes in the conversations. According to Braun and Clarke (2006) and Nowell et al. (2017), this approach is flexible and well-structured in identifying extracts to generate themes manually. Though Holloway and Todres (2003) describe this manual process as time-consuming, it was helpful to us as we were able to identify the latent meanings of the information shared by the study participants. The generated themes were discussed among the authors of the study to agree on themes to be used for the study.

Data analysis

Two of the researchers manually coded the interview transcripts by extracting key information and assigning codes to them. Line-by-line open coding was used in generating the final themes of the study (Khandkar, 2009). To ensure trustworthiness, an inter-coder agreement was reached (i.e., the researchers used the interview data in a similar manner). The four main themes generated from the conversations were: (1) misrepresentation of the purpose of ChatGPT and chatbots, (2) lack of digital knowledge and technical skills and the usability of ChatGPT, (3) lack of exposure and advocacy (accuracy) of ChatGPT, and (4) ChatGPT and AI enthusiasm (see the next section).

Study 2 (quantitative phase)

Research design

A quantitative survey design was used to investigate the awareness and acceptance of ChatGPT as an AI conversational tool for education within five months of its launch. First, we wanted to assess whether academics' awareness of ChatGPT improved two months after the qualitative study. Furthermore, we wanted to assess the acceptance and use of ChatGPT by academics as a pedagogical tool. The survey helped in probing further into the day-to-day use of ChatGPT by academics (Creswell et al., 2007). The survey design enabled the researchers to administer a questionnaire to the study sample to examine their attitudes, perceptions, and behaviors about ChatGPT (Creswell, 2018).

Participants

Academics from the qualitative study and their colleagues formed the sample of the study. The academics in Study 1 were contacted to answer the questions on the survey and were encouraged to invite colleagues who were somewhat exposed to ChatGPT and were willing to participate in the study. That is, all the 34 academics in the qualitative study were re-selected and 16 academics were new recruits based on their willingness to join the study. A convenience sampling technique was used to recruit the academics included in the study because they were accessible during the time of the study and voluntarily opted to participate in the study. They were duly informed of their right to withdraw from the study at any time. In all, 50 academics responded to the survey. Out of the 50 participants who answered the survey, 23, representing 46%, were aged 45 and above (Figure 1). 43 (86%) were male, and seven (14%) were female (Figure 2). Twenty-six academics, constituting 56%, were from the field of education (Figure 3), 84% of them used ChatGPT-3.5, and 16% used the plus version (GPT-4) (Figure 4).

Figure 1-4 presents the demographic information of participants.

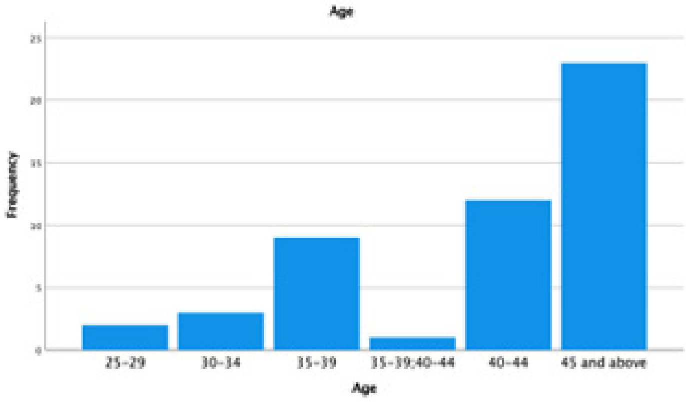


Figure 1: Age of academics

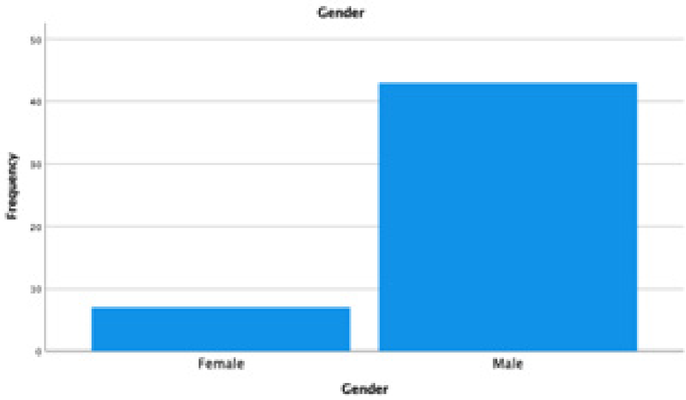


Figure 2: Gender of academics.

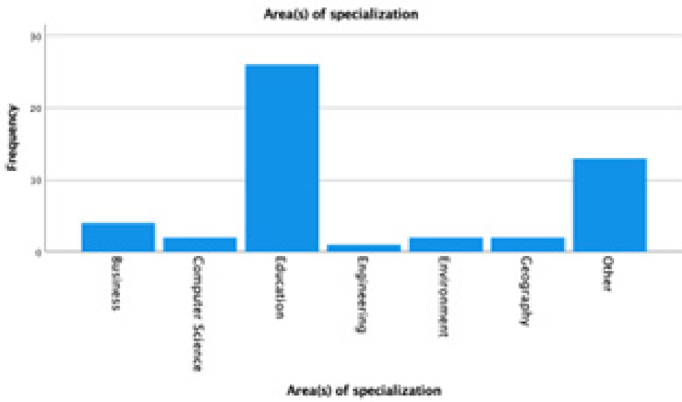


Figure 3: Area of specialization of academics.

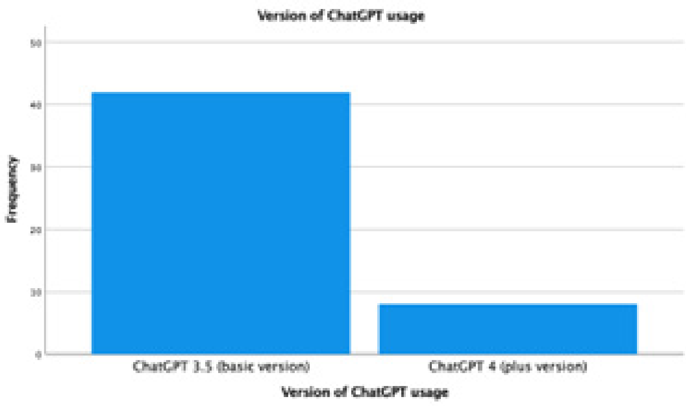


Figure 4: Version of ChatGPT used by academics.

Ethical considerations

Participants' informed consent was gained before distributing the online survey. We adhered to all ethical principles in the Declaration of Helsinki.

Data collection and instrumentation

Google Forms served as the online platform for the survey items. The survey questions consisted of two sections. Section A took the demographic information of the participants, and Section B measured the awareness and acceptance of ChatGPT by academics. The survey questions (Section B) were developed based on the Technology Acceptance Model (TAM) survey. We modified the items on the TAM survey to suit the context of this paper. Four items (i.e. job relevance, perceived usefulness, perceived ease of use, and perceived enjoyment) comprising fourteen questions (14) made up the final items on the survey. The final questions which were used for collecting data were agreed upon by the authors of the study. The internal consistency of the survey was also checked. A Cronbach's alpha value of .876 was obtained, indicating that the items on the survey were reliable for eliciting data from the study participants. All items are measured on a seven-point Likert scale (where 1 = strongly disagree; 2 = moderately disagree, 3 = somewhat disagree, 4 = neutral (neither disagree nor agree), 5 = somewhat agree, 6 = moderately agree, and 7 = strongly agree). Tables 1-4 present the ratings of participants in the study. A link to the online survey was distributed to academics from the three institutions in the country, and they were encouraged to invite colleagues to answer the questions in the survey. Overall, the data collection took a month to complete.

Data analysis

Descriptive statistics (mean, standard deviation, and percentages) were used in presenting the data. Additionally, correlational analysis was performed between the main variables (the five items) that made up the survey.

Qualitative results

Theme 1: Misrepresentation of the purpose of ChatGPT and Chatbots

A thematic analysis of the interview data revealed that as of January 31, 2023, the academics in this study lacked awareness of ChatGPT. Most of the participants in the study had not heard about ChatGPT. In some instances, the interviewer had to explain the meaning of a "chatbot", not to mention ChatGPT, to some of the academics. Although many academics might have interacted with chatbots before through a visit to a blog, an official website, or an online store, they did not know the term "chatbot" or what it stands for. A probe into the inquiry about ChatGPT and what chatbots, in general, stand for suggests that academics in this study perceived that their prior conversations with chatbots were with human operators. Two of the academics perceived ChatGPT at the time as a social media mobile application available on the App Store or Google Play Store. While an app version of ChatGPT might be in sight in the near future, it is currently web-based in nature. Some of the academics who had heard about ChatGPT did not pay attention to it, did not sign up, or were doubtful about creating an account.

> What are chatbots and what do they do? Personally, I do not know what ChatGPT is. If it is an app, I can download it from my Google Play Store and see what it is. (AP6)

> Is ChatGPT a new app? I might have heard about chatbots, but I have not taken the time to read about them to know their main purpose. But after our talk, I will check it out. (L7)

> Yeah, I have interacted with a chatbot before from a marketing store, but I thought I was talking to a service personnel or customer care. This is interesting. So, what is the main purpose of ChatGPT? (P3)

The narratives, thus far, depict a complete lack of knowledge and conceptual understanding of ChatGPT, an application that has, in a short space of time, impacted the educational and other sectors of human life. Gauging by Rogers' DIT (2003), this evidence puts the study participants in the laggards' group of adopters as against, for instance, Rudolph et al. (2023b) who interrogated and challenged the intelligence of chatbots. Our study was interested in how the knowledge gap impacted the technical skills and usability of the system among the participants. This is presented under the ensuing sub-theme.

Theme 2: Lack of digital knowledge and technical skills and the usability of ChatGPT

A majority of the academics inquired about how they could use ChatGPT. Questions were asked if it could be used offline or whether the Internet is needed to access it. Also, some of the academics inquired if they had to have some technical skills or should have prior knowledge about chatbots before they could effectively use ChatGPT. Questions were also asked about best practices in the use of ChatGPT in educational settings to avoid issues relating to plagiarism.

> Is it something I can use like Grammarly to enhance my work, and do I have to have some special skills to be able to use it? (L13)

> Right now, I do not know how useful it can be for me. But I will check it out and see how I can use it in my research and teaching. (P3)

Though the study participants may be categorized as laggards, one will not be wrong to identify them as people who have an interest in adopting technology. The challenges with data cost and digital gadgets in the study setting might be a factor working against technology adoption.

> Can I download the ChatGPT and use it offline? If I do not have mobile data or access to WIFI, how can I use it? (SL6)

Theme 3: Lack of exposure and advocacy (accuracy) of ChatGPT

Because most of the participants lacked knowledge of chatbots and generative conversational AI user experience, their primary source of information about ChatGPT was the interviewer who interacted with them. Questions were asked by some of the academics concerning the accuracy of ChatGPT. Sample conversations the first author or other Twitter users had with ChatGPT were sent to the academics for their perusal. As an early adopter of ChatGPT, the first author expounded on his experiences with ChatGPT to the academics.

> Oh, that is cool! But if it can give me responses to my questions, how do I know that the answers are correct? (L5)
>
> Because I have not used it before and I am just hearing about it, maybe, you can tell me some of your experiences when you asked ChatGPT a question. Were you satisfied with the quality of the responses it gave you? (SL3)
>
> I am interested to know more about ChatGPT and how it can help me in my work. But can the output from it be trusted, or do I have to do more investigations on my query after getting a response from ChatGPT? (AP, 1)

Theme 4: ChatGPT and AI enthusiasm

During the interviews, the researcher asked each of the academics to read about ChatGPT on Google. It was observed that the academics in the study expressed a high sense of enthusiasm about the potential of ChatGPT and what it could mean for education. The academics were in awe of how AI is shaping the educational landscape and other sectors of society. In one of the conversations, an academic texted that he was happy to know about ChatGPT and will introduce it to his students and sister who was studying in one of the tertiary institutions in Ghana.

> Wow, this is good. ChatGPT is able to give exact responses from the blogs I read. When I get to my office, I will try and register and start using it. Thank you! (SL3)
>
> I think ChatGPT will bring significant changes to how we educate students. I am very happy about how technology is driving education. We just need to use it for our benefit. (AP9)
>
> “I will try and introduce ChatGPT to my colleagues. I know they will be surprised by its abilities. After briefly searching about it, I am impressed. I think both teachers and students can benefit from it.” (P2)

Similar to the reflections under theme 2, the participants’ awareness level is low, which might be the key factor driving their late acceptance of the innovative AI-powered chatbots. However, that did not affect their enthusiasm for its future usage, a depiction of the second stage (persuasion or interest) of Rogers’ (2003) DIT.

Quantitative results

Most of the academics used a specific version of ChatGPT because of cost or relevance to their work (Figure 5). Webinars and social media platforms other than Facebook and Twitter were the main media of exposure to ChatGPT for a majority of the academics (Figure 6). Most academics spend one hour daily using ChatGPT (Figure 7). Most of the participants considered themselves basic users of ChatGPT (Figure 8).

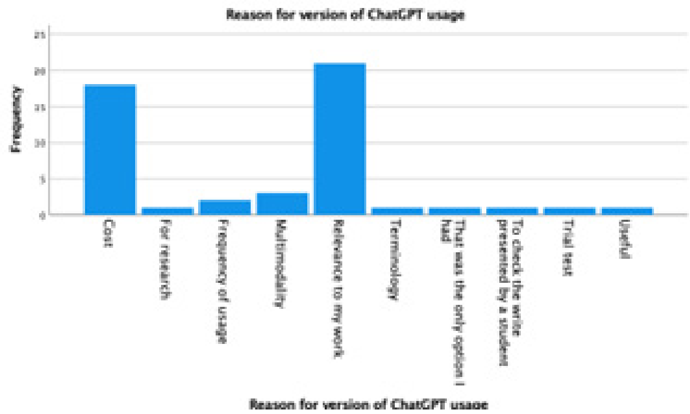


Figure 5: Reason for ChatGPT use by academics.

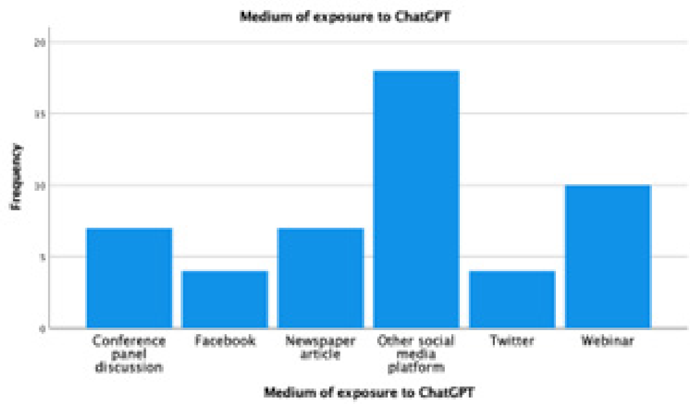


Figure 6: Medium of exposure to ChatGPT.

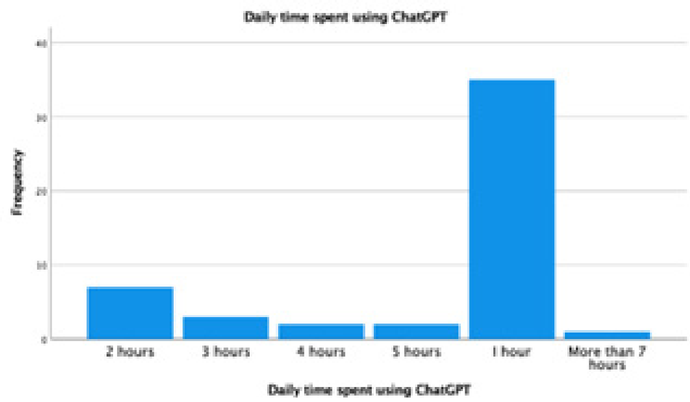


Figure 7: Daily time spent using ChatGPT by academics.

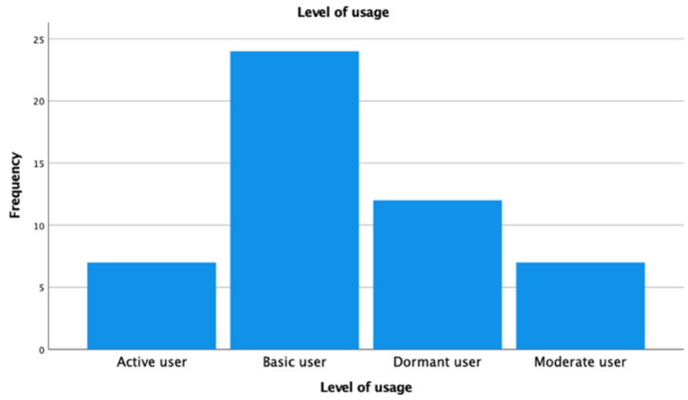


Figure 8: Level of usage of ChatGPT by academics.

Table 1: Descriptive statistics of job relevance of ChatGPT.

Variable	n	SD	MD	SWD	N	SWA	MA	SA	M	SD
Job Relevance	**50**								**4.51**	**1.89**
1. Are you in agreement or disagreement that the use of ChatGPT is important to your work?		10%	6%	18%	14%	6%	20%	26%		
2. Are you in agreement or disagreement that the usage of ChatGPT is relevant to your work?		12%	10%	12%	10%	10%	22%	24%		
3. The use of ChatGPT is pertinent to my various teaching and research-related tasks.		14%	6%	14%	12%	24%	14%	16%		

Note.

n (number of participants), SD (Strongly disagree), MD (Moderately disagree), SWD (Somewhat disagree), N (Neutral), SWA (Somewhat disagree), MA (Moderately agree), SA (Strongly agree), M (mean), and SD (Standard deviation)

From Table 1, it can be observed that a higher percentage of the participants moderately agreed (20%) or strongly agreed (26%) that ChatGPT is important to their work. However, a majority of the participants also moderately agreed (22%) or strongly agreed (24%) that ChatGPT is not relevant to their work. That is, while they recognized the importance of ChatGPT in aiding them in completing their work, they did not perceive its relevance. This could be because many academics do not actively use AI tools in their work or are prone to resist integrating new technology in their work.

Table 2: Descriptive statistics of perceived usefulness of ChatGPT.

Variable	n	SD	MD	SWD	N	SWA	MA	SA	M	SD
Perceived Usefulness	**50**								**4.12**	**1.89**
4. Using ChatGPT improves my performance at school.		18%	8%	16%	10%	18%	22%	8%		
5. Using ChatGPT in my learning increases my productivity.		14%	10%	14%	12%	16%	20%	14%		
6. Using ChatGPT enhances my effectiveness at school.		18%	8%	12%	18%	20%	14%	10%		
7. I find the ChatGPT to be useful in my learning.		16%	6%	8%	12%	28%	18%	12%		

Note.

n (number of participants), SD (Strongly disagree), MD (Moderately disagree), SWD (Somewhat disagree), N (Neutral), SWA (Somewhat disagree), MA (Moderately agree), SA (Strongly agree), M (mean), and SD (Standard deviation).

It can be seen that from Table 2, although a higher percentage of the participants moderately agreed that ChatGPT improves their performance (22%), a higher percentage also strongly disagreed that ChatGPT improves their performance at school (18%). The number of academics who strongly agreed that ChatGPT enhances their effectiveness at school (10%) is less than those who strongly disagreed that it does not enhance their effectiveness at school (18%). Similarly, only 12% of the academics perceived ChatGPT to be useful in their learning while 16% strongly disagreed. Generally, the perceived usefulness of ChatGPT among academics was somewhat low (M = 4.12, SD = 1.89).

Table 3: Descriptive statistics of perceived ease of use of ChatGPT.

Variable	n	SD	MD	SWD	N	SWA	MA	SA	M	SD
Perceived Ease of Use	**50**								**4.75**	**1.67**
8. My interaction with the ChatGPT is clear and understandable.		10%	2%	14%	10%	32%	22%	10%		
9. Interacting with the ChatGPT does not require a lot of my mental effort.		12%	2%	10%	16%	18%	30%	12%		
10. I find the ChatGPT to be easy to use.		10%	6%	6%	16%	8%	32%	22%		
11. I find it easy to get the ChatGPT to do what I want it to do.		12%	2%	8%	14%	14%	28%	22%		

Note.

n (number of participants), SD (Strongly disagree), MD (Moderately disagree), SWD (Somewhat disagree), N (Neutral), SWA (Somewhat disagree), MA (Moderately agree), SA (Strongly agree), M (mean), and SD (Standard deviation)

The Table illustrates that academics rated the perceived ease of use of ChatGPT as fairly low (M = 4.75, SD = 1.67). Nonetheless, most of the academics moderately agreed that their interaction with ChatGPT was clear and understandable (22%) as opposed to only 10% who strongly disagreed. The number of academics who moderately agreed that interacting with ChatGPT does not require mental effort was higher than those who strongly disagreed (12%). A similar finding can be seen in the rating between how academics perceived ChatGPT to be easy to use and easy to get ChatGPT to carry out commands.

Table 4. Descriptive statistics of perceived enjoyment of ChatGPT.

Variable	n	SD	MD	SWD	N	SWA	MA	SA	M	SD
Perceived Enjoyment	**50**								**4.80**	**1.68**
12. I find using ChatGPT to be enjoyable.		8%	0%	10%	20%	22%	24%	16%		
13. The actual process of using ChatGPT is pleasant.		8%	0%	10%	18%	22%	22%	20%		
14. I have fun using ChatGPT.		12%	2%	8%	20%	22%	18%	18%		

Note.

n (number of participants), SD (Strongly disagree), MD (Moderately disagree), SWD (Somewhat disagree), N (Neutral), SWA (Somewhat disagree), MA (Moderately agree), SA (Strongly agree), M (mean), and SD (Standard deviation).

Table 4 demonstrates that a higher percentage of academics moderately agree (24%) and strongly agree (26%) that ChatGPT is enjoyable to use. A similar finding can be seen in their ratings of how pleasant and fun ChatGPT is perceived to be. A higher number of academics strongly agreed that ChatGPT is pleasant (18%) and fun (18%) as opposed to those who did not.

Table 5. Correlational analysis of study variables.

	Awareness	Job Relevance	Perceived Usefulness	Perceived Ease of Use	Perceived Enjoyment
Awareness	1	.634**	.688**	.431**	.523**
Job Relevance	.634**	1	.761**	.490**	.724**
Perceived Usefulness	.688**	.761**	1	.283*	.523**
Perceived Ease of Use	.431**	.490**	.283*	1	.801**
Perceived Enjoyment	.523**	.724**	.523**	.801**	1

$^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$

Based on the information in Table 5, there is a significant correlation between Awareness and Job Relevance – r = 0.634** (p < 0.01), Awareness and Perceived Usefulness: r = 0.688** (p < 0.01), Awareness and Perceived Ease of Use: r = 0.431** (p < 0.01), Awareness and Perceived Enjoyment: r = 0.523** (p < 0.01). A positive significant correlation was found between Job Relevance and Perceived Usefulness: r = 0.761** (p < 0.01), Job Relevance and Perceived Ease of Use: r = 0.490** (p < 0.01), Job Relevance and Perceived Enjoyment: r = 0.724** (p < 0.01). A significant Pearson correlation coefficient was found between Perceived Usefulness and Perceived Ease of Use: r = 0.283* (p < 0.05), Perceived

Usefulness and Perceived Enjoyment: r = 0.523** ($p < 0.01$), Perceived Ease of Use and Perceived Enjoyment: r = 0.801** ($p < 0.01$). Overall, a significant positive correlation was found among all the variables. This means that an increase in one variable results in a subsequent increase in another variable. For example, regarding ChatGPT awareness, the more an academic is aware of ChatGPT, the more they find it relevant and useful to their work, easy to use and enjoyable when using it.

Discussion of findings

After the launch of the AI-powered conversational agent, ChatGPT, we have discovered an intense excitement and fear for the use of the generative pretrained transformer tool in teaching. With reference to what ChatGPT is and the awareness of academics in Ghanaian educational institutions, several issues emerged. We found that most of the participants did not have a conceptual understanding of ChatGPT and how it could be applied to learning, teaching and personal development. A possible reason for this could be that although ChatGPT is a new phenomenon that has created hype for the past few months, there was little research published about it during the time of the interviews. Now, the hype has generated a lot of academic research publications on AI-conversational agents uses in education. Many of the studies have reported significant results of ChatGPT as an AI language tool that creates opportunities for learning and teaching (Alshurafat, 2023; Baidoo-Anu & Owusu Ansah, 2023). Moreover, a few studies by African scholars located at universities outside the continent had published on this phenomenon (Tlili et al., 2023a; Baidoo-Anu & Owusu Ansah, 2023; Alshurafat, 2023). Studies revealed that ChatGPT is a "game-changer" that will transform all assessment protocols (Stokel-Walker & Van Noorden, 2023), and ChatGPT-based learning is a motivator for teaching (Ali et al., 2023). These studies became the drivers for Africans to launch the stage to start implementing AI machines for learning and teaching in daily practices. The study participants had misrepresentations of the purpose of using ChatGPT and chatbots in learning and teaching. Most of the participants in the study had not heard about ChatGPT. Based on the understanding of participants, this lecturer echoed the following view "Is ChatGPT a new app? I might have heard about chatbots, but I have not taken the time to read about them to know their main purpose. But after our talk, I will check it out" (L7).

In view of the lack of knowledge of participants, most had misconceptions of what artificial intelligence is, in particular, ChatGPT as a language tool and how to use it for teaching. They viewed it as a social media mobile application which one could download from Google Play Store. Because of a lack of digital literacy skills, academics quizzed if they had to have some technical skills or should have prior knowledge about chatbots before they could effectively use ChatGPT for teaching and learning. Moreover, studies indicated Africans must be aware and embrace innovative technologies such as ChatGPT, which has transformative potential in education and equip learners with transferable skills needed in the labour market (Santandreu et al., 2023; Rudolph et al., 2023b; Tlili et al., 2023a; Carpenter et al., 2022; UNESCO, 2021). It was evident that many of the participants lacked user experience. Their primary source of information about ChatGPT was from the authors during the data collection stage. The awareness of new technologies has been underscored by several scholars to foster positive attitudes among users towards the rapid adoption of technology (Dinev & Hu, 2007; Pandey et al., 2021).

We were to find out how academics experienced when exposed to ChatGPT as a language tool. Based on Rogers' (2003) DIT, the academics in this study fall in the laggards or, at best, late majority category. This situation has been enabled by factors such as the generally slow progress and late recognition of technological innovation in many African and other developing economies (Adarkwah, 2021; Agyei & Voogt, 2012; International Finance Corporation, 2019). In the case of this research, most of the participants were unaware of the AI facility and what it could do. This situation is encapsulated in a reflection shared by a participant, "Yeah, I have interacted with a chatbot before from a marketing store, but I thought I was talking to a service personnel or customer care. This is interesting. So, what is the main purpose of ChatGPT?" (P3).

Despite the lack of awareness of the academics and the limitations which had compounded the situation, it was interesting to note that the participants expressed enthusiasm after they had been informed about ChatGPT and how it could transform their teaching and learning. This is a depiction of their persuasion and interest (Rogers, 2003) in the technology after a mental assessment of its capacity to create more meaningful and impactful teaching and learning environments. This enthusiasm and interest are in consonance with assertions by researchers (such as Chaka, 2023; Dinev & Hu, 2007; Pandey et al., 2021) that positive attitudes toward new technologies could foster innovative classrooms. In addition, such innovative technologies, especially ChatGPT, are touted as transforming education generally and equipping students with transferrable skills, which are requirements to fit into the ever-competitive 21st-century labour market.

Furthemore, this research demonstrated academics' perspectives in Ghana about ChatGPT and its acceptance as an innovative technology. Considering the poverty of the academics' awareness of ChatGPT in the qualitative phase of the study, they demonstrated limited knowledge about the use of ChatGPT or how it can facilitate their teaching and academic growth. This leverages promoting technological awareness and digital skills for the university teachers to be able to optimally function in today's job market through different channels such as newsletters and blogs, technology podcasts, technology conferences and forums, social media, technology clubs, technology professional development, and partnerships and technology centres (Adarkwah & Huang, 2023; Bizclik, 2021; Moore, 2022).

Although they believe that ChatGPT is important to their profession, the majority found that it is not relevant to their work. This could be because many academics do not actively use AI tools in their work or are prone to resist integrating new technology in their work (Van Wyk et al., 2023; Rogers, 2003). This also could be because of the novelty effect of this technology, especially because they lack the awareness of it and have a positive attitude to possibly adopt it.

The findings of this research raise a debate about using Chatbots and academic performance when there is a contract in the academic's opinions. Even this debate extended to the potential of ChatGPT to enhance their effectiveness at school and to be useful in their learning (Santandreu et al., 2023; Rudolph et al., 2023b). This could relate to the nature of their use to serve their learning objectives, their majors, or their attitudes toward technological tools. It is early to ultimately judge its effectiveness and usefulness in their learning after a few months of its launch, though Tlili et al. (2023b) have questioned the quality of reviews conducted on the phenomenon. Therefore, we suggest conducting follow-up studies to verify their perception of using it when they are fully aware of AI technologies in education.

Since academics rated the perceived ease of use of ChatGPT as fairly low, this is consistent with theme two of the qualitative study that emphasized the lack of technical skills and usability (Van Wyk et al., 2023). Meanwhile, their interaction with ChatGPT was clear and understandable and did not require mental effort. This emphasizes the importance of a thorough understanding how to use it in terms of user experience (Mogavi et al., 2023). Everyone can easily interact with the conversational AI to ask follow-up questions and find answers, but lacking the awareness and knowledge of implementation could affect the ease of use perception. Thus, we suggest arranging institutional training for academics to introduce the mechanisms and nature of AI technologies supported by the responsible use and ways of integration in education.

Although the general findings of this research indicated the minimal awareness of ChatGPT by academics and mixed views about its relevance and usefulness in work-related tasks, a higher percentage agreed that ChatGPT is enjoyable to use, pleasant and fun (Van Wyk et al., 2023). It can be explained in light of satisfying their needs and curiosity about this technology. They possibly used to use search engines and refine the search results for placing information. Unlike this way, they simply felt that it could save their time and effort instead of browsing hundreds of websites and resources, hence reflecting on their perceptions of enjoyment. Drawing on the basic psychological needs theory as a sub-theory of a human motivation macro-theory known as self-determination theory (Ryan & Deci, 2000), satisfaction of the need for autonomy is crucial for motivation. Thus, they enjoyed trying the new tool or functions because they were motivated to use it, generated by their curiosity.

According to the research result, we suggest extending the study context to include a wide range of participants from different majors in Ghana, considering the factors affecting their adoption of the AI tools in education (e.g. the academic background, age, the prior experience of using technological tools, etc.).

Ways of promoting techno-trend awareness and fostering technology acceptance in Ghana

Technological awareness and acceptance in this AI era are important because of the rapid use of different types of digital tools in work-based and educational settings. It is almost difficult to refrain from using technology in everyday life in our modern society. An individual has to possess digital skills to be able to function optimally in today's job market. The labour market requires workers to be digitally literate to ensure high work productivity. Those who are proficient with technology are often perceived as having more career opportunities than those who are not digital natives. Human-machine collaboration, integrating technology into teaching, technology for inclusivity, etc. all indicate how technology is pivotal in this fourth industrial revolution. Thus, technology trend awareness can foster the rapid uptake of novel technologies, creativity, and innovation. Below, we enumerate ways education systems, particularly higher education, can promote technology trend awareness.

Newsletters and blogs: Institutions of higher education (IHE) can subscribe to newsletters and blogs that provide information on technological trends and share it with their academic staff and students. Some of the world's top technology newsletters include Technology Magazine, TechCrunch, The Other Valleys, Dense Discovery, and CB Insights (Bizclik, 2021).

Technology podcasts: IHE can create a technology podcast that relays information on novel innovations with learners. Additionally, there are several technological podcasts that can be made available in school libraries or reading rooms for academics. An example of such a podcast is This Week in Tech (TWiT). A podcast could be a great source of technology news and information (Moore, 2022). Podcasting forms part of the wider expansion and diversification of digital technology in education in response to the need for increased student engagement, the incorporation of alternative methods of inquiry and epistemologies into curricula, and the promotion of innovative and extensive dissemination of knowledge and research (Moore, 2022).

Technology conferences and forums: Research about the latest technologies and the best practices on their use can be shared during organized conferences that focus on technology. Scholars can make poster presentations or participate in technology exhibitions during academic conferences. For instance, at the first author's university, a forum is been organized on ChatGPT. Fisher and Purcal (2017) talked about how e-learning trends appear on top of the agenda of scientific conferences. Such forums could be replicated in Ghanaian educational institutions through collaborations with universities in developed economies, which are early adopters of technological innovations.

Good social media use: Fostering good social media use by encouraging academics to follow pages and groups that provide information on technological trends and education. For example, news about ChatGPT and other AI products is often discussed on Twitter and LinkedIn pages. Academics can also watch news about technology on YouTube. Online communities like Quora and Reddit's technology can provide academics with the latest news on technological products. Moon and Hadley (2014) mentioned that news organizations make use of Twitter as a good source of diverse kinds of information. Wilson (2000) adds that librarians are able to monitor technology trends with podcasts, RSS and Twitter (rebranded as X).

Technology clubs/groups: IHE can create technology associations in the form of clubs and groups where insights can be shared about the latest trends in technology.

Technology professional development: Professional training programmes on the use of technology can be organized for academics through seminars and workshops (Adarkwah & Huang, 2023). Academics can receive real-time or hands-on training on how to use new technologies like ChatGPT. The introduction of new technologies can be difficult for teachers to grasp its use, making technology professional development urgent (King, 2002). Martin et al. (2010) also found effective technological professional development to be correlated with high-quality lesson plans and learning performance.

Partnerships and technology centers: IHE can partner with both domestic and international tech companies to provide information about novel technologies and their use (Amponsah & Bekele, 2022). Technological centers can be created with the sole purpose of educating academics about latest technologies. Computer science programmes and information technology communication (ICT) centers can be leveraged to provide news about ground-breaking technology to academics.

Conclusion, implications, and limitations

The study's findings serve as a foundational assessment of recent literature on ChatGPT and why the discussions on ChatGPT are a classic case scenario of the slow pace of technological growth in IHE in most emerging economies such as Ghana. The conversations with 34 academics presented in this study suggest that they lacked knowledge about ChatGPT. Specifically, academics had limited knowledge about what AI chatbots mean and inquired about the purpose of ChatGPT, how it can be used, and the extent to which outputs in response to user commands are accurate. Generally, the academics in the study had a high enthusiasm about the educational possibilities ChatGPT can afford after being exposed to the chatbot.

Moreover, the quantitative phase of the study also confirmed that participants were not highly aware of ChatGPT even after several months had elapsed after its launch and the public hype about ChatGPT was gaining ground. Regarding the job relevance and perceived usefulness of ChatGPT, there were conflicting views among academics. Nonetheless, most of the academics perceived the interface of ChatGPT to be fun to use and easy to navigate.

From a theoretical implication perspective, the study serves as a springboard for researchers in Ghana and peer countries to join the discussions on ChatGPT and assess the readiness of faculty, staff and students to use AI chatbots like ChatGPT for educational purposes. Another research area to focus on is how IHE can create or revise their educational policies to include advanced chatbots like ChatGPT for pedagogical purposes. Policies can focus on academic integrity issues and means of providing easy access to ChatGPT Plus, which is a payable/subscription version of ChatGPT. Also, regarding human-machine collaboration, IHE can focus on how to effectively integrate ChatGPT into teaching. IHE can create practical guides on the safe use of ChatGPT from scratch or make use of a publicly available handbook for educators.

Practically, the extent of the 'unawareness' of intelligent chatbots like ChatGPT among academics in Ghana is a wakeup call for university leaders and educators to find various means to create technological trend awareness in IHE and other levels of education which could benefit from the promising features of ChatGPT. Without promoting technology trend awareness, the likelihood of fostering the rapid adoption of innovative technologies would be near impossible. Some of the key strategies suggested earlier include a subscription to technology newsletters and blogs, good use of social media for learning new tech information on X and LinkedIn, the provision and creation of technology podcasts, technology professional development for academics, and organizing tech conferences and forums, among others.

A limitation of the study is that the review approach chosen was not systematic in nature. A more rigorous approach would be needed to focus on presenting findings on different dimensions of ChatGPT. Also, a large-scale quantitative study on ChatGPT awareness and readiness to use in educational settings would be beneficial. The limitations notwithstanding, this study has broken grounds for academics, IHEs and stakeholders in education in Ghana and similar countries to rethink higher education to expedite the rate of adoption of generative AI tools such as ChatGPT.

Acknowledgement

We would like to thank the academics who voluntarily participated in this study and shared their candid views to enrich the findings.

Data availability statement

Rhe datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Abramski, K., Citraro, S., Lombardi, L., Rossetti, G., & Stella, M. (2023). Cognitive network science reveals bias in GPT-3, GPT-3.5 Turbo, and GPT-4 Mirroring math anxiety in high-school students. *Big Data and Cognitive Computing, 7*(3), 124. https://doi.org/10.3390/bdcc703012424

Adarkwah, M. A. (2021). "I'm not against online teaching, but what about us?”: ICT in Ghana post Covid-19. *Education and Information Technologies, 26*(2), 1665–1685. https://doi.org/10.1007/s10639-020-10331-z

Adarkwah, M. A., & Huang, R. (2023). Technology addiction, abduction and adoption in higher education: Bird’s eye view of the ICT4AD policy in Ghana 20 years on. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13352

Agyei, D. D., & Voogt, J. (2012). Developing technological pedagogical content knowledge in pre-service mathematics teachers through collaborative design. *Australasian Journal of Educational Technology, 28*(4), 4. https://doi.org/10.14742/ajet.827

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. (2023). Impact of ChatGPT on learning motivation: teachers and students' voices. *Journal of English Studies in Arabia Felix, 2*(1), 41-49.

Amponsah, S., & Bekele, T. A. (2022). Exploring strategies for including visually impaired students in online learning. *Education and information technologies,* 1-23.

Alshurafat, H. (2023). *The usefulness and challenges of chatbots for accounting professionals: Application on ChatGPT.* SSRN 4345921.

Auty, R. M. (1994). Industrial policy reform in six large newly industrializing countries: The resource curse thesis. *World Development, 22*(1), 11–26. https://doi.org/10.1016/0305-750X(94)90165-1

Aydın, Ö., & Karaarslan, E. (2022). *OpenAI ChatGPT generated literature review: Digital twin in healthcare* (SSRN Scholarly Paper 4308687). https://doi.org/10.2139/ssrn.4308687

Baidoo-Anu, D., & Owusu Ansah, L. (2023). *Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning*. SSRN 4337484.

Bizclik. (2021, August 12). *Top ten technology newsletters.* https://technologymagazine.com/top10/top-ten-technology-newsletters

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Carpenter, J. P., Kerkhoff, S. N., & Wang, X. (2022). Teachers using technology for co-teaching and crowdsourcing: The case of global read aloud collaboration. *Teaching and Teacher Education, 114*, 103719. https://doi.org/10.1016/j.tate.2022.103719

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching, 6*(2), 1-11. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.12

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating. Ensuring academic integrity in the era of ChatGPT. *EdArXiv*. https://doi.org/10.35542/osf.io/mrz8h

Creswell, J. W. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th edition). SAGE Publications, Inc. https://us.sagepub.com/en-us/nam/research-design/book270550

Creswell, J. W., Hanson, W. E., Clark Plano, V. L., & Morales, A. (2007). Qualitative research designs: Selection and implementation. *The Counseling Psychologist, 35*(2), 236–264. https://doi.org/10.1177/0011000006287390

Creswell, J. W., Shope, R., Clark, V. L. P., & Green, D. O. (2006). How interpretive qualitative research extends mixed methods research. *Research in the Schools, 13*(1), 1–11.

Czarniawska, B. (2004). *Narratives in social science research.* SAGE Publications, Ltd. https://doi.org/10.4135/9781849209502

Dearing, J. W., & Cox, J. G. (2018). Diffusion of innovations theory, principles, and practice. *Health Affairs, 37*(2), 183-190.

Dinev, T., & Hu, Q. (2007). The centrality of awareness in the formation of user behavioral intention toward protective information technologies. *Journal of the Association for Information Systems, 8*(7). https://doi.org/10.17705/1jais.00133

Fisher, K. R., & Purcal, C. (2017). Policies to change attitudes to people with disabilities. Scandinavian Journal of Disability Research, 19(2), Article 2. https://doi.org/10.1080/15017419.2016.1222303

Frieder, S., Pinchetti, L., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., Chevalier, A., & Berner, J. (2023). *Mathematical capabilities of ChatGPT* (arXiv:2301.13867). arXiv. https://doi.org/10.48550/arXiv.2301.13867

Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2022). *How does chatgpt perform on the medical licensing exams? The implications of large language models for medical education and knowledge assessment* medRxiv. https://doi.org/10.1101/2022.12.23.22283901

Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., ... & Vandrik, S. (2023). *Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education: A guide for students and lecturers* (No. 02-2023). Hohenheim Discussion Papers in Business, Economics and Social Sciences.

Hanson, W. E., Creswell, J. W., Clark, V. L. P., Petska, K. S., & Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology, 52*(2), 224–235. https://doi.org/10.1037/0022-0167.52.2.224

Holloway, I., & Todres, L. (2003). The status of method: flexibility, consistency and coherence. *Qualitative Research, 3*(3), 345-357.

Hosseini, M., & Horbach, S. P. (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review, 8*(1), 4.

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2), 1-8. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.2

International Finance Corporation. (2019). *Digital skills in sub-Saharan Africa spotlight on Ghana*. https://www.ifc.org/wps/wcm/connect/ed6362b3-aa34-42ac-ae9f-c739904951b1/Digital+Skills_Final_WEB_5-7-19.pdf?MOD=AJPERES

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*. EdArXiv. https://doi.org/10.35542/osf.io/5er8f

Khan, R. A., Jawaid, M., Khan, A. R., & Sajjad, M. (2023). ChatGPT-Reshaping medical education and clinical management. *Pakistan Journal of Medical Sciences, 39*(2), 605.

Khandkar, S. (2009). *Open coding*. University of Calgary. https://pages.cpsc.ucalgary.ca/~saul/wiki/uploads/CPSC681/open-coding.pdf

King, K. P. (2002). Educational technology professional development as transformative learning opportunities. *Computers & Education, 39*(3), 283–297. https://doi.org/10.1016/S0360-1315(02)00073-8

Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability, 15*(7), 5614. https://doi.org/10.3390/su15075614

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., Leon, L. D., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health, 2*(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198

Martin, W., Strother, S., Beglau, M., Bates, L., Reitzes, T., & McMillan Culp, K. (2010). Connecting instructional technology professional development to teacher and student outcomes. *Journal of Research on Technology in Education, 43*(1), 53–74. https://doi.org/10.1080/15391523.2010.10782561

Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1), 16-30. https://doi.org/10.37074/jalt.2023.6.1.34

Mogavi, R. H., Deng, C., Kim, J. J., Zhou, P., Kwon, Y. D., Metwally, A. H. S., ... & Hui, P. (2023). *Exploring user perspectives on chatgpt: Applications, perceptions, and implications for ai-integrated education*. arXiv preprint arXiv:2305.13114.

Moon, S. J., & Hadley, P. (2014). Routinizing a new technology in the newsroom: Twitter as a news source in mainstream media. *Journal of Broadcasting & Electronic Media, 58*(2), 289–305. https://doi.org/10.1080/08838151.2014.906435

Moore, T. (2022). Pedagogy, podcasts, and politics: What role does podcasting have in planning education? *Journal of Planning Education and Research*, *0*(0). https://doi.org/10.1177/0739456X221106327

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1609406917733847.

OpenAI. (2023). *ChatGPT: Optimizing language models for dialogue*. https://openai.com/blog/chatgpt/

Pandey, V., Misra, N., Greeshma, R., Astha, A., Jeyavel, S., Lakshmana, G., Rajkumar, E., & Prabhu, G. (2021). Techno trend awareness and its attitude towards social connectedness and mitigating factors of COVID-19. *Frontiers in Psychology,* 12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.637395

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior, 140*, 107600. https://doi.org/10.1016/j.chb.2022.107600

Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching, 6*(2), 1-7. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.4

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed). Free Press.

Rospigliosi, P. A. (2023). Artificial intelligence in teaching and learning: What questions should we ask of ChatGPT?. *Interactive Learning Environments, 31*(1), 1-3.

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. Journal of Applied Learning and Teaching, 6(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist, 55*(1), 68.

Santandreu, D. C., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching, 6*(2), 1-15. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.20

Stokel-Walker, C., & Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature, 614*(7947), 214-216.

Susnjak, T. (2022). *ChatGPT: The end of online exam integrity?* arXiv. https://doi.org/10.48550/arXiv.2212.09292

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023a). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Tlili, A., Huang, R., Mustafa, M. Y., Zhao, J., Bozkurt, A., Xu, L., ... & Burgos, D. (2023). Speaking of transparency: Are all Artificial Intelligence (AI) literature reviews in education transparent?. *Journal of Applied Learning and Teaching, 6*(2), 1-12. https://doi.org/10.37074/jalt.2023.6.2.15

UNESCO. (2021). *Reimagining our futures together: A new social contract for education*. United Nations Educational, Scientific and Cultural Organization,. https://unesdoc.unesco.org/ark:/48223/pf0000379707.locale=en

Vaishya, R., Misra, A., & Vaish, A. (2023). ChatGPT: Is this version good for healthcare and research? *Diabetes & Metabolic Syndrome: Clinical Research & Reviews,* 102744. DOI: https://doi.org/10.1016/j. dsx.2023.102744

Van Wyk, M. M., Adarkwah, M., & Amponsah, S. (2023). Why all the hype about ChatGPT? Academics' views of a Chat-based conversational learning strategy at an open distance e-learning institution. *Open Praxis, 15*, 214–225. https://doi.org/10.55982/openpraxis.15.3.563

Wilson, E. V. (2000). Student characteristics and computer-mediated communication. *Computers & Education, 34*(2), 67–76. https://doi.org/10.1016/S0360-1315(99)00041-X

Zhai, X. (2022). *ChatGPT user experience: Implications for education* (SSRN Scholarly Paper 4312418). https://doi.org/10.2139/ssrn.4312418

Appendices

Appendix 1: Qualitative phase (interview questions).

This interview aims to investigate your general awareness and acceptance of chatbots, specifically ChatGPT, in education. The interview contains 5 open-ended questions which have been derived from the objectives of the study. Anonymity will be ensured, and all answers will be confidential. Please answer each question with sufficient information and details to ensure the usefulness of this interview and study. You are welcomed to also probe into the questions. Thanks for your time.

Section A
What is your position at the university?
What is your gender?
Section B

1. What do you know about chatbots in general?

2. What is your experience in using chatbots?

3. How familiar are you with ChatGPT as a generative AI tool for education?

4. What is your experience in using ChatGPT?

5. In your opinion, to what extent can chatbots, particularly ChatGPT, be beneficial for your work?

Appendix 2: Quantitative phase.

Section A

Thank you very much for accepting this invitation to participate in this survey which aims at investigating the acceptance of ChatGPT for educational purposes. The survey has two main sections; 1. Demographic information and 2. Questions. The survey questions are split into nine (9) sections with each consisting of a number of questions. In total, it takes an average of five (5) minutes to complete the questionnaire. All answers will remain confidential. Please, select the most appropriate answers. Thank you!

Demographic Data

This section collects your personal information; age, gender, and discipline at school. Please indicate/select the appropriate answer.

1. What is your age?

☐ 20-24
☐ 25-29
☐ 30-34
☐ 35-39
☐ 40-44
☐ 45 and above

2. What is your gender?

☐ Male
☐ Female

3. What is your area of specialization?

☐ Education
☐ Business
☐ Computer Science
☐ Geography
☐ Engineering
☐ Environment
☐ Philosophy
☐ Law
☐ Health and Medicine
☐ Other

4. What version of ChatGPT do you use?

☐ ChatGPT 3.5 (basic version)

☐ ChatGPT 4 (plus version)

5. Why do you use the version of ChatGPT you use?

Cost

☐ Frequency of usage

☐ Relevance to my work

☐ Multimodality

6. How did you get to know about ChatGPT and its use?

☐ Webinar

☐ Newspaper article

☐ Conference panel discussion

☐ Twitter

☐ Facebook

☐ Other social media platforms

7. On the average how much time do you spend on ChatGPT each day?

☐ 1 hour

☐ 2 hours

☐ 3 hours

☐ 4 hours

☐ 5 hours

☐ 6 hours

☐ 7 hours

☐ More than 7 hours

Section B

Questions on acceptance of ChatGPT

All items are measured on a 7-point Likert scale (where 1: strongly disagree; 2: moderately disagree, 3: somewhat disagree, 4: neutral (neither disagree nor agree), 5: somewhat agree, 6: moderately agree, and 7: strongly agree), except Use of ChatGPT, which was measured using a multiple choice option.

Variable	SD	MD	SWD	N	SWA	MA	SA
Job Relevance							
1. Are you in agreement or disagreement that the use of ChatGPT is important to your work?							
2. Are you in agreement or disagreement that the usage of ChatGPT is relevant to your work?							
3. The use of ChatGPT is pertinent to my various teaching and research-related tasks.							

Variable	SD	MD	SWD	N	SWA	MA	SA
Perceived Usefulness							
4. Using ChatGPT improves my performance at school.							
5. Using ChatGPT in my learning increases my productivity.							
6. Using ChatGPT enhances my effectiveness at school.							
7. I find the ChatGPT to be useful in my learning.							

Variable	SD	MD	SWD	N	SWA	MA	SA
Perceived Ease of Use							
8. My interaction with the ChatGPT is clear and understandable.							
9. Interacting with the ChatGPT does not require a lot of my mental effort.							
10. I find the ChatGPT to be easy to use.							
11. I find it easy to get the ChatGPT to do what I want it to do.							

Variable	SD	MD	SWD	N	SWA	MA	SA
Perceived Enjoyment							
12. I find using ChatGPT to be enjoyable.							
13. The actual process of using ChatGPT is pleasant.							
14. I have fun using ChatGPT.							

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools

Chaka Chaka[A]

[A] *Professor, University of South Africa, Pretoria, South Africa*

Keywords

ChatGPT;
Copyleaks AI Content Detector;
generative AI;
Giant Language model Test Room;
GPTZero,
higher education;
OpenAI Text Classifier;
Writer.com's AI Content Detector.

Correspondence

chakachaka8@gmail.com [A]

Article Info

Received 24 June 2023
Received in revised form 21 July 2023
Accepted 22 July 2023
Available online 24 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.12

Abstract

This paper set out to test the accuracy of five AI content tools, GPTZero, OpenAI Text Classifier, Writer.com's AI Content Detector, Copyleaks AI Content Detector, and Giant Language model Test Room, to detect AI-generated content in the responses generated by ChatGPT, YouChat, and Chatsonic. The responses were generated from these three AI chatbots using English prompts related to applied English language studies. Then, the ChatGPT-generated responses were Google-translated into German, French, Spanish, Southern Sotho, and isiZulu, and inputted into GPTZero for it to detect the AI-generated content in them. Additionally, the ChatGPT-generated responses Google-translated into German, French and Spanish were inputted into Copyleaks AI Content Detector for it to detect the AI-generated content in them. For the ChatGPT-, YouChat-, and Chatsonic-generated responses, Copyleaks AI Content Detector emerged as the top-most performing AI content detector among the five AI content detectors. It was followed by OpenAI's AI Text Classifier. Concerning the ChatGPT-generated responses that were Google-translated into five languages, GPTZero misidentified all of them as human-produced. For the ChatGPT-generated responses that were Google-translated into German, French and Spanish, Copyleaks AI Content Detector correctly identified three of the German-translated texts, five of the French-translated texts, and all the Spanish-translated texts as AI-generated. Thus, it is evident from this paper that all five AI content detectors seem not yet fully ready to accurately and convincingly detect AI-generated content from machine-generated texts in different contexts. This has dire consequences for AI-generated plagiarism in academic essay writing.

Introduction

The launch of ChatGPT, a generative artificial intelligence (AI) chatbot owned by OpenAI (OpenAI, 2022), on 30 November 2022, had a domino effect in cyberspace and in the real-life world. It not only rattled the AI world in which generative AI chatbots, which before ChatGPT were relatively unknown, suddenly emerged or announced their presence (Chaka, 2023a; Eliaçik, 2023a; Hetler, 2023; Kanran, 2023), but it also led to the emergence of AI content detection tools intended to detect and to differentiate between AI-generated and human-written texts. One such AI content detection tool, which was launched late in 2022 as a direct consequence of ChatGPT, is GPTZero. The first part of its name is directly linked to the last part of ChatGPT's name. Much more will be said about GPTZero below. In a manner almost resembling what happened after ChatGPT was released, similar AI detection tools emerged or announced their presence in the aftermath of GPTZero's launch. Examples of such tools are AI Text Classifier, Giant Language Model Test Room (GLTR), Writer.com's AI Content Detector, and Copyleaks AI Content Detector (Lim, 2023; Outlook Spotlight, 2023; Chrome, 2023; Copyleaks, 2023). Again, much more will be said about these tools below.

All these AI-powered content detection tools emerged when there were assertions that no current AI content tool could detect AI plagiarism in ChatGPT-generated responses (Chaka, 2023b; Cutcliffe, 2022; Heilweil, 2022). While these tools can be used to detect what Lim (2023) calls AI-assisted content in different text types in general, it is AI-assisted academic content that is the focus of this paper. This is more so as immediately after the release of ChatGPT, some schools and universities reacted by saying that they would ban it because of the temptation it had for students to use it in school- or university-level essays (Anders, 2023; Barnett, 2023; Caren, 2022; Ceres, 2023; Harris, 2022; Hern, 2022; Somoye, 2023; Stokel-Walker, 2022; Wingard, 2023). Of course, there were some academic and science journals that were said to have taken a stance to ban ChatGPT as well (Sample, 2023). So, at issue here is AI-assisted academic content that tends to characterise responses produced by generative AI chatbots such as ChatGPT and others similar to it. This type of plagiarism is a grave concern for schools and universities.

Literature review

With the advent of generative AI-powered large language model (LLM) chatbots, which was heralded by ChatGPT's release in November 2022, there has been a growing number of scholarly papers that focus on and explore these chatbots. Examples of such scholarly papers include, but are not limited to: Alser & Waisberg (2023), Chaka (2023a), Cotton et al. (2023), Ifelebuegu (2023), Popenici (2023), Rasul et al. (2023), Rudolph et al. (2023a, b), Sullivan et al. (2023), and Yeadon et al. (2023). Some of these papers are published papers, while others are preprints, a publication pattern that almost resembles that of papers published during the COVID-19 pandemic (Chaka, 2020). Among these two streams of scholarly papers, there are those that explore the risks posed by ChatGPT for academic integrity concerning student assessment (see Ifelebuegu, 2023; Khalil & Er, 2023; Perkins, 2023; Rudolph et al., 2023a, b; Sullivan et al., 2023; Ventayen, 2023; Yeadon et al., 2023). But the critical issue with regard to academic integrity for most educational institutions is detecting plagiarism and distinguishing AI-generated content from human-written content. This is more so for both student essay writing and scholarly writing. In addition to the AI content detection tools specified in the preceding section, other tools include OriginalityAI, Content At Scale, Kazan SEO, GPT-2 Output Detector (Outlook Spotlight, 2023), Crossplag AI Content Detector (Lim, 2023), Claude AI, AI Writing Check, GPT Radar, and CatchGPT (Wiggers, 2023). Additional tools are Corrector App AI Content Detector, Plagibot, CopyScape, Winston AI, Writefull GPT Detector, Turnitin (Uzun, 2023), SciSpace, Hive Moderation, Hello Simple AI (Awan, 2023), PlagiarismCheck, Check For AI, DetectGPT, Compilation, and Go Winston (Weber-Wulff et al., 2023).

Since most of these AI content detectors are new, not much research has been conducted to evaluate their efficacy, accuracy, and reliability in terms of distinguishing between the content generated by current AI-powered LLM chatbots and the content written by humans. So, this is a new and growing area that still needs a lot of research. Of the few scholarly papers focusing on this area, a lot of them are preprints. Five such preprints are Aremu (2023), Cai and Cui (2023), Guo et al. (2023), Ventayen (2023), and Weber-Wulff et al. (2023). Two of these papers, Aremu (2023 and Weber-Wulff et al. (2023), have some relevance to the current paper. These two papers are briefly reviewed by discussing only aspects of them that have some bearing on this paper.

Aremu's (2023) paper investigated the capability of six AI text detectors, Sapling AI, Crossplag AI Content Detector, OpenAI Text Classifier, ZeroGPT, GPTZero, and Content At Scale, to accurately identify different essay types written by humans and those generated by AI (ChatGPT). The essay types in question were argumentative, descriptive, expository, and narrative essays. Their sample numbers were as follows: argumentative = 13; descriptive = 17; expository = 11; and narrative = 11. These sample numbers were split almost equally between the two datasets: human-written and AI-generated essay types. The prompts for the four essay types were as follows, respectively: Gun control; A day at the beach; The benefits of regular exercise; and A journey towards self-discovery. The human-written essay samples were obtained from the Internet, and were pre-2022 (before the advent of ChatGPT), while the AI-generated essays were sourced from ChatGPT by using the same four prompts with their attendant enhancements. In the main, these AI detectors performed well in accurately recognising human-written essays. In contrast, they performed poorly in detecting ChatGPT-generated and enhanced essays. Crossplag and Content At Scale outperformed the other AI detectors by accurately identifying human-authored essays with consistency and reliability, while ZeroGPT and GPTZero outdid the other detectors in terms of identifying ChatGPT-generated essays. This indicates their being robust and resistant to content deception (Aremu, 2023).

Weber-Wulff et al.'s (2023) paper employed 14 AI detection tools to examine their accuracy and error types in distinguishing between human-written text and AI-generated (ChatGPT-generated) text. These tools consisted of 12 publicly available AI detection tools and two commercial plagiarism detection tools. They were: Check For AI; Compilatio; Content at Scale; Crossplag; DetectGPT; Go Winston; GPT Zero; GPT-2 Output Detector Demo; OpenAI Text Classifier; PlagiarismCheck; Turnitin; Writeful GPT Detector; Writer; and ZeroGPT. All of these tools were non-premium versions. The paper used 54 test cases that were divided into the following five categories of English-language files:

- human-written;
- human-written in a non-English language with a resultant AI/machine translation to English;
- AI-generated text;
- AI-generated text with resultant human manual edits; and
- AI-generated text with resultant AI/machine paraphrase.

The human-written test cases were produced by nine people (eight researchers and one collaborator), and represented diverse disciplines such as academic integrity, computer science, civil engineering, economics, history, linguistics, and literature. They were written in Bosnian, Czech, German, Latvian, Slovak, Spanish, and Swedish and were machine-translated into English using DeepL (3 cases) and Google Translate (6 cases). Different prompts were used to generate AI texts through ChatGPT. Two additional texts were generated from ChatGPT using fresh prompts. One set of them was manually edited by exchanging words with their synonyms or by re-ordering sentence parts. The other set was automatically rewritten by employing an AI-powered tool, Quillbot. In terms of detection accuracy across all text cases, Turnitin (ranked 1) and Compilatio (ranked 2) topped the other tools, while PlagiarismCheck (ranked 13) and Content at Scale (ranked 14) were the most poorly-performing tools. The paper concludes that its findings failed to confirm the accuracy claims made by the detection tools it used, as these tools are unsuitable for providing evidence of academic misconduct. It also concludes that these tools are amenable to gaming, especially through paraphrasing and machine translation (Weber-Wulff et al., 2023).

Research problem

With the rising number of generative AI-powered LLM chatbots, there are growing concerns about the risks these chatbots pose to academic integrity by academics and educational institutions. To address these concerns, a number of online AI content detection tools have been released following the launch of ChatGPT. All these tools make bold claims (mostly undercut by concomitant disclaimers) about their accuracy rate and their reliability in detecting AI-generated content (see Chaka, 2023a, b; Chrome, 2023; Copyleaks, 2023; Kirchner et al., 2023; Outlook Spotlight, 2023; Tech Desk, 2023; Tyrrell, 2023; Weber-Wulff et al., 2023; Wiggers, 2023). Amid this burgeoning number of AI content detection tools, there is a need to evaluate the accuracy and reliability of these tools to differentiate between AI-generated content and human-produced content. This is critical as their efficacy in doing so will help academics and educational institutions know when student content is human-written and when it is AI-generated. The distinction between the content generated by an AI tool and the one produced by a human, or what Uzun (2023) calls the author factor, becomes trickier to determine as manipulating any form of content tends to elude most of the currently available AI detection tools (see Aremu, 2023; Cai & Cui, 2023; Guo et al., 2023; Uzun, 2023; Ventayen, 2023; Weber-Wulff et al., 2023). Related to the author factor is the content factor, the validity and reliability of the content produced, both AI-generated and human-written content (see Uzun, 2023).

Against this background, the purpose of this paper is three-fold: to test the accuracy of five AI content detection tools to detect the content generated by three AI chatbots, ChatGPT, YouChat, and Chatsonic, in its original English version; to evaluate the accuracy of one of these five AI content detection tools to detect the German, French, Spanish, Southern Sotho, and isiZulu versions of this content as machine-translated by Google Translate; and to test the accuracy of one of these five AI content detection tools to detect the German, French, and Spanish versions of this content as machine-translated by Google Translate. The five AI content detection tools are: GPTZero, OpenAI Text Classifier, Writer.com's AI Content Detector, Copyleaks AI Content Detector, and Giant Language Model Test Room. Relatedly, the paper's research questions are:

- What is the accuracy of the five AI content detection tools (GPTZero, OpenAI Text Classifier, Writer.com's AI Content Detector, Copyleaks AI Content Detector, and Giant Language Model Test Room) in detecting the content generated by ChatGPT, YouChat, and Chatsonic, in its original English version?
- What is the accuracy of GPTZero in detecting the German, French, Spanish, Southern Sotho, and isiZulu versions of this content as machine-translated by Google Translate?
- What is the accuracy of Copyleaks AI Content Detector in detecting the German, French, and Spanish versions of this content as machine-translated by Google Translate?

As pointed out above, there is currently a paucity of research that has been conducted in the area of study highlighted above. Thus, the current paper attempts to make a contribution to this area of study.

Reviewing of the five AI content detectors

GPTZero

GPTZero is an AI content detection tool built by a senior computer science student at Princeton University shortly after the release of ChatGPT. As its name indicates, it is intended to detect whether a text generated by ChatGPT is AI-generated or human-written (Chaka, 2023a; Ofgang, 2023; Tech Desk, 2023; Tyrrell, 2023). Of course, in this sense, it has a wider application beyond the ChatGPT-generated text to text generated by other generative AI tools, including ordinary human-written responses or essays that have nothing to do with AI generation. Therefore, it can also be referred to as an AI content detection app.

How, then, does it detect whether a text is AI-generated or human-produced? It does so by identifying two measures: perplexity and burstiness. Perplexity measures a text's randomness. The understanding here is that a human-written text displays randomness or chaoticness and, thus, is likely to perplex or be unfamiliar to a language model such as GPTZero. The higher the perplexity of the text, the higher the likelihood that it is human-written. The converse is true: the lower the text's perplexity, the lower the likelihood that it is human-written. This lower perplexity index signals that a text is AI-generated. Burstiness measures the complexity of sentences or how highly varied sentence usage is in a text. The belief here is that humans are prone to varying the types and the length of their sentences when they write, while machines are not. So, burstiness relates to sentence variability or sentence bursting (Chaka, 2023a; Ofgang, 2023). Most importantly, GPTZero sometimes highlights or flags an AI-generated text in yellow in any given sample and allocates perplexity and burstiness scores to text samples. Higher scores for both measures indicate that a text is human-generated, while lower scores for both measures signal that a text is AI-generated. One of the drawbacks of this tool is that it sometimes misclassifies or misrecognises portions of a text as either AI-generated or human-generated, even in instances where that is not the case (Tyrrell, 2023). So, it is not 100% per cent accurate (Chaka, 2023b).

OpenAI AI Text Classifier

OpenAI AI Text Classifier is an AI detector owned by OpenAI, which also owns ChatGPT. It was released at the beginning of 2023 after the launch of ChatGPT in November 2022. Its main function is to differentiate between AI-generated and human-written text (Eliaçik, 2023b; Ismail, 2023; Tyrrell, 2023). In one of its blogs, its mother tech company asserts that it has "trained a classifier to distinguish between text written by a human and text written by AIs from a variety of providers" (Kirchner et al., 2023, par. 1). It also makes some disclaimers that it is not feasible to fully reliably detect every AI-generated text and that its classifier is not yet fully reliable. It, then, points out that when it tested its classifier in one use case, it had a 26% true positives rate (it correctly identified 26% of AI-generated text) and a 9% false positives rate (it misidentified 9% of human-produced text as AI-generated). According to OpenAI, some of the limitations its text classifier has are as follows:

- Unreliability on shorter texts having fewer than 1,000 characters;
- Only the first 5,000 characters are displayed in the free version;
- Sometimes, the classifier misidentifies longer texts and wrongly labels human-produced text as AI-generated;
- The classifier currently works better on English texts and has a high degree of unreliability on texts written in other languages;
- Unreliability to identify predictable text, especially identifying whether the first 1,000 prime numbers are AI-written or not;
- Edited AI-generated text can evade the classifier; and
- Poor detection of text fine-tuned outside the original training data (Eliaçik, 2023b; Ismail, 2023; Kirchner et al., 2023; Tyrrell, 2023).

Writer.com's AI Content Detector

This AI content detector tool, which is owned by Writer.com, is touted as reliable (Outlook Spotlight, 2023). Unlike most of its peers, it is a no-sign-up or a no-create-an-account tool for usage. It evaluates a text and identifies (by calculating) how much of it is likely AI-generated through percentage scores. It has a 1,500-character limit per text/prompt. Text can be added to this detector by pasting or writing it or by providing a URL of the intended text. The AI tool does not have a 100% accuracy rate, and sometimes, it can be tricked by certain texts (Help Center, 2023; see Lim, 2023). It can also be used for editing and generating text, and its parent company, Writer.com, has offerings such as products (e.g., Grammarly alternative, ChatGPT alternative, and Jasper alternative) and resources (e.g., Inclusive language and AI content generator) (Help Center, 2023; Outlook Spotlight, 2023).

Copyleaks AI Content Detector

Copyleaks AI Content Detector is a free-to-use AI tool that can determine whether a text is generated by AI chatbots like ChatGPT and many others or whether a text is written by a human. According to Copyleaks, this tool has, among others, the following differentiating features:

- A 99.12% detection accuracy rate
- In-depth, detailed analysis
- Detecting GPT-J, GPT-3, GPT-3.5, ChatGPT, GPT-4, and other related AI language models

- Detecting AI content written in multiple languages such as English, Spanish, Polish, Italian, and a few other languages, with more other languages being currently considered

- Verifying the authenticity of social media posts, online news articles, online reviews, etc. (Chrome, 2023; Copyleaks, 2023).

Giant Language Model Test Room

Giant Language Model Test Room (GLTR) is an online tool that employs an algorithm capable of detecting any content related to AI-generated text produced by AI chatbots. It executes a forensic inspection of language model elements on texts to establish whether they are AI- or human-generated. It is supported by a database of predicted words, in which such predicted words are highlighted in green, yellow, and red. The more predicted words a text has, the more likely that it is AI-generated than human-generated. It can also analyse a text for its realness. Its major drawback is that it works better on GPT-2 texts than on GPT-3 texts produced by bots such as ChatGPT (Lim, 2023; Outlook Spotlight, 2023).

All of the five AI content detectors reviewed above were employed in this paper in their free-to-use or non-premium versions. As pointed out earlier, Weber-Wulff et al.'s (2023) paper also evaluated the efficacy of fourteen AI detection tools in their non-premium versions.

Methodology

The paper used an exploratory study design. One key aspect of this study design, which resonates with the present paper, is exploring a topic or an area that has not been studied before (Chaka, 2023a; Elman et al., 2020; Singh, 2021). Evaluating the efficacy, accuracy and reliability of AI content detection tools in differentiating between AI-generated content and human-written content is still a less researched area.

Data collection process

The data collection process for this paper consisted of two stages. In the first stage, the content was generated using ChatGPT, YouChat, and Chatsonic. This content was generated by inputting three sets of English prompts into these three AI chatbots, with each set of prompts for each AI chatbot. The prompts were queried to the three AI chatbots on two different dates. ChatGPT's prompts were queried on 31 January 2023, while the prompts for YouChat, and Chatsonic were inputted on 07 March 2023. This time-lapse was occasioned by the fact that I only became aware of the last two AI chatbots in March 2023 (see Chaka, 2023a). The prompts for these three AI chatbots were based on some of the aspects of applied English language studies (AELS). The latter is one of the areas of my research interests. These prompts are indicated below.

ChatGPT's prompts

- What are decolonial applied English language studies?
- What is critical southern decoloniality?
- Who are the authorities on decolonial linguistics?
- Who are the leading scholars of critical southern decoloniality?
- What is translanguaging?
- What is the difference between translanguaging, multilanguaging, and languaging?

YouChat's prompts

- What are decolonial applied English language studies?
- What is critical southern decoloniality?
- What are Chaka's (2020) views of translanguaging?
- Who are the authorities on decolonial linguistics?
- What is translanguaging?
- What are the latest theories for translanguaging, multilanguaging, and languaging?
- What is the difference between translanguaging, multilanguaging, and languaging?

Chatsonic's prompts

- What is decolonial applied linguistics?
- What are decolonial applied English language studies?
- What is critical southern decoloniality?
- What are Chaka's (2020) views of translanguaging?
- Who are the authorities on decolonial linguistics?
- What is translanguaging?
- What is the difference between translanguaging, multilanguaging, and languaging?

In the second stage, the responses generated from the three AI chatbots were inputted into the five AI content detectors mentioned earlier in three phases from 30 March 2023 to 02 April 2023. During the first phase, the English-only responses were fed into the five AI content detectors. In the second phase, the ChatGPT-generated responses were machine-translated into five languages using Google Translate and inputted into GPTZero. The five languages were German, French, Spanish, Southern Sotho, and isiZulu. The reason for choosing GPTZero for the translated responses is that it was the only AI detector that recognised all five languages at the time of conducting the study. The Southern Sotho that Google Translate uses is the Lesotho orthography of the Sesotho language and not the South African Sesotho orthography. In respect of isiZulu, Google Translate refers to it as Zulu. Henceforth, the paper uses (isi)Zulu to indicate that Zulu has an isi- prefix. During the third phase, the Google-translated German, French, and Spanish responses were fed into Copyleaks AI Content Detector. Currently, Copyleaks AI Content Detector does not support Southern Sotho and (isi)Zulu.

The three sets of AI-generated English responses and the ChatGPT-generated English responses that were Google-translated into the five languages specified above constituted the datasets for this study. After they had been generated and translated, all these datasets were copied and transferred to their respective MS Word files in their original forms. They were not tampered with or manipulated, except that the ChatGPT-generated English responses were Google-translated into the five specified languages. So, they were inputted into the five AI detection tools in their original generated and translated versions.

Results

Detection of the ChatGPT-, YouChat-, and Chatsonic-generated English responses by five AI content detectors

All the ChatGPT-, YouChat-, and Chatsonic-generated responses were subjected to the five AI content detectors, GPTZero, OpenAI AI Text Classifier, Writer.com's AI Content Detector, Copyleaks AI Content Detector, and GLTR for them to detect AI-generated content from these three sets of responses. Concerning ChatGPT-generated responses, all five AI content detection tools yielded their detection results, as illustrated in Table 1. For example, GPTZero correctly classified four texts as AI-generated, while it was indecisive about two texts. Its lowest and highest perplexity scores were 38 and 90. The same kind of classification pattern was yielded by OpenAI AI Text Classifier. Writer.com's AI Content Detector classified five texts inaccurately as human-generated, while its classification of one text was accurate. Its lowest and highest percentages for human-generated content were 1% and 99%. In contrast, Copyleaks AI Content Detector classified five texts accurately, but classified one text inaccurately. Its lowest and highest probability percentages for AI-generated texts were 94% and 99.8%, while its probability percentage for human-generated text was 19.5%. For GLTR, it correctly classified one text as machine-generated, but misclassified five texts. The idea of classified is a proxy for predicted as these tools attempt to predict whether a given text response is AI- or human-generated more than just classifying a given text.

Table 1: Detection of the ChatGPT-, YouChat-, and Chatsonic-generated English responses by five AI content detectors.

Name of AI Content Detector	ChatGPT-Generated Responses[1]	YouChat-Generated Responses	Chatsonic-Generated Responses	Correct detection vs. Incorrect Detection	Indecisiveness	Ranking All the Five AI Content Detectors
Copyleaks AI Content Detector	5 texts' classification (prediction) highly accurate; 1 text's classification inaccurate **NB:** Lowest and highest probability percentages for AI-generated texts: 94% and 99.8% **NB:** Probability percentage for human-generated text: 19.5%	5 texts' classification (prediction) highly accurate; 2 texts' classification inaccurate **NB:** Lowest and highest probability percentages for AI-generated texts: 99.6% and 99.9% (3 texts) **NB:** Lowest and highest probability percentages for human-generated texts: 19.4% and 19.8%	5 texts' classification (prediction) highly accurate; 3 texts' classification inaccurate **NB:** Lowest and highest probability percentages for AI-generated texts: 92.3% and 99.9% (3 texts) **NB:** Lowest and highest probability percentages for human-generated texts: 20% and 98.6%	15 vs. 6	0	1
OpenAI AI Text Classifier	4 texts' classification (prediction) accurate; 2 texts' classification indecisive	3 texts' classification (prediction) accurate; 2 texts' classification indecisive; 2 texts' classification inaccurate	2 texts' classification (prediction) almost accurate; 5 texts' classification indecisive; 1 text's classification inaccurate	9 vs. 3	9	2
GPTZero	4 texts' classification (prediction) accurate; 2 texts' classification indecisive **NB:** Lowest and highest perplexity scores: 38 and 90	1 text's classification (prediction) accurate; 6 texts' classification inaccurate **NB:** Lowest and highest perplexity scores: 40 and 167	3 texts' classification (prediction) accurate; 1 text's classification indecisive; 4 texts' classification inaccurate **NB:** Lowest and highest perplexity scores: 32 and 118	8 vs. 10	3	3
Writer.com's AI Content Detector	5 texts' classification (prediction) inaccurate; 1 text's classification very nearly accurate **NB:** Lowest and highest percentages for human-generated content: 1% and 99%	1 text's classification (prediction) 100% accurate; 3 texts' classification almost accurate; 3 texts' classification inaccurate (one of which was 100% inaccurate) **NB:** Lowest and highest percentages for human-generated content: 0% and 100%	2 texts' classification (prediction) highly and almost accurate; 6 texts' classification inaccurate (two of which were 100% inaccurate) **NB:** Lowest and highest percentages for human-generated content: 2% and 100%	7 vs. 14	0	4
GLTR	1 text machine-generated; 5 texts misclassified	All the 7 texts misclassified	6 texts misclassified; 2 indecisive	1 vs. 18	2	5

Pertaining to YouChat-generated responses, GPTZero classified one text correctly as AI-generated, while it misclassified the six other texts as human-written. Its lowest and highest perplexity scores for all these texts were 40 and 167. OpenAI AI Text Classifier classified three texts accurately as AI-generated but misclassified two texts. It was indecisive about two texts. Writer.com's AI Content Detector correctly classified four texts as AI-generated, but misclassified three texts as human-generated. Its lowest and highest percentages for human-generated content for these texts were 0% and 100%. For its part, Copyleaks AI Content Detector correctly classified five texts as AI-generated and incorrectly classified the other two texts as human-generated. Its lowest and highest probability percentages for AI generated texts were 99.6% and 99.9%, respectively, with three texts having a 99.9% tie. Its lowest and highest probability percentages for human-generated texts were 19.4% and 19.8%. In contrast to the other four AI detectors, GLTR misidentified all seven texts as human-written.

With reference to Chatsonic-generated responses, GPTZero correctly classified three texts as AI-generated, but misclassified four texts as human-generated. It was indecisive about one text. It recorded the lowest and highest perplexity scores for these eight texts as 32 and 118. OpenAI AI Text Classifier identified two texts correctly as AI-generated but detected one text incorrectly. It was indecisive about five more texts. Writer.com's AI Content Detector correctly identified two texts as AI-generated but misidentified six texts as human-written. Its lowest and highest percentages for human-generated content were 2% and 100%. In this regard, Copyleaks AI Content Detector correctly classified five texts as AI-generated but misclassified three texts as human-written. Its lowest and highest probability percentages for AI-generated texts were 92.3% and 99.9%, with three texts having a 99.9% tie. However, its lowest and highest probability percentages for human-generated texts were 20% and 98.6%, respectively. Again, in contrast to the other four AI detectors, GLTR misclassified six texts, while it was indecisive about two texts.

When the five AI content detectors were judged together in terms of their overall correct identification of the three sets of responses generated by the three AI tools, they rank as follows: Copyleaks AI Content Detector (1); OpenAI AI Text Classifier (2); GPTZero (3); and Writer.com's AI Content Detector (4), and GLTR (5) (see Table 1).

Detection by GPTZero of the ChatGPT-generated responses Google-translated into German, French, Spanish, Southern Sotho, and (isi)Zulu

In this section, what is at issue is not the accuracy and correctness of the Google-translated texts for all five languages but rather GPTZero's ability to classify them correctly as machine-generated texts. All the ChatGPT-generated responses were subjected to GPTZero for it to detect if they were AI-generated or human-written (see Table 2). GPTZero incorrectly classified all the translated texts in all five languages as human-written. Its high and lowest perplexity scores for the texts translated into each of these languages were as follows: 110 and 2,478 (German); 57 and 221 (French); 108 and 361 (Spanish); 602 and 1,715 (Southern Sotho); and 651 and 938 ((isi)Zulu).

Table 2: Detection by GPTZero of the ChatGPT-generated responses Google-translated into German, French, Spanish, Southern Sotho, and (isi)Zulu.

Name of AI Content Detector	Google-German-Translated ChatGPT-Generated Responses	Google-French-Translated ChatGPT-Generated Responses	Google-Spanish-Translated ChatGPT-Generated Responses	Google-Southern Sotho-Translated ChatGPT-Generated Responses	Google-(isi)Zulu-Translated ChatGPT-Generated Responses
GPTZero	All 6 the texts' classification (prediction) inaccurate **NB:** All high perplexity scores, with the lowest and highest scores being 110 and 2,478.	All 6 the texts' classification (prediction)inaccurate **NB:** Lowest and highest perplexity scores being 57 and 221.	All 6 the texts' classification (prediction) inaccurate **NB:** Lowest and highest perplexity scores being 108 and 361.	All 6 the texts' prediction inaccurate **NB:** All high perplexity scores, with the lowest and highest scores being 602 and 1,715.	All 6 the texts' classification (prediction) inaccurate **NB:** All high perplexity scores, with the lowest and highest scores being 651 and 938.

Detection by Copyleaks AI Content Detector of the ChatGPT responses Google-translated into German, French, and Spanish

Similarly, here, all the ChatGPT-generated responses, which were Google-translated into the three languages as mentioned above, were subjected to Copyleaks AI Content Detector for it to detect whether they were AI-generated or not (see Table 3). This AI content detector correctly classified three German-translated texts as AI-generated but misclassified three texts as human-written. Its lowest and highest probability percentages for AI-generated texts were 83.7% and 99.9%, while its lowest and highest probability percentages for human-generated texts were 14.8% and 53.4%. It, then, correctly identified five French-translated texts as AI-generated but misidentified one text as human-written. Here, its lowest and highest probability percentages for AI-generated texts were 94.9% and 99.9%, with three texts having a tie at 99.9%. Its probability percentage for the human-generated text was 6.9%. Lastly, Copyleaks AI Content Detector correctly classified all the Spanish-translated texts as AI-generated. Its lowest and highest probability percentages for these texts were 99. 5% and 99.9%, with two texts and four texts tied at 99.5% and 99.9%, respectively.

Table 3: Detection by Copyleaks AI Content Detector of the ChatGPT responses Google-translated into German, French, and Spanish.

Name of AI Content Detector	Google-German-Translated ChatGPT-Generated Responses	Google-French-Translated ChatGPT-Generated Responses	Google-Spanish-Translated ChatGPT-Generated Responses
Copyleaks AI Content Detector	2 texts' classification (prediction) highly accurate; 1 text's classification almost accurate; 3 texts' classification inaccurate **NB:** Lowest and highest probability percentages for AI generated texts: 83.7% and 99.9% (2 texts) **NB:** Lowest and highest probability percentages for human-generated texts: 14.8% and 53.4%	5 texts' classification (prediction) highly accurate; 1 text's classification inaccurate **NB:** Lowest and highest probability percentages for AI generated texts: 94.9% and 99.9% (3 texts) **NB:** Probability percentage for human-generated text: 6.9%	All the 6 texts' classification highly accurate. **NB:** Lowest and highest probability percentages for AI generated texts: 99. 5% (2 texts) and 99.9% (4 texts)

Discussion

For ChatGPT-generated responses, Copyleaks AI Content Detector had more correct classifications of the texts than the other four AI detectors. It was followed by GPTZero and OpenAI AI Text Classifier, which were joint second. In terms of misclassifications of texts (incorrect classifications of texts), GLTR topped the other four AI detectors with eighteen misclassifications; it was followed by Writer.com's AI Content Detector with fourteen misclassifications. AI detectors with the joint-most indecisive texts were GPTZero and OpenAI AI Text Classifier. GPTZero had a perplexity score of 90 for one of its AI-generated texts, which is a high score given that AI-generated texts are supposed to have a lower perplexity index compared to human-written texts (Chaka, 2023a; Ofgang, 2023; Tech Desk, 2023; Tyrrell, 2023). This, in a way, highlights an element of shakiness related to equating a high perplexity with human-only-written texts in an instance where machines, too, can generate texts with a high perplexity index (Heel, 2023; Lim, 2023; Wiggers, 2023). Writer.com's AI Content Detector recorded 1% and 99% as its lowest and highest percentages for human-generated content for two texts apiece. This means it identified the first text as 99% AI-generated, while it recognised the second text to be 1% AI-generated. These are two extremely contrasting detection rates for these texts when considering that all the texts in this set were ChatGPT-generated. For Copyleaks AI Content Detector, its probability percentage for a human-generated text was 19.4%, which is a bit high for texts that were exclusively machine-generated.

In relation to YouChat-generated responses, again, Copyleaks AI Content Detector had more correct identifications of the texts in this set than the other four AI detectors. It was followed by OpenAI AI Text Classifier. The other AI detectors had more misclassifications of texts than the correct classifications, with GLTR racking up the most misclassified texts. Only one AI detector (OpenAI AI Text Classifier) had two undecided texts. The highest perplexity score that GPTZero had for this set of texts is 167, which is a very high score for texts that were machine-generated. The concern raised above about a high perplexity as an indicator of human-produced texts, applies here, too (Tech Desk, 2023; Iyer, 2023). In this context, Writer.com's AI Content Detector had 0% and 100% as its lowest and highest percentages for human-generated content: 0% and 100%. As is the case with the previous instance, these are two diametrically opposed detection rates for texts that were ChatGPT-generated. Copyleaks AI Content Detector recorded 19.4% and 19.8% as its lowest and highest probability percentages for human-generated texts. Again, these are high probability percentages for machine-generated texts.

Concerning Chatsonic-generated responses, the same trend as the one characterised above applies with minor variations. For example, Copyleaks AI Content Detector still had the most correct identifications of the texts in this set, but with GPTZero following it. Both Writer.com's AI Content Detector and GLTR had the most joint misidentified texts, followed by GPTZero. OpenAI AI Text Classifier had the most undecided texts. GPTZero had the highest perplexity score of 167, which, again, is a high score for texts that were machine-generated. Writer.com's AI Content Detector recorded the contrasting lowest and highest percentages of 2% and 100% for human-generated content. Copyleaks AI Content Detector's highest probability percentage of 98.6% for human-generated text was the highest ever for these machine-generated texts.

Overall, of the five AI content detectors tested across the three sets of texts, OpenAI AI Text Classifier and GLTR appeared to be most consistent in their detection rates if the indecisiveness of texts and the misclassification of texts are, respectively, used factors. In contrast, Copyleaks AI Content Detector tended to be the most consistent of the five AI content detectors if the correct identification of texts is used as a factor. Moreover, Copyleaks AI Content Detector trumped all the other four AI content detectors for the most correctly classified texts. GLTR had the highest text misclassification rate and could classify only one text correctly. As such, it ranked last among the five AI content detectors. In an instance in which seven AI text detectors, which included OpenAI AI Text Classifier, GPTZero and Copyleaks, were tested to detect AI-generated content created by Claude (a generative AI tool similar to ChatGPT), GPTZero was the top consistent performer. It was followed by ChatGPT and OpenAI AI Text Classifier. The writing samples were based on prompts related to different writing genres (Wiggers, 2023). However, in the current paper, GPTZero was the third-best performing AI content detector (see Table 1). As mentioned earlier, Aremu's (2023) paper that tested the detection capabilities of six AI detection tools found both ZeroGPT and GPTZero to have a higher level of deception robustness and resistance than the other four AI detection tools.

The fact that the five AI content detectors recognised some of the AI-generated texts inputted to them as human-written points to their propensity to false negativity. In their paper, Weber-Wulff et al. (2023) also found that fourteen of the AI detection tools they evaluated were prone to false negatives: they mistook AI-generated texts for human-produced texts. They call this tendency "misattributing" AI-generated texts to humans. This is what the present paper has referred to as misclassification and misidentification.

As mentioned earlier, as regards the ChatGPT responses that were Google-translated into the five aforementioned languages, GPTZero misidentified all of them as human-produced. One major reason it misidentified all these translated texts is the higher perplexity scores it assigned to them. This is particularly the case with the German, Southern Sotho and (isi)Zulu texts, whose highest perplexity scores were 2,478, 1,715 and 938, respectively. This, more than what has been said earlier, highlights the shakiness and, at times, the unreliability of a higher perplexity as an indicator of human-only-written texts. This is more so given that machine-translated texts can have inordinately higher perplexity scores, such as the ones for the five languages in this paper.

In contrast, Copyleaks AI Content Detector correctly identified three of the German-translated texts, five of the French-translated texts, and all the Spanish-translated texts as AI-generated. Again, here, Copyleaks AI Content Detector outperformed GPTZero, an outcome that contrasts with that of Wiggers' testing (2023) of seven AI content detectors in which GPTZero was a top performer.

Even though in the current study no texts were deliberately manipulated through editing, paraphrasing, or effecting an extra space (a space bar) between words, there are studies that have discovered that text manipulation reduces the detection efficacy of AI detection tools. For instance, Cai and Cui (2023) found that effecting a mere single space, what they call an extra space, results in text detection evasion. If this is the case, this points to one of the inherent problems with current AI detection tools: their lack of reliability and consistency in accurately differentiating between AI-generated texts and human-produced texts. Part of this problem might have to do with the algorithmic configuration of these AI detection tools, which assumes that distributional gaps exist between AI-generated and human-written content. Once these distributional gaps are destabilised by, for example, intentionally adding single space characters before commas in AI-generated content, these tools tend to misrecognise the content output (see Cai & Cui, 2023). In addition, Guo et al. (2023) point out that by removing indicating words such as Nope, My take is, and Hmm, from human-written content and by removing I regret to hear that, I'm an AI assistant, and Here are steps to follow, from an AI-generated content, most AI detection tools are likely to be tricked in their detection capability. These inherent algorithmic shortcomings are likely to be there in the premium versions of these AI detection tools as well.

Implications and limitations

With the release of ChatGPT, as a generative AI chatbot, for public use, a lot of AI content detectors were instantly launched, even though others could have been there before the launch of ChatGPT. Some of these AI content detectors, such as GPTZero and OpenAI AI Text Classifier, were specifically intended to detect AI-generated content from ChatGPT (Eliaçik, 2023a, b; Ismael, 2023; Iyer, 2023: Kirchner et al., 2023; Lim, 2023; Ofgang, 2033). Nevertheless, with their instant launch, these AI content detectors seem not yet fully ready to accurately and convincingly detect AI-generated content from machine-generated texts in different contexts. Most of them seem to be beset by the algorithmic shortcomings mentioned above. This is one of the implications emanating from the five AI content detectors tested in this paper. Additionally, the five AI content detectors were not able to distinguish, in clear-cut terms, between AI-generated texts and human-produced texts. All they could do was to make estimates in percentages (e.g., Writer.com's AI Content Detector and Copyleaks AI Content Detector) or in probabilistic terms such as likely (e.g., GPTZero) and probability (e.g., Copyleaks AI Content Detector), or in combined percentages and probabilistic terms (e.g., Copyleaks AI Content Detector). Others, such as GLTR, used estimating histograms. Educational institutions, academic staff, and students are impatiently waiting for an AI content detector that will precisely, accurately, and correctly detect AI-generated and human-written texts every time they apply them to student writing and to academic essay writing. They are not interested in percentage and probabilistic estimates.

A major limitation of this paper is that it used free-to-use AI content detectors or the non-premium versions of some of these AI content detectors. In fact, when the data for this study were collected, all five AI content detectors had only free versions that were available to the public. Nonetheless, most of them now do have premium or paid-for versions. The tricky thing about the premium versions of these AI detection tools is that one has to have a paid-for subscription with them for one to be able to access and use them. This becomes almost impossible if a researcher wants to evaluate more of them at the same time. But it seems implausible that the premiums versions of these AI detection tools are free of the two algorithmic shortcomings mentioned above. Mostly, what their premiums versions boast of are differentiators such as increased word/character counts and uploading multiple full-text files as part of premium benefits. In the main, these differentiators are, at best, mechanical, and, at worst, not game-changing. One of the things needed to help improve the AI detection efficacy of AI tools is improved super-intuitive machine learning algorithms that can detect sophisticated and subtle stylometric patterns built into language use (see Uzun, 2023).

References

Alser, M., & Waisberg, E. (2023). Concerns with the usage of ChatGPT in academia and medicine: A viewpoint. *American Journal of Medicine Open, 9*(100036), 1-2. https://doi.org/10.1016/j.ajmo.2023.100036

Aremu, T. (2023). *Unlocking Pandora's box: Unveiling the elusive realm of AI text detection*. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4470719_code5947956.pdf?abstractid=4470719&mirid=1

Anders, B. A. (2023). Why ChatGPT is such a big deal for education. *C2C Digital Magazine*. https://scalar.usc.edu/works/c2c-digital-magazine-fall-2022---winter-2023/why-chatgpt-is-bigdeal-education

Awan, A. A. (2023). Top 10 tools for detecting ChatGPT, GPT-4, Bard, and Claude. *KDnuggets*. https://www.kdnuggets.com/2023/05/top-10-tools-detecting-chatgpt-gpt4-bard-llms.html

Barnett, S. (2023). ChatGPT is making universities rethink content. *Wired*. https://www.wired.com/story/chatgpt-college-university-content/#intcid=_wired-amp-bottom-recirc_8fe776e7-19f4-4523-9ba8-1b1d23637669_text2vec1

Cai, S., & Cui, W. (2023). *Evade ChatGPT detectors via a single space*. Arxiv. https://arxiv.org/pdf/2307.02599.pdf

Caren, C. (2023). AI writing: The challenge and opportunity in front of education now. *Turnitin*. https://www.turnitin.com/blog/ai-writing-the-challenge-andopportunity-in-front-of-education-now

Ceres, P. (2023). ChatGPT is coming for classrooms. Don't panic. *Wired*. https://www.wired.com/story/chatgpt-is-coming-for-classrooms-dont-panic/#intcid=_wired-bottom-recirc_9c0d2ac5-941b-45c7-b9ac-7ac221fc2e33_wired-content-attribution-evergreen

Chaka, C. (2020). Higher education institutions and the use of online instruction and online tools and resources during the COVID-19 outbreak – An online review of selected U.S. and SA's universities. *Research Square,* 1-46. Preprint. https://doi.org 10.21203/rs.3.rs-61482/v1

Chaka, C. (2023a). Generative AI chatbots - ChatGPT versus YouChat versus Chatsonic: Use cases of selected areas of applied English language studies. *International Journal of Learning, Teaching and Educational Research, 22*(6), 1-19. https://doi.org/10.26803/ijlter.22.6.1

Chaka, C. (2023b). Stylised-facts view of fourth industrial revolution technologies impacting digital learning and workplace environments: ChatGPT and critical reflections. *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.1150499.

Chrome. (2023). *AI content detector – Copyleaks*. https://chrome.google.com/webstore/detail/ai-content-detector-copyl/gplcmncpklkdjiccbknjjkoidpgkcakd

Copyleaks. (2023). *ChatGPT and AI content detection*. https://copyleaks.com/blog/chatgpt-and-ai-content-detection

Cotton, D. R. E., Cotton, P. A., & Shipway, L. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 60,* 1-13. https://doi.org/10.1080/14703297.2023.2190148

Cutcliffe, J. (2022). ChatGPT, chatbots and artificial intelligence in education. *Ditch That Textbook*. https://ditchthattextbook.com/ai/

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://journals.sfu.ca/jalt/index.php/

jalt/article/view/875

Eliaçık, E. (2023a). Best ChatGPT alternatives: What awaits us beyond ChatGPT? *DataConomy.* https://dataconomy.com/2023/02/beyond-chatgpt-alternatives-best-jasper/

Eliaçık, E. (2023b). *AI Text Classifier: OpenAI's ChatGPT detector indicates AI-generated text.* https://dataconomy.com/2023/02/openai-ai-text-classifier-chatgpt-detector/

Gerring, J., Mahomey, J., & Elman, C. (2020). Introduction. In C. Elman, J. Gerring & J. Mahoney (Eds.), *The production of knowledge: Enhancing progress in social science* (pp. 1-14). Cambridge University Press.

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). *How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection.* Arxiv. https://arxiv.org/pdf/2301.07597.pdf

Harris, M. (2022). ChatGPT—The game-changing app every teacher should know about. *Learners Edge*. https://www.learnersedge.com/blog/chatgpt-the-game-changing-app-everyteacher-should-know-about

Heilweil, R. (2022). AI is finally good at stuff, and that's a problem. *Vox*. https://www.vox.com/recode/2022/12/7/23498694/ai-artificial-intelligence-chat-gpt-openai

Help Center. (2023). *AI content detector.* https://support.writer.com/article/205-ai-content-detector

Hern, A. (2022, December 4). AI bot ChatGPT stuns academics with essay-writing skills and usability. *The Guardian.* https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability

Hetler, A. (2023). Bard vs. ChatGPT: What's the difference? *TechTarget.* https://www.techtarget.com/whatis/feature/Bard-vs-ChatGPT-Whats-thedifference

Ismail, M. (2023). OpenAI releases AI classifier to check for Chat GPT written articles, still needs work. *TechNave*. https://technave.com/gadget/OpenAI-releases-AI-classifier-to-check-for-Chat-GPT-written-articles-still-needs-work-33313.html

Iyer, A. (2023). After GPTZero, Turnitin is developing tool to identify AI-generated text. *Analyticsindiamag.* https://analyticsindiamag.com/after-gptzero-turnitin-is-developing-a-tool-to-identify-ai-generated-text/

Kamran, S. (2023). *Microsoft Bing vs Google Bard: ChatGPT cloning battle has begun*. https://techcommunity.microsoft.com/t5/itops-talk/microsoft-bing-vs-googlebard-chatgpt-cloning-battle-has-begun/m-p/3737979

Khalil, M., & Er, E. (2023). *Will ChatGPT get you caught? Rethinking of plagiarism detection.* Arxiv. https://arxiv.org/pdf/2302.04335.pdf

Kirchner, J. H., Ahmad, L., Aaronson, S., & Leike, J. (2023). New AI classifier for indicating AI-written text. *OpenAI*. https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text

Lim, H. (2023). 5 content detection tools to tell if content is written by ChatGPT. *HongKiat*. https://www.hongkiat.com/blog/chatgpt-content-detection-tools/

Ofgang, E. (2023). What is GPTZero? The ChatGPT detection tool explained by its creator. *TechLearning*. https://www.techlearning.com/news/what-is-gptzero-the-chatgptdetection-tool-explained

Outlook Spotlight. (2023). Best AI content detection tools: Free ChatGPT output detector. *Outlook India.* https://www.outlookindia.com/outlook-spotlight/best-ai-content-detection-tools-free-chatgpt-output-detector-news-256773

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice, 20*(2), 1-26. https://doi.org/10.53761/1.20.02.07

Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://journals.sfu.ca/jalt/index.php/jalt/article/view/883

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://journals.sfu.ca/jalt/index.php/jalt/article/view/787

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://journals.sfu.ca/jalt/index.php/jalt/article/view/689

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://journals.sfu.ca/jalt/index.php/jalt/article/view/771

Sample, I. (2023). Science journals ban listing of ChatGPT as co-author on papers. *The Guardian*. https://www.theguardian.com/science/2023/jan/26/science-journals-ban-listing-of-chatgpt-as-co-author-on-papers

Singh, A. (2021). *An introduction to experimental and exploratory research.* https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3789360_code3438088.pdf?abstractid=3789360&mirid=1

Somoye, F. L. (2023). *Can universities detect ChatGPT?* https://www.pcguide.com/apps/can-universities-detect-chat-gpt/

Stokel-Walker, C. (2022). AI bot CHATGPT writes smart essays — Should professors worry? *Nature*. https://doi.

org/10.1038/d41586-022-04397-7.

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching, 6*(1), 31-40. https://journals.sfu.ca/jalt/index.php/jalt/article/view/731

Tech Desk. (2023). *GPTZero to help teachers deal with ChatGPT-generated student essays*. https://indianexpress.com/article/technology/gptzero-app-helpsteachers-catch-chatgpt-8378062/

Tyrrell, J. (2023). *ChatGPT screening: OpenAI text classifier versus GPTZero app*. https://techhq.com/2023/02/chatgpt-screening-openai-text-classifier-versusgptzero-app/

Ventayen, R. J. M. (2023). OpenAI ChatGPT generated results: Similarity index of artificial intelligence-based contents. *Advances in Intelligent Systems and Computing,* 1-6. http://dx.doi.org/10.2139/ssrn.4332664

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O. Šigut, P., & Waddington, L. (2023). *Testing of detection tools for AI-generated text*. arXiv https://doi.org/10.48550/arXiv.2306.15666

Wiggers, K. (2023). Most sites claiming to catch AI-written text fail spectacularly. *TechCrunch*. https://techcrunch.com/2023/02/16/most-sites-claiming-to-catch-ai-written-text-fail-spectacularly/

Wingard, J. (2023). ChatGPT: A threat to higher education? *LinkedIn*. https://www.linkedin.com/pulse/chatgpt-threat-higher-education-dr-jason-wingard

Yeadon, W., Inyang, O-O., Mizouri, A., Peach, A., & Testrow, C. P. (2023). The death of the short-form physics essay in the coming AI revolution. *Physics Education, 58*(035027), 1-13.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence

Ebrahim Mohammadkarimi[A]

[A] Language and Development Center, University of Raparin, Kurdistan, Iraq

Keywords

Academic dishonesty;
Artificial Intelligence (AI);
students' writings;
teachers' perceptions.

Correspondence

Ebrahim.karimi@uor.edu.krd [A]

Article Info

Received 26 June 2023
Received in revised form 16 July 2023
Accepted 16 July 2023
Available online 17 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.10

Abstract

This research study examines teachers' perceptions of academic dishonesty in the writings of EFL students in the context of AI. The study involved 67 teachers who provided their perspectives through questionnaires and interviews. The findings indicate a mixed perception among teachers regarding the benefits of AI technologies for students, with some acknowledging advantages while others expressed concerns about its impact on academic integrity. Teachers unanimously agreed on the negative influence of AI on students' commitment to academic honesty, perceiving it as enabling dishonesty and hindering skill development. The study highlights the role of teachers in detecting AI-generated assignments and emphasizes the need for addressing ethical implications. Strategies identified include problem-solving activities, plagiarism detection tools, and integration of AI in teaching practices. While some teachers acknowledged challenges in detecting AI-related academic dishonesty, the study underscores the importance of comprehensive training and support for teachers to utilize AI effectively while preserving academic integrity. The study concludes by calling for institutions and policymakers to prioritize ethical considerations and develop guidelines for the responsible use of AI in education.

Introduction

Artificial intelligence (AI) is called a "double-edged sword" (Shah, 2023, p. 1) and "a friend yet a foe" (Lim et al., 2023, p. 3). There are both possibilities and problems for educators and students in this age of AI. A troubling rise in instances of academic dishonesty in student papers is one of the most important difficulties that instructors confront today. Students are becoming more innovative in their techniques of duplicating others' work as AI progresses. The issue of academic integrity, and the need for teachers to protect it, is at an all-time high (Kleebayoon & Wiwanitkit, 2023; Lim et al., 2023).

Academic dishonesty is defined as any kind of cheating or unethical behavior in the classroom that breaches fairness and honesty principles (Sevimel-Sahin, 2023, p. 308). AI's advancement has provided students with a wealth of tools for creating realistic-looking homework with little effort. Some of the AI-powered tools that have made academic misconduct simpler than ever for today's students include essay generators, online essay mills, and custom writing services (Crawford et al., 2023).
Teachers have an important role in helping students avoid and address academic dishonesty in their work. Their ability to detect plagiarism or cheating is important to upholding the educational system's standards and authenticity. However, the advancement of powerful AI tools has made it more difficult for academics to consistently uncover incidents of academic dishonesty. Teachers may struggle to distinguish between genuine effort and fraud as a consequence of learners' capacity to adapt AI-generated content to seem as if it was their original work (Farrokhnia et al., 2023; Sullivan et al., 2023).

Furthermore, academic writing is now seen differently as a result of the widespread use of AI technologies in education. Students could be tempted to use these tools as shortcuts to getting good scores or doing assignments on time, which would diminish the value of critical thinking, research abilities, and creativity. The long-term impacts on students' learning and their capacity to acquire critical intellectual and analytical abilities that are crucial to their personal and professional development raise questions in this regard (Rudolph et al., 2023a; Sison et al., 2023).

In addition, it is impossible to ignore the ethical ramifications of AI in education. While AI technologies can improve educational opportunities and speed up academic development, they also have the potential to undermine the ideals of authenticity and integrity that are at the core of education. As teachers attempt to find a balance between using technology for instructional objectives and respecting the norms of academic integrity, they must wrestle with the moral quandaries raised by AI (Creely et al., 2023; Singer et al., 2023).

Teachers must alter their teaching techniques and evaluation methodologies to counteract academic dishonesty in this age of AI. To successfully identify and prevent academic dishonesty, educators should remain up to date on the newest AI technology and approaches. Additionally, they need to build a classroom climate that emphasizes ethical conduct and teaches students about the penalties for academic dishonesty (Chan & Lee, 2023; Koraishi, 2023).

As AI continues to transform the educational environment, instructors confront the problem of tackling academic dishonesty in student writing. The introduction of AI technology has compromised the ideals of academic integrity, making it easier for students to participate in plagiarism and other sorts of cheating. However, instructors may maintain integrity by remaining educated, advocating ethical behavior in the classroom, and conducting proactive interventions (Adiguzel et al., 2023; Kooli, 2023).

The significance of this research lies in studying instructors' views on academic dishonesty in student writing in the AI age. With the fast progress of AI technology, academic misbehavior has grown more sophisticated, providing new hurdles for educators. By analyzing instructors' viewpoints and experiences surrounding this topic, the research intends to add to the current literature on academic integrity. It seeks to give recommendations for educators, educational institutions, and policymakers on how to encourage ethical standards and protect the quality of academic writing in the era of AI.

Literature review

Academic misconduct, encompassing acts of plagiarism and cheating, has long been a pressing concern within educational institutions. However, the emergence of artificial intelligence (AI) technologies has brought forth a transformed landscape of academic dishonesty, presenting novel challenges for educators. In 2023, a substantial number of scholarly articles (e.g. Bishop, 2023; Chan & Hu, 2023; Chan & Tsi, 2023; Chen, 2023; Cotton et al., 2023; Dergaa et al., 2023; Fitria, 2023; Huang & Tan, 2023; Khalil & Er, 2023; Limna et al., 2023; Manley, 2023; Möller, 2023; Perkins, 2023; Rudolph et al., 2023a; Shen et al., 2023; Zhou et al., 2023) have been published, delving into the application of AI, specifically ChatGPT, in aiding students with their writing endeavors. This surge of scholarly contributions underlines the significance and relevance of the subject matter under scrutiny.

The proliferation of AI-powered tools has considerably widened the array of opportunities for students to engage in academic dishonesty. As Manley (2023) asserts, the accessibility of essay mills, paraphrasing software, and other AI-driven resources has facilitated plagiarism and the creation of ostensibly original content by students. Consequently, educators confront the arduous task of identifying and addressing instances of academic misconduct that are growing increasingly sophisticated and evasive.

Teachers often find themselves at the forefront of detecting occurrences of academic dishonesty in students' written work. However, the advent of AI technologies has rendered this responsibility progressively intricate. Cotton et al. (2023) shed light on the challenges faced by teachers in discerning authentic work and content generated with the assistance of AI tools. The utilization of advanced algorithms and natural language processing capabilities inherent in AI tools poses difficulties for educators in identifying instances of plagiarism.

The prevalence of AI tools in education holds implications for pedagogy and students' learning outcomes. An excessive reliance on AI-generated content may impede the cultivation of critical thinking and research skills among students. AI might be perceived by students as a shortcut to academic success, circumventing the requisite intellectual engagement fundamental to authentic learning. Educators must contemplate the potential long-term consequences of AI-driven academic dishonesty on students' academic growth (Chan & Hu, 2023; Chan & Tsi, 2023).

The ethical ramifications of AI in education should not be overlooked. Authors such as Perkins (2023) emphasize the imperative for educators to grapple with the moral dilemmas associated with AI-powered academic dishonesty. While AI technologies offer opportunities for enhanced learning experiences, they can also undermine the principles of integrity and authenticity within the educational realm. Educators must navigate these ethical considerations and ensure responsible and ethical utilization of AI in the learning process.

In addressing academic dishonesty within the AI era, educators assume a vital role in promoting academic integrity. By implementing strategies that foster a culture of honesty and ethical behavior, educators can actively discourage students from engaging in dishonest practices. Zhang et al. (2023) suggest that integrating educational interventions, such as instructing proper citation and referencing techniques, can facilitate students' understanding of the value of original work and the consequences of plagiarism.

Effectively combating academic dishonesty in the AI era necessitates support from educational institutions and policymakers. Opportunities for professional development can equip educators with the knowledge and skills required to identify instances of AI-driven academic misconduct. Furthermore, educational institutions should establish robust academic integrity policies that explicitly address the use of AI tools and the consequences of academic misconduct (Möller, 2023; Perkins, 2023).

In a review article, Rudolph et al. (2023a) explored the applications of ChatGPT and its association with higher education in general and, specifically, assessment, teaching and learning. Following an explanation of ChatGPT's functionality and a summary of its advantages and disadvantages, they concentrated on the technology's implications for higher education and talked about the future of instruction, evaluation, and learning in the context of AI chatbots like ChatGPT. They reviewed applications that are directed towards students, teachers, and systems, as well as possibilities and dangers, and they placed ChatGPT within the framework of current artificial intelligence in education research. They offered advice for students, professors, and higher education institutions in their article's conclusion.

Abdullayeva and Musayeva (2023) investigated how ChatGPT can affect learners' writing abilities. The article provided instances of how ChatGPT has been utilized in education and highlighted the advantages and possible disadvantages of utilizing it to improve students' writing abilities. These instances involve the AI writing teacher M-Write from the University of Michigan as well as AI-powered writing tools such as Grammarly and Hemingway Editor. They concluded that ChatGPT had the potential to change how writing is taught.

In related research, Mohamed (2023) investigated faculty members' perspectives on the possibility of ChatGPT to improve English as a foreign language (EFL) instruction. As the main strategy for gathering data for the research, in-depth interviews with faculty members were conducted. The findings of the interviews showed that the faculty members' views on ChatGPT's effectiveness were divided. While some academics praised ChatGPT for delivering quick and correct answers to a broad variety of queries, others voiced the concern that it would impede students' growth in research and critical thinking abilities and perhaps promote prejudices or false information.

Furthermore, Dergaa et al. (2023) investigated the possible advantages and challenges of natural language processing (NLP) technologies such as ChatGPT in research publications and academic writing, emphasized the ethical considerations present when employing these instruments, and looked at the effect that they might have on the authenticity and credibility of academic work. This research involves a literature evaluation of relevant academic publications published in peer-reviewed journals indexed in the first quartile of Scopus. The search utilized terms such as "natural language processing", "academic writing", "AI-generated text", and "ChatGPT". The evaluation was performed utilizing a quasi-qualitative technique, which involves reading and critically analyzing the sources and finding relevant data to support the study objectives. The study concluded that ChatGPT and other NLP technologies have the potential to boost academic writing and research efficiency. Nevertheless, its utilization raises questions regarding its influence on the authenticity and legitimacy of academic work.

Finally, in Limna et al.'s (2023) empirical study, the perspectives of educators and students on the usage of ChatGPT in education are investigated. These researchers utilized a qualitative research strategy, employing in-depth interviews to acquire data. A purposive sample strategy was used to pick ten instructors and 15 students from various institutions in Thailand. The data obtained were assessed employing content analysis and NVivo. The results indicated that instructors and students usually had a favorable impression of utilizing ChatGPT in education. The chatbot was considered a beneficial tool for delivering fast feedback, addressing queries, and offering assistance to students. Notwithstanding, the data highlighted some issues with the usage of ChatGPT in teaching, including concerns about the accuracy of the information provided by the chatbot and the potential for losing personal interaction with teachers.

Despite the growing recognition of the challenges posed by academic dishonesty in students' writings in the era of AI, there is a notable gap in the literature regarding teachers' reflections on this issue. While existing research has touched upon the changing landscape of academic dishonesty, detection methods, and its impact on pedagogy, there is limited exploration of teachers' personal experiences, perspectives, and strategies in addressing AI-driven academic misconduct. Understanding the unique insights and reflections of teachers can provide valuable information for developing effective interventions, policies, and support systems that specifically cater to their needs in combating academic dishonesty in the AI era. Bridging this gap in the literature will contribute to a more comprehensive understanding of the issue and facilitate the implementation of targeted approaches to maintain academic integrity. Therefore, this investigation seeks to answer the following research questions:

1. How do teachers perceive the use of AI by their students for academic purposes?

2. How do teachers reflect upon the impact of AI on the prevalence of and commitment to academic dishonesty in students' writings?

3. What strategies and approaches do teachers employ to address and deter AI-driven academic dishonesty, and how do they reflect on the effectiveness of these interventions in maintaining academic integrity?

Methodology

Participants

Participants in the study were 67 teachers from different universities across Iraqi Kurdistan. There are more than 30 universities in Iraqi Kurdistan, of which more than half are private. Although availability sampling was used to recruit participants, it was attempted to find a similar number of participants from state and private universities and from both genders. Teachers were teaching EFL students at different levels. They had various academic backgrounds and teaching experiences and were of both genders. Table 1 illustrates the demographic information of these participants.

Table 1. Demographic information of teachers.

N	Teaching experience (average)	University		Gender		Age (average)
		State	Private	Female	Male	
67	8	35	32	30	37	34

Research design

This research triangulated the data collection tools by using a semi-structured interview and a questionnaire. Both questionnaires and interview questions were approved by the ethical committee of the researcher's university. The interviews were conducted both online (for those who were from faraway cities) and face-to-face (for those who were near the researcher) in the English language. The main themes of the interviews were research questions. The questionnaire was designed by the researcher. Its reliability was checked by Cronbach`s alpha, and it was 0.87, which is a good result. It was sent to the participants online via Google Docs. The participants submit their responses to the questionnaire anonymously. The process of collecting data started at the beginning of May 2023 and lasted about three weeks. The questionnaire consisted of three sections. There were six questions in the first section that asked about the general perceptions of teachers toward the use of generative AI technologies such as ChatGPT and Google Bard. The second section included three questions asking about teachers' reflections on the impact of AI on academic dishonesty in students' writings. There were three questions in the last section, which asked about detecting such academic dishonesty.

Data analysis

The responses to the questionnaires were collected from Google Docs. For analyzing the questionnaires, SPSS 25 was used. Of the 67 teachers who responded to the questionnaires, only 23 agreed to participate in interviews. Participants were 14 males and 9 females from both state and private universities. Their average age was 33, and their average teaching experience was seven years. Each interview lasted about 20–25 minutes, and they were audio-recorded with the participants` verbal permission. The recorded data were transcribed, and thematic analysis was used to evaluate the data.

Results

Questionnaire

The questionnaire employed a five-point Likert scale spanning from 'strongly disagree' (1) to 'strongly agree' (5) to measure participants' responses. The results of the responses to the questionnaires are illustrated in three sections. The first six questions show the perceptions of participants about students' use of AI.

According to the responses, teachers overall had positive attitudes toward AI technologies, including ChatGPT and Google Bard. Participants mentioned that they all used these technologies and believed that they were useful. They mostly noted that AI was a great and available resource for students (strongly agree and agree = 79%), and they were aware that their students used it for academic purposes (strongly agree and agree = 83%). In contrast, only some of them believed that AI might make the students better writers (strongly agree and agree = 45%) and that it could threaten their academic performance (strongly agree and agree = 82%). For Item 6, apart from 27% strongly agreeing or agreeing that they would incorporate AI into their teaching approaches in the future, 22% held a neutral position, which seems to be interesting in the questionnaire.

As could be observed from items 7–9, teachers without doubt all agreed (strongly agree and agree) on the negative impact of AI on the academic integrity of their students. They believed that AI made academic dishonesty more accessible and tempting for students and had a negative effect on the development of general and transferrable skills. Therefore, institutions and policymakers should urgently pay attention to the ethical implications of AI-powered academic

Table 2. Teachers' perceptions on the use of AI.

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Perceptions on the use of AI					
1. I personally use AI technologies such as ChatGPT and Google Bard, and I believe they are useful for me.	45 (67%)	22 (33%)	0 (0%)	0 (0%)	0 (0%)
2. I believe generative AI technologies like ChatGPT and Google Bard are great resources for students, given they are accessible 24/7.	21 (31%)	32 (48%)	0 (0%)	6 (9%)	8 (12%)
3. I am aware of using AI technologies such as ChatGPT and Google Bard by my students for academic purposes.	45 (67%)	11 (16%)	2 (4%)	9 (13%)	0 (0%)
4. I believe AI technologies like ChatGPT and Google Bard can help students become better writers.	24 (36%)	6 (9%)	6 (9%)	23 (34%)	8 (12%)
5. I believe AI technologies such as ChatGPT can endanger students' overall academic performance.	38 (57%)	17 (25%)	8 (12%)	4 (6%)	0 (0%)
6. I anticipate incorporating AI technologies such as ChatGPT and Google Bard into my teaching approaches in the future.	6 (9%)	12 (18%)	15 (22%)	18 (27%)	16 (24%)

Table 3. Teachers' reflection about the use of AI.

Reflection					
7. AI has made academic dishonesty more accessible and tempting for students.	39 (58%)	28 (42%)	0 (0%)	0 (0%)	0 (0%)
8. AI will negatively impact my students' development of general or transferrable skills such as collaboration, problem-solving, and leadership qualities.	39 (58%)	28 (42%)	0 (0%)	0 (0%)	0 (0%)
9. The ethical implications of AI-powered academic dishonesty require urgent attention from educational institutions and policymakers.	39 (58%)	28 (42%)	0 (0%)	0 (0%)	0 (0%)

Table 4. Teachers' detection of AI-generated writing.

Detection					
10. Teachers can already accurately identify a student's usage of generative AI technologies to complete an assignment.	2 (3%)	0 (0%)	5 (7%)	34 (51%)	26 (39%)
11. Teachers encounter challenges in identifying instances of AI-enabled academic dishonesty in students' writings.	26 (39%)	34 (51%)	5 (7%)	0 (0%)	2 (3%)
12. Currently, plagiarism detectors, such as Turnitin and Ithenticate can fully detect the writings of AI technologies, including ChatGPT and Google Bard.	4 (6%)	6 (9%)	19 (28%)	26 (39%)	12 (18%)

dishonesty.

The last three items of the questionnaire asked about teachers' detection of AI-generated writing. Nearly all (strongly disagree and disagree = 90%) of the participants admitted that they cannot detect the use of AI-generated writings in assignments, and the same number claimed that they faced challenges in identifying such writings. Finally, only a few percent (strongly agree and agree = 15%) of them believed that they could detect AI-generated writings via plagiarism detection tools such as Turnitin and Ithenticate.

Interviews

In order to gain a deep understanding of the responses, the main sections of the questionnaire were the primary questions in the interview, while several sub-questions were asked. Unlike questionnaires, responses in interviews were clearer and more reasonable. Based on the questions, the following themes are classified as teachers' perceptions toward the use of AI technologies, their reflections on the impact of these technologies on students' academic dishonesty, and detecting this academic dishonesty.

Teachers' perceptions toward the use of AI technologies

Unlike questionnaires, in interviews, teachers mostly acknowledged that AI is beneficial for students. However, there were a few teachers who believed that AI poses certain challenges and risks for students that need to be carefully addressed or even avoided.

> In my opinion, AI is not only useful but also necessary for students; they can learn a lot from it if they don't abuse it (Teacher 21).
>
> For sure, AI technologies like ChatGPT are beneficial for all students (Teacher 3).
>
> AI does not help students because they simply copy and paste; they don`t try to learn from it (Teacher 15).

The impact of AI on the commitment to academic dishonesty

All teachers agreed that factors including easy accessibility to AI and the failure and negligence of teachers in detecting AI-generated assignments led to students' commitment to academic dishonesty.

> ChatGPT, Google Bard, and other AI technologies are always available, easily seducing students to commit plagiarism (Teacher 19).
>
> Since AI-generated assignments cannot be detected easily, most of the students use these technologies for their assignments (Teacher 16).
>
> Some teachers do not pay attention to students' assignments, which have been generated by AI; this leads to more academic dishonesty (Teacher 7).

Strategies to deter AI-driven academic dishonesty

Teachers expressed divergent opinions and presented a variety of strategies to effectively tackle and identify instances of academic dishonesty arising from AI implementation. These strategies encompassed incorporating problem-solving, critical-thinking activities, real-life and personal examples, encouraging students to express their own ideas, using plagiarism checker software such as Turnitin, using AI to write the assignment and comparing with students' writings, incorporating AI in teaching, and writing the first draft in the classroom. However, there were a few teachers who believed that detecting academic dishonesty was challenging.

> There are some strategies, such as giving students assignments that require problem-solving and critical thinking, including their ideas, and providing real and personal examples (Teacher 11).
>
> Recently, Turnitin software can detect AI writing, though partially (Teacher 4).
>
> Teachers should make themselves familiar with AI technologies and can give their own assignments to ChatGPT or Bard, and then they can find out if their students did the same thing (Teacher 8).
>
> We should tell our students that we are aware of such technologies, and even we should incorporate using these technologies in our teaching practice and motivate our students to use them not just to copy from them but to learn from them (Teacher 17).
>
> I usually ask my students to do brainstorming on the topic in class and write the first draft and then complete it at home; for the next session, I compare the draft of brainstorming with the final draft that they do at home. It reduces academic dishonesty (Teacher 14).
>
> It's not always easy to identify academic dishonesty in students' writings because I'm aware that some students use AI to write assignments and then paraphrase them with paraphrasing tools (Teacher 12).

Discussion

The primary aim of this study was to investigate the perceptions of teachers regarding academic dishonesty in the writings of English as a Foreign Language (EFL) students in the context of AI. Employing both questionnaires and interviews, the research sought to elucidate the responses to the key research inquiries.

According to the findings of the survey, a portion of teachers expressed a belief in the advantageous nature of AI technologies for their students, acknowledging their utilization by the students. With the exception of a minority of teachers, these assertions were further validated during the interviews. This outcome aligns with previous studies conducted by Limna (2023) and Mohamed (2023), where participants described these technologies as either a "friend" or a "foe" (Lim et al., 2023, p. 3).

The second part of the questionnaire investigated the impact of using AI to be academically dishonest. Unquestionably, teachers unanimously concurred regarding the adverse influence of AI on the academic integrity of their students. They held the belief that AI had amplified the accessibility and allure of academic dishonesty for students, impeding the cultivation of fundamental general and transferable skills. Consequently, it is imperative for institutions and policymakers to promptly address the ethical implications arising from AI-driven academic dishonesty. Furthermore, the interviews brought to light that the failure of teachers to identify AI-generated assignments is another contributing factor to students engaging in academic dishonesty. These findings agree with various studies (Chan & Hu, 2023; Chan & Tsi, 2023; Dergaa et al., 2023; Manley, 2023), where the collective concern regarding the prevalence of academic dishonesty was evident. Nevertheless, this viewpoint diverges from the perspective that labels AI as a "double-edged sword" (Shah, 2023), as despite certain challenges associated with AI implementation, the study participants, as well as other studies, overwhelmingly acknowledged the numerous advantages that AI offers to students.

The final section of the questionnaire and interview inquired about approaches to address and detect cases of academic dishonesty stemming from AI adoption. These strategies encompassed the integration of problem-solving and critical-thinking exercises, utilizing real-life and personal illustrations, promoting student expression of original ideas, employing plagiarism detection tools like Turnitin, utilizing AI to generate assignments for comparison with students' work, incorporating AI in teaching practices, and engaging in classroom-based initial drafting. Nonetheless, a minority of teachers acknowledged the difficulties associated with detecting instances of academic dishonesty. These findings align closely with the findings of Cotton et al. (2023), who propose strategies including educating students on plagiarism, mandating the submission of initial and final drafts, employing plagiarism detection software, and monitoring and regulating students' utilization of AI. Furthermore, the findings exhibit partial concurrence with the research conducted by Pickell and Doak (2023), who refer to the process of identifying AI-generated writings through plagiarism detection tools as an "endless cat-and-mouse game." This perspective is rooted in the notion that as AI continues to advance, plagiarism detection tools strive to adapt and identify such writings through evolving methods.

Conclusion and recommendations

This research gives useful insights into the perspectives of instructors about academic dishonesty in the writings of EFL students in the setting of artificial intelligence (AI). The results reveal that instructors possess an understanding of the potential benefits connected with AI technology while simultaneously expressing worry over its deleterious consequences for academic integrity. This research underlines the relevance of institutions and legislators addressing the ethical considerations associated with AI-driven academic dishonesty as well as giving aid to instructors in recognizing and avoiding such occurrences.

The solutions presented in this work, including the integration of problem-solving activities and the deployment of plagiarism detection tools, provide realistic ways for efficiently addressing AI-related academic dishonesty. Nonetheless, it is vital to recognize the limitations inherent in this research, such as the comparably small sample size and the lack of student opinions. Subsequent research attempts should strive to enlarge the participant cohort, integrate the opinions of students, and assess the usefulness of the outlined solutions in minimizing AI-driven academic dishonesty.

Overall, this research contributes to the existing body of literature concerning academic dishonesty in the era of artificial intelligence, offering valuable insights for educators, institutions, and policymakers in the advancement of academic integrity and the responsible utilization of AI technologies within the realm of education.

This research on teachers' reflections on academic dishonesty in EFL students' writings in the era of AI has several limitations that need to be considered. The study's participant pool consisted of only 67 teachers, which may not provide a comprehensive representation of the broader population of educators. The perspectives and insights of other teachers who were not part of the study may have been overlooked, potentially limiting the generalizability of the findings.

Another limitation is the exclusion of students' perceptions from the research. By solely focusing on the reflections and viewpoints of teachers, the study overlooks the valuable insights and experiences of the students themselves. Incorporating the perceptions of students could have enriched the analysis and provided a more holistic perspective on the issue of academic dishonesty. These limitations highlight the need for careful interpretation of the study's results and should guide future research endeavors aiming to further investigate teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence.

The findings of this study have several implications for practice and further research. Firstly, the recognition of teachers' perceptions regarding the advantages of AI technologies highlights the need for incorporating these tools effectively into the educational process. Educators should be provided with training and support to harness the potential benefits of AI while also addressing concerns related to academic dishonesty. This implication aligns with the findings of multiple scholarly investigations conducted by Cotton et al. (2023), Rasul et al. (2023), and Rudolph et al. (2023b).

Secondly, the unanimous agreement among teachers on the negative impact of AI on academic integrity emphasizes the urgency of addressing this issue. Institutions and policymakers should prioritize developing ethical guidelines and policies that promote responsible and ethical use of AI technologies in education. This aligns with the scholarly perspective presented by Chan (2023).

Moreover, the recognition of teachers' responsibility in identifying AI-generated assignments as a contributing factor to academic dishonesty emphasizes the significance of enhancing teachers' awareness and proficiency in identifying and addressing instances of plagiarism facilitated by AI. It is crucial to develop professional development initiatives and resources aimed at equipping teachers with the requisite skills and knowledge to effectively identify and prevent academic dishonesty stemming from AI. This is consistent with the implications derived from the study conducted by Rudolph et al. (2023b).

Additionally, the research identifies a number of techniques, such as the incorporation of problem-solving and critical-thinking activities, utilization of plagiarism detection tools, and integration of AI in teaching practices, which offer practical approaches to tackling and mitigating AI-related academic dishonesty. Educators and educational institutions could explore using these tactics, customizing them to their individual settings and needs.

References

Abdullayeva, M., & Musayeva, Z. M. (2023, April). The impact of Chat GPT on students' writing skills: an exploration of AI-assisted writing tools. *International Conference of Education, Research and Innovation 1*(4), 61-66. https://zenodo.org/badge/DOI/10.5281/zenodo.7876800.svg

Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology, 15*(3), ep429. https://doi.org/10.30935/cedtech/13152

Bishop, L. (2023). A computer wrote this paper: What ChatGPT means for education, research, and writing. *Research, and Writing* (January 26, 2023). https://dx.doi.org/10.2139/ssrn.4338981

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*(1), 1-25. https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y., & Lee, K. K. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and Millennial Generation teachers?. *arXiv preprint arXiv:2305.02878*. https://doi.org/10.48550/arXiv.2305.02878

Chan, C. K. Y., & Tsi, L. H. (2023). The AI revolution in education: Will AI replace or assist teachers in higher education?. *arXiv preprint arXiv:2305.01185*. https://doi.

org/10.48550/arXiv.2305.01185

Chen, T. J. (2023). ChatGPT and other artificial intelligence applications speed up scientific writing. *Journal of the Chinese Medical Association, 86*(4), 351-353. https://doi.org/10.1097/JCMA.0000000000000900

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International,* 1-12. https://doi.org/10.1080/14703297.2023.2190148

Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice, 20*(3), 02.

Creely, E., Henriksen, D., & Henderson, M. (2023, March). Artificial intelligence, creativity, and education: Critical questions for researchers and educators. In *Society for information technology & teacher education international conference* (pp. 1309-1317). Association for the Advancement of Computing in Education (AACE).

Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. B. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport, 40*(2), 615-622. https://doi.org/10.5114/biolsport.2023.125623

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International, 15*(1), 1-15. https://doi.org/10.1080/14703297.2023.2195846

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. In *ELT Forum: Journal of English Language Teaching, 12*(1), 44-58. https://doi.org/10.15294/elt.v12i1.64069

Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: writing better scientific review articles. *American Journal of Cancer Research, 13*(4), 1148-1154.

Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking of plagiarism detection. *arXiv preprint arXiv:2302.04335.*

Kleebayoon, A., & Wiwanitkit, V. (2023). Artificial intelligence, chatbots, plagiarism and basic honesty: comment. *Cellular and Molecular Bioengineering, 16*(2), 173-174. https://doi.org/10.1007/s12195-023-00759-x

Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability, 15*(7), 5614. https://doi.org/10.3390/su15075614

Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology, 3*(1), 55-72.

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education, 21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching, 6*(1), 64-74. https://doi.org/10.37074/jalt.2023.6.1.32

Manley, S. (2023). The use of text-matching software's similarity scores. *Accountability in Research, 30*(4), 219-245. https://doi.org/10.1080/08989621.2021.1986018

Mohamed, A. M. (2023). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: Perceptions of EFL Faculty Members. *Education and Information Technologies, 28*(6), https://doi.org/10.1007/s10639-023-11917-z

Möller, A. (2023). An analysis of university academic integrity policies in New Zealand. *Journal of Further and Higher Education, 47*(3), 338-350. https://doi.org/10.1080/0309877X.2022.2130195

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice, 20*(2), 07. https://doi.org/10.53761/1.20.02.07

Pickell, T. R., & Doak, B. R. (2023). *Five ideas for how professors can deal with GPT-3*... For now. Digital commons. https://digitalcommons.georgefox.edu/cgi/viewcontent.cgi?article=1432&context=ccs

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Sevimel-Sahin, A. (2023). Academic integrity in online foreign language assessment: What does current research tell us?. In R. S. Suman, S. Moccia, K. Chinnusamy, B. Singh, & R. Regin (Eds.), *Handbook of research on perspectives in foreign language assessment* (pp. 306-328). IGI Global.

Shah, F. A. (2023). Is Chat-GPT a silver bullet for scientific manuscript writing?. *Journal of Postgraduate Medical Institute, 37*(1), 1-2. https://doi.org/10.54079/jpmi.37.1.3219

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology, 307*(2), e230163. https://doi.org/10.1148/radiol.230163

Singer, J. B., Báez, J. C., & Rios, J. A. (2023). AI creates the message: Integrating AI language learning models into social work education and practice. *Journal of Social Work Education, 59*(2), 294-302. https://doi.org/10.1080/10437797.2023.2189878

Sison, A. J. G., Daza, M. T., Gozalo-Brizuela, R., & Garrido-Merchán, E. C. (2023). ChatGPT: More than a weapon of mass deception, ethical challenges and responses from the human-centered artificial intelligence (HCAI) perspective. *arXiv preprint arXiv:2304.11215*. https://doi.org/10.48550/arXiv.2304.11215

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching, 6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Zhang, Y., Chu, S. K. W., Liu, Y., & Lu, X. (2023). Effects of a hybrid training for plagiarism prevention module on plagiarism-free academic writing in higher education. *Educational Technology & Society, 26*(2), 1–18. https://www.jstor.org/stable/48720992

Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *arXiv preprint arXiv:2303.11812*. https://doi.org/10.48550/arXiv.2303.11812

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Constructing and testing the psychometrics of an instrument to measure the attitudes, benefits, and threats associated with the use of Artificial Intelligence tools in higher education

Muayyad Ahmad[A]
Fadwa Alhalaiqa[B]
Maha Subih[C]

A Professor, Clinical Nursing Department, School of Nursing, University of Jordan
B Professor, Psychiatric/Mental Health Nursing, College of Nursing, Qatar University
C Associate Professor, School of Nursing, Al-Zaytoonah University of Jordan

Keywords

Artificial intelligence;
attitudes;
benefits;
higher education;
reliability;
threats;
validity.

Correspondence

mma4@ju.edu.jo [A]

Article Info

Received 22 September 2023
Received in revised form 2 November 2023
Accepted 3 November 2023
Available online 3 November 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.36

Abstract

Under the acceleration in the body of information regarding AI technology and the paucity of instruments that assess the views and reactions of consumers, we have constructed this instrument to measure the attitudes, benefits, and threats (ABT) toward using Artificial Intelligence (AI) tools in higher education. Google Form was used in August of 2023 to collect data from students and teachers at higher education institutions in 11 Asian and African countries. After the ABT instrument obtained a sufficient score in content validity, additional statistical analyses were done. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were performed. This study included 503 participants who are familiar with AI tools. Over 56% have Bachelor's degrees and 35% have Master's or Doctoral degrees. The most popular AI tool was ChatGPT. One model out of six models created for the factor structure of the 35 items that measure attitudes, benefits, and threats was chosen. The selected model provides the highest explained variance (55.6%). The CFA, using AMOS software, demonstrated that the fit indices were satisfactory for the adopted model. Attitude (15), benefits (6), and threats (14 items) are the three factors of the model. The CFA supports the EFA with the ABT three-factor structure model. The high factor loadings and communalities suggest that the factors are reliable and valid measures of the attitude, benefits, and threats toward AI tools among highly educated personnel.

Introduction

In recent years, Artificial intelligence (AI) has made tremendous advances, resulting in a vast collection of tools and applications (Ismail et al., 2023; Soori et al., 2023). The field of AI-based education and research has entered a brand-new phase of rapid development (Yagi et al., 2023). The enhancement of research and educational efficiency and precision is one of the main advantages (Makeleni et al., 2023). However, few instruments test how individuals perceive, react, and comprehend the new technologies that are continuously growing.

Early identification of a good attitude can assist in establishing the conditions for a successful implementation (Jones et al., 2022). Positive attitudes are usually associated with increased adoption rates (Munianday et al., 2022). A positive mindset can promote the quicker and more effective application of AI tools (Jones et al., 2022). If consumers have a negative attitude, these insights can help designers and developers make the necessary modifications (Lin & Shi, 2022). Analysing attitudes makes it easier to identify ethical difficulties, which is essential for developing AI responsibly. However, the gathering of attitude data may involve sensitive information that could be exploited if not appropriately safeguarded (Almaghrabi & Bugis, 2022). Quantitative threat assessments assist companies in minimising dangers and optimising returns by determining which risks to address first and allocating resources effectively (Żebrowski et al., 2022).

To fully exploit the economic and societal advantages of AI technologies, it is vital to comprehend and quantify their benefits. By understanding the specific benefits, businesses may better align AI projects with their strategic objectives and improve their long-term planning (Allioui & Mourdi, 2023). In a digital world that is always getting bigger, knowing AI strengths could give the person an edge over competitors from all over the world (Duong et al., 2022; Perifanis & Kitsios, 2023). Furthermore, the analysis of AI's positive consequences, such as health gains and environmental benefits, may influence public opinion and legislation (Littman et al., 2022). Having a firm understanding of the benefits and threats enables a more comprehensive approach to threats assessment (Tepylo et al., 2023). If companies know how people feel about their AI products, they can market them better (Haleem et al., 2022). Geographical and cultural differences might be considered when customising AI solutions for various markets (Salo-Pöntinen & Saariluoma, 2022).

In the literature, there are few articles discussing the attitudes toward AI tools. One article, for instance, proposes the development and validation of the AI Attitude Scale (AIAS), a brief self-report instrument designed to assess public perceptions of AI technology (Grassini, 2023). Many reasons necessitate the development of an instrument to assess attitudes toward AI in higher education. First, it can assist educators in comprehending students' attitudes toward AI to create appropriate curricula and educational materials (Moldt et al., 2023). Second, it can assist researchers in analysing the impact of AI on higher education and identifying areas in need of improvement (Escotet, 2023). In our study, we have focused on students at the higher education level and the faculty members as well.

AI tools can analyse student performance and behaviour, identify knowledge gaps, and provide individualised support and feedback to enhance learning outcomes (Alqahtani et al., 2023). Thus, it is necessary to measure the benefits of AI tools in higher education to comprehend their impact on students' learning outcomes and identify improvement areas.

There are numerous articles discussing the threats of AI tools. Two articles, for instance, discuss the threats of using AI for cybersecurity, such as the need for substantial investments in computing power, memory, and data (Hassoulas et al., 2023; Saeed et al., 2023). Other articles discuss the disadvantages of artificial intelligence, such as ethical concerns regarding bias and privacy, security risks posed by hacking, and a lack of human-like creativity and empathy (Huang et al., 2023; Wach et al., 2023). However, finding the threats associated with using AI tools from the viewpoints of students and faculty members is essential in higher education. Therefore, we used a systematic strategy to explore the literature in order to develop an instrument that measures the attitudes, benefits, and threats related to the use of AI tools in higher education among students and teachers.

This study's purposes were to: 1) construct an instrument to measure attitudes, benefits and threats (ATB); 2) examine the factor structure of the ATB instrument using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

Methods

Participants and settings

Data were collected from 503 students and staff members at higher education institutions in Asia and Africa using Google Forms during August 2023. Participants came from 11 different countries, including two from Africa (Egypt and Sudan) and nine from Asia (Jordan, Palestine, Lebanon, Saudi Arabia, Iraq, Thailand, India, Philippines, and Kuwait). The eligibility criteria were being a graduate or undergraduate student or a faculty member at a university. Participants were required to be able to read English because the instrument was written in English.

Ethical considerations

The study was authorised by the Institutional Review Board (IRB) at the School of Nursing/University of Jordan. The first page of the questionnaire contains information regarding the research purpose, methodology, participants' right to decline participation, and assurance of confidentiality. An email address was provided for members of the study team to receive and respond to inquiries from anticipated participants. Informed consent was gained by selecting "yes" in response to the question "Are you willing to participate in this study?" The data were saved on the desktop of the principal investigator (PI), and only approved members of the study team had access to them.

Instrument

The research instrument has two components—first, the sociodemographic and personal characteristics. Participants' age, gender, level of education, frequency of AI tool use, and nationality were collected. The second component consists of three subscales evaluating attitudes, benefits and threats of using AI tools. The research team developed the instrument to measure the ABT associated with teachers' and students' use of AI tools in higher education settings. The research team did a comprehensive evaluation of the literature, and then each member of the team extracted and categorised essential features under the titles' attitudes, benefits, and threats. The three proposed drafts were combined, and redundant text was removed. Following this, psychometric tests were conducted.

Data analysis

For descriptive statistics and EFA, IBM SPSS 29.01 was used (IBM, 2023b). IBM AMOS 26.0 was used to develop the CFA using structural equation modeling (IBM, 2023a). Data are visualised in tables and figures.

Psychometrics of the instrument

Seven items assessed the benefits of AI technologies, 16 items assessed the threat, and 17 items assessed the attitudes. Three professionals in higher education were consulted to obtain the content validity index (CVI): one in computer technology and artificial intelligence, one is a professor in nursing with a subspecialty in health informatics, and one is a professor in medical education. The panel of experts assessed the applicability of each item on the instrument. The CVI is then calculated using the average of the expert assessments. Five items were eliminated from the study because their CVI scores were below 0.70 or they were irrelevant. The remaining 35 items were reviewed by five specialists, including three from the initial panel and two from the physics and sociology departments. Each item's minimum score was 0.85, and the overall CVI score for the scale was 0.95. Each item was scored using a 5-point Likert scale ranging from strongly disagree (0) to strongly agree (4).

Exploratory factor analysis (EFA) and CFA were used to test the construct validity of the study scale. The 35 items were divided into three subscales: Attitudes (15 items), benefits (6 items), and threats (14 items). The overall explained variance for this study was 55.6%. The Cronbach's alpha coefficient was calculated for each of the three subscales and for the overall scale. The benefits subscale score was 0.82, the threat subscale score was 0.91, and the attitudes subscale score was 0.90. In addition, the scale's overall reliability was 0.93.

Results

Participants in this study were highly educated and came from 11 different countries in Asia and Africa, with the majority coming from the Middle East. There were a total of 503 participants. About 56% of them have a Bachelor's degree, and over 35% have a Master's or Doctoral degree. Women constituted almost 58% of the sample. Almost a quarter of the sample reported using AI technologies on a daily or weekly basis. The participants' ages ranged from 18 to 69 years (Table 1).

Table 1: Descriptive statistics for the study sample (N=503).

Characteristics	Frequency	%
Age (Mean=30.9 , SD=11.3)		
Gender		
Male	246	48.9
Female	257	51.1
Education		
Diploma	37	7.4
Bachelor	266	52.9
Master	108	21.5
PhD	92	18.3
Frequency of using AI tools		
Daily	35	7.4
Weekly	88	17.5
Monthly	62	12.3
Rarely	316	62.8
Country		
Egypt	30	6.0
India	17	3.4
Iraq	53	10.5
Jordan	136	27.0
Kuwait	63	12.5
Lebanon	100	19.9
Palestine	67	13.3
Philippine	5	1.0
Saudi Arabia	15	3.0
Sudan	17	3.4

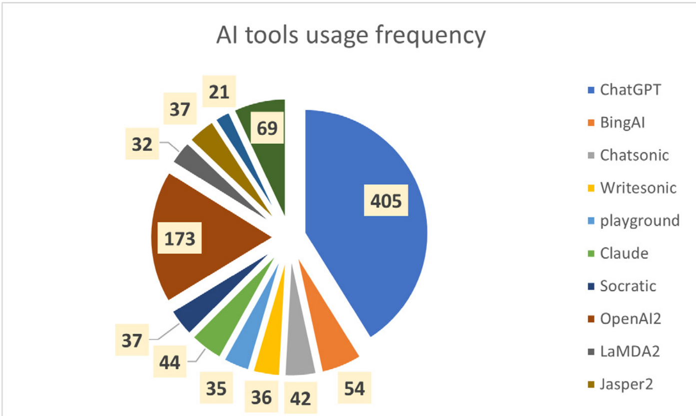


Figure 1: The 12 AI tools frequency usage among the study sample.

Exploratory factor analysis

Exploratory Factor Analysis (EFA) is a statistical technique similar to Principal Component Analysis (PCA) that is used to reduce data from numerous variables to fewer dimensions (Vitoratou et al., 2023). Both are utilised for dimensionality reduction, but their approaches and interpretations are fundamentally distinct (Schreiber, 2021). Principal Component Analysis aims to maximise variance and does not concern itself with explaining the data. It transforms the original variables into a new set of uncorrelated variables (principal components). Furthermore, Principal Axis Factoring (PAF) seeks to uncover latent links (‘factors’ or latent variables) between observed variables. In contrast to PCA, it is intended to model the underlying structure, and is typically used to identify a theory or construct (Schreiber, 2021).

The Principal Component Analysis (PCA) and Principal Axis Factoring (PAF) were utilised with various rotation settings, in addition to using Eigenvalues greater than 1 and limiting the number of output factors to three (Table 2). Most factors with high item loadings, clean loading (difference between two loadings on the same factor should be greater than 0.20), and good overall model fit constitute the best EFA model (Liao et al., 2023).

Therefore, in this study, we have chosen model six in Table 2. Model six was conducted through PCA with Oblimin rotation and Kaiser Normalisation. The model has 55.6% of the total variance explained. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .93. The Bartlett's Test of Sphericity had a Chi-Square = 8169 ($p<.001$). A significant Bartlett's test of sphericity (p-value < 0.05) indicates that the correlation matrix is not an identity matrix. The two items with unclean loading were allocated to the suitable factor based on theoretical reasoning (Dautle & Farrell, 2023). Thus, one was allocated under the threats factor and the other under the attitudes factor.

Table 2: Descriptive statistics for the 6 models in EFA.

Model	Extraction and Rotation	Extraction Either Eigenvalue or fixed number of factors	Number of factors	Total variance explained	Number of items with un-clean loading^
1	PCA, No Rotation	Eigenvalue>1	6	54.3%	20
2	PCA, No Rotation	Fixed number=3	3	47.8%	7
2	PCA, Varimax with Kaiser Normalisation	Eigenvalue>1	6	55.3%	8
3	PCA, Varimax with Kaiser Normalisation	Fixed number=3	3	47.1%	3
4	PCA, Equamax with Kaiser Normalisation	Fixed number=3	3	47.1%	9
5	Principal Axis Factoring, Equamax with Kaiser Normalisation	Fixed number=3	3	42.3%	10
6	PCA, Oblimin with Kaiser Normalisation	Fixed number=3	3	55.6%	2

^Unclean loading means the difference in loading between the same items is less than .20

Table 3 presents the means and standard deviations for all the study items. The range of means for each of the 35 items could range from 0 to 4.

Table 4 presents the loading of items on the three factors of the instrument. The analysis was conducted through PCA with Oblimin rotation and Kaiser Normalisation. The three factors are labelled as attitudes (15 items), benefits (6 items), and threats (14 items). The loadings for the three factors were significant. The commonalities for all factors were also high, suggesting that the factors explained a significant amount of the variance in the observed variables.

Table 3: The means and standards deviation for the 35 items in the instrument.

Items*	Mean	SD
B1 Easy to use	2.78	.914
B2 Save time	2.97	.906
B3 Accessible with low cost	2.69	.961
B4 Help students to ask questions and interact with the material at their own pace	2.67	.956
B5 AI tools are user-friendly	2.59	.842
B6 I know that AI tools are used in education and research	2.80	.986
T1 Lack of human interaction	2.76	1.045
T2 Legal issue (e.g. copyright issues, authorship)	2.63	1.044
T3 Decrease creativity and critical thinking	2.82	1.059
T4 AI tools does not replace practical training	2.80	.982
T5 Security concerns	2.59	1.025
T6 Technical issue	2.51	.994
T7 Over-reliance on technology	2.71	1.028
T8 Ethical dilemma/concerns such as plagiarism	2.58	.951
T9 Threats of AI tools: Need Internet all the time	2.87	.994
T10 Difficulty in handling complex task in research	2.50	.961
T11 Threats of AI tools: Inaccurate/incorrect or biased information	2.41	.940
T12 Over-detailed, redundant, excessive content	2.34	.967
T13 Using AI tools will reduce skills and abilities of person who use it (e.g., writing skills, critical thinking …..etc)	2.67	.991
T14 I see AI tools as a threat to human ethics	2.34	1.019
A1 AI tools content can be used if properly cited and documented	2.71	.875
A2 Authors should have proper knowledge on how to use AI tools	2.70	.895
A3 I recommend AI tools to a friend or colleague		
A4 I'm interested in using of a premium version of AI tools with advanced features	2.43	.977
A5 AI tools has a positive impact on my education/learning	2.51	.888
A6 There is a need for specific training on how to use AI tools in order for them to be useful.	2.63	.964
A7 I suggest providing adequate information on establishing ethical guidelines for the use of AI tools.	2.68	.920
A8 I think AI tools should be included in the study curricula	2.41	1.003
A9 To improve AI applications in the real world, it is essential to encourage researchers to be honest and transparent about their methods.	2.76	.886
A10 I review and edit the response that generated by AI tools before using them in my work	2.71	.869
A11 AI tools can be listed as an author based on its significant contribution	2.45	.958
A12 I feel comfort with ethical and responsible use of AI-generated content from AI tools.	2.34	.965
A13 AI tools could enhance research (e.g., assisting the researchers in framing the sentences, improving the content drafted by the authors.	2.63	.902
A14 I think the responses generated by AI tools are overall easy and coherent	2.68	.988
A15 I trust the information that I read and see on AI tools?	2.21	.916

*B=Benefits, T=Threats, A=Attitude; SD=Standard deviation.

Table 4: Pattern matrix and items loading on the three factors.

Items*	Attitudes	Threats	Benefits
A8	.701		
A1	.696		
A4	.694		
A5	.686		
A3	.676		
A11	.623		
A2	.607		
A6	.604		
A12	.591		
A7	.566		
A15	.551		
A9	.518		
A10	.476		
A13	.451		
A14	.402		
T1		-.755	
T2		-.748	
T6		-.733	
T5		-.707	
T3		-.688	
T4		-.666	
T11		-.659	
T7		-.655	
T12		-.635	
T10		-.627	
T8		-.556	
T14		-.507	

T9	-.507	
T13	-.497	
B1		-.782
B2		-.724
B3		-.697
B6		-.626
B4		-.477
B5		-.402

*The letter and number correspond to the same items in Table 3.

Confirmatory Factor Analysis

The CFA was conducted using AMOS 26.0 (IBM, 2023a). The fit indices were all within acceptable ranges, suggesting that the model fit the data. The fit indices for the study instrument are presented in Table 5.

Table 5: Fitting indices for the 3-factor model.

Fitting index	Index Value	Thresholds
Chi-squared test of model fit (**CMIN**)	2053.557*	
CMIN/DF	3.687*	< 2 to < 5
Root mean square error of approximation (RMSEA)	.07	< 0.05 (good), 0.05–0.08 (acceptable) < 0.10 (poor but sometimes acceptable)
Goodness-of-fit index (GFI)	.92	> 0.90 (acceptable) > 0.95 (good)
Adjusted goodness-of-fit index (AGFI)	.91	> 0.80 (acceptable) > 0.90 (good)
Comparative fit index (CFI)	.90	> 0.90 (acceptable) > 0.95 (good)
Tucker-Lewis index (TLI)	.90	> 0.90 (acceptable) > 0.95 (good)

*p<.001

With confirmatory factor analysis, one can determine the efficiency of the construct. It is a crucial phase and analysis in structural equation modelling (SEM). Standardised Confirmatory Factor Analysis for the three factors with a 35-item structure model is depicted in Figure 2.

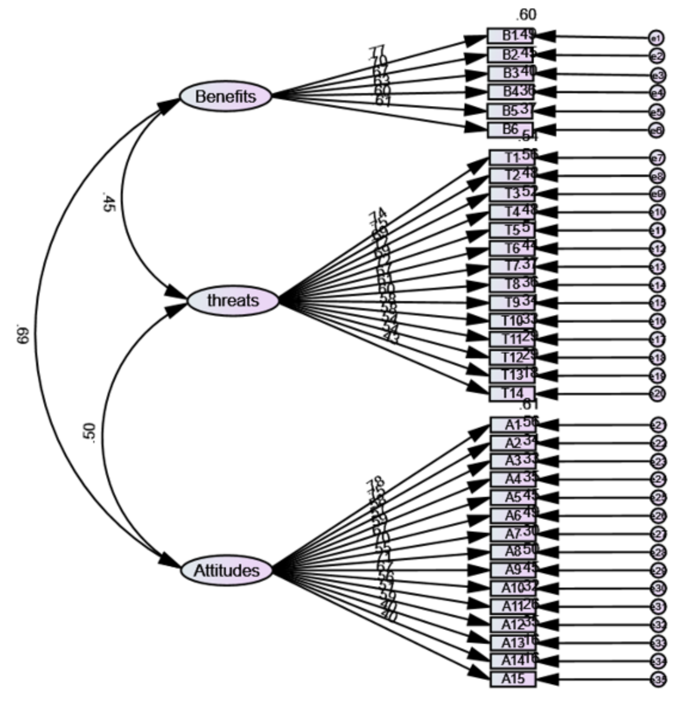


Figure 2: Standardised Confirmatory Factor Analysis for the three factors 35-item structure model.

Discussion

In this study, we have constructed the ABT instrument and examined its psychometric properties across a large, heterogeneous sample of university students and faculty members from 11 Asian and African nations. It is projected that the use of AI tools will continue to grow worldwide. Better customer experiences can be offered through the deployment of AI technologies that can be personalised to give each client the information and services they require (Chaturvedi & Verma, 2023), which may increase customer satisfaction (Chaka, 2023; Cui & van Esch, 2023).

The existing literature is lacking in providing measuring instruments for the perceptions of university students and academic staff toward AI technologies (Mantello et al., 2023). Thus, it was necessary to develop the ABT instrument to contribute to the body of knowledge in this rapidly developing field.

Using standard EFA approaches, a preliminary investigation of the measurement properties of the scale was done. This method is suitable for the first phases of empirical research when exploration is the major objective, and there are no theoretical models available (Mantello et al., 2023). Consequently, it produces more precise data on the acceptability of the specified instrument. However, exploratory factor models do not generate explicit test statistics for assessing convergent and discriminant validity like CFA does (Ahmad et al., 2018). Therefore, the CFA methodology of structural equation modelling was used for measuring unobserved (latent) variables (Dhaene & Rosseel, 2023; Navandar et al., 2023).

The three-factor model of this study has explained more than half of the variance (55.6%), the highest proportion among the six models under EFA. Additionally, the items with clean loading were superior to the other models. The CFA has supported the ABT structural model examined in this study. Moreover, the internal consistency coefficients for the three subscales and the entire instrument were high. Therefore, the authors of this study recommend administering the ABT to students and teachers in higher education to gauge their attitudes, benefits, and threats toward AI tools.

Conclusion

The ABT instrument structure was examined using both EFA and CFA methodologies. It was determined that the 35-item scale with the three-factor model is concise, valid, and empirically verified. The findings of this study can be used to assess the attitudes, benefits, and threats toward AI tools among students and faculty members at high education levels, and possibly other sectors in the community.

References

Ahmad, M. M., Alzayyat, A., & Al-Gamal, E. (2018). Psychometric properties of the coping behavior inventory among students of Arabic universities. *Journal of Nursing Measurement, 26*(1), 176-185. https://doi.org/10.1891/1061-

3749.26.1.176

Allioui, H., & Mourdi, Y. (2023). Unleashing the potential of AI: Investigating cutting-edge technologies that are transforming businesses. *International Journal of Computer Engineering and Data Science (IJCEDS), 3*(2), 1-12. https://ijceds.com/ijceds/article/view/59

Almaghrabi, N. S., & Bugis, B. A. (2022). Patient confidentiality of electronic health records: A recent review of the Saudi literature. *Dr. Sulaiman Al Habib Medical Journal, 4*(3), 126-135. https://doi.org/10.1007/s44229-022-00016-9

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., & Al Yami, M. S. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy, 19*(8), 1236-1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching, 6*(2), 1-11. https://doi.org/10.37074/jalt.2023.6.2.12

Chaturvedi, R., & Verma, S. (2023). Opportunities and challenges of AI-driven customer service. In: Sheth, J. N., Jain, V., Mogaji, E., & Ambika, A (Eds.), *Artificial Intelligence in customer service: The next frontier for personalized engagement* (pp. 33-71). Palgrave Macmillan.

Cui, Y., & van Esch, P. (2023). Artificial Intelligence in customer service strategy for seamless customer experiences. In: Sheth, J. N., Jain, V., Mogaji, E., & Ambika, A (Eds.), *Artificial Intelligence in customer service: The next frontier for personalized engagement* (pp. 73-97). Palgrave Macmillan.

Dautle, S., & Farrell, S. (2023, June). Using EFA to determine factor structure of a computer-based version of the purdue spatial visualization test: Rotations (PSVT: R). In *2023 ASEE Annual Conference & Exposition*.

Dhaene, S., & Rosseel, Y. (2023). An evaluation of non-iterative estimators in the Structural after Measurement (SAM) approach to Structural Equation Modeling (SEM). *Structural Equation Modeling: A Multidisciplinary Journal,* 1-15. https://doi.org/10.1080/10705511.2023.2220135

Duong, P. N., Kien, D. T., & Khoa, B. T. (2022). Do knowledge management systems motivate and satisfy the academic staff in higher education Institutions? *Journal of Applied Learning and Teaching, 5*(2), 84-94. https://doi.org/10.37074/jalt.2022.5.2.16

Escotet, M. Á. (2023). The optimistic future of Artificial Intelligence in higher education. *Prospects,* 1-10. https://doi.org/10.1007/s11125-023-09642-z

Grassini, S. (2023). Development and validation of the AI attitude scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1191628

Haleem, A., Javaid, M., Qadri, M. A., Singh, R. P., & Suman, R. (2022). Artificial intelligence (AI) applications for marketing: A literature-based study. *International Journal of Intelligent Networks, 3*, 119-132. https://doi.org/10.1016/j.ijin.2022.08.005

Hassoulas, A., Powell, N., Roberts, L., Umla-Runge, K., Gray, L., & Coffey, M. (2023). Investigating marker accuracy in differentiating between university scripts written by students and those produced using ChatGPT. *Journal of Applied Learning & Teaching, 6*(2), 1-7. https://doi.org/10.37074/jalt.2023.6.2.13

Huang, R., Tlili, A., Xu, L., Ying, C., Zheng, L., Metwally, A. H. S., Ting, D., Chang, T., Wang, H., & Mason, J. (2023). Educational futures of intelligent synergies between humans, digital twins, avatars, and robots-the iSTAR framework. *Journal of Applied Learning & Teaching, 6*(2), 1-16. https://doi.org/10.37074/jalt.2023.6.2.33

IBM. (2023a). *IBM AMOS statistics for Windows, Version 26.0. Armonk*. IBM Corp.

IBM. (2023b). *IBM SPSS statistics for Windows, Version 29.0. Armonk*. IBM Corp.

Ismail, F., Tan, E., Rudolph, J., Crawford, J., & Tan, S. (2023). Artificial intelligence in higher education. A protocol paper for a systematic literature review. *Journal of Applied Learning and Teaching, 6*(2), 1-8. https://doi.org/10.37074/jalt.2023.6.2.34

Jones, O., Matin, R., van der Schaar, M., Bhayankaram, K. P., Ranmuthu, C., Islam, M., Behiyat, D., Boscott, R., Calanzani, N., & Emery, J. (2022). Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: A systematic review. *The Lancet Digital Health, 4*(6), e466-e476. https://doi.org/10.1016/S2589-7500(22)00023-1

Liao, J., Zhong, T., Cai, R., Tang, D., Sun, W., Yu, R., Zhou, L., & Wan, L. (2023). Psychometric validation of the Chinese version of the Shirom-Melamed burnout questionnaire among parents of children with cancer. *Journal of Pediatric Nursing, 69*, 101-107. https://doi.org/10.1016/j.pedn.2023.01.004

Lin, B., & Shi, L. (2022). Do environmental quality and policy changes affect the evolution of consumers' intentions to buy new energy vehicles. *Applied Energy, 310*, 118582. https://doi.org/10.1016/j.apenergy.2022.118582

Littman, M. L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., Hadfield, G., Horowitz, M. C., Isbell, C., & Kitano, H. (2022). *Gathering strength, gathering storms: The one hundred year study on artificial intelligence (AI100) 2021 study panel report.* https://doi.org/10.48550/arXiv.2210.15767

Makeleni, S., Mutongoza, B. H., & Linake, M. A. (2023). Language education and Artificial Intelligence: An exploration of challenges confronting academics in global south universities. *Journal of Culture and Values in Education, 6*(2), 158-171. https://doi.org/10.46303/jcve.2023.14

Mantello, P., Ho, M.-T., Nguyen, M.-H., & Vuong, Q.-H. (2023). Bosses without a heart: Socio-demographic and cross-cultural determinants of attitude toward emotional AI in the workplace. *AI & Society, 38*(1), 97-119. https://doi.org/10.1007/s00146-021-01290-1

Moldt, J.-A., Festl-Wietek, T., Madany Mamlouk, A., Nieselt, K., Fuhl, W., & Herrmann-Werner, A. (2023). Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. *Medical Education Online, 28*(1), 2182659. https://doi.org/10.1080/10872981.2023.2182659

Munianday, P., Radzi, A. R., Esa, M., & Rahman, R. A. (2022). Optimal strategies for improving organizational BIM capabilities: PLS-SEM approach. *Journal of Management in Engineering, 38*(3), 04022015. https://doi.org/10.1061/(ASCE)ME.1943-5479.000103

Navandar, Y. V., Patel, D., Dhamaniya, A., Velmurugan, S., & Bari, C. (2023). Users perception based service quality analysis at toll plazas using structural equation modeling. *Case Studies on Transport Policy, 13,* 101053. https://doi.org/10.1016/j.cstp.2023.101053

Perifanis, N.-A., & Kitsios, F. (2023). Investigating the influence of artificial intelligence on business value in the digital era of strategy: A literature review. *Information, 14*(2), 85. https://doi.org/10.3390/info14020085

Saeed, S., Altamimi, S. A., Alkayyal, N. A., Alshehri, E., & Alabbad, D. A. (2023). Digital transformation and cybersecurity challenges for businesses resilience: Issues and recommendations. *Sensors, 23*(15), 6666. https://doi.org/10.3390/s23156666

Salo-Pöntinen, H., & Saariluoma, P. (2022). Reflections on the human role in AI policy formulations: How do national AI strategies view people? *Discover Artificial Intelligence, 2*(1), 3. https://doi.org/10.1007/s44163-022-00019-3

Schreiber, J. B. (2021). Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy, 17*(5), 1004-1011. https://doi.org/10.1016/j.sapharm.2020.07.027

Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, A review. *Cognitive Robotics, 3*, 54-70. https://doi.org/10.1016/j.cogr.2023.04.001

Tepylo, N., Straubinger, A., & Laliberte, J. (2023). Public perception of advanced aviation technologies: A review and roadmap to acceptance. *Progress in Aerospace Sciences, 138,* 100899. https://doi.org/10.1016/j.paerosci.2023.100899

Vitoratou, S., Uglik-Marucha, E., Hayes, C., & Pickles, A. (2023). *A comprehensive guide for assessing measurement tool quality: The Contemporary Psychometrics (ConPsy) Checklist.* https://doi.org/10.31234/osf.io/t2pbj

Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J., & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review, 11*(2), 7-24. https://doi.org/10.15678/EBER.2023.110201

Yagi, M., Yamanouchi, K., Fujita, N., Funao, H., & Ebata, S. (2023). Revolutionizing spinal care: Current applications and future directions of artificial intelligence and machine learning. *Journal of Clinical Medicine, 12*(13), 4188. https://doi.org/10.3390/jcm12134188

Żebrowski, P., Couce-Vieira, A., & Mancuso, A. (2022). A Bayesian framework for the analysis and optimal mitigation of cyber threats to cyber-physical systems. *Risk Analysis, 42*(10), 2275-2290. https://doi.org/10.1111/risa.13900

Appendix

Attitudes, benefits, and threats associated with the use of Artificial Intelligence tools in higher education

Please answer each of the following questions about what you know, how you feel, and what you do with AI tools. (Please note that there is no best answer; we just want to know your opinion about each item.)

	Attitudes (15 items)	**Strongly disagree 0**	**Disagree 1**	**Neutral 2**	**Agree 3**	**Strongly agree 4**
A1	AI tools content can be used if properly cited and documented					
A2	Authors should have proper knowledge on how to use AI tools					
A3	I recommend AI tools to a friend or colleague					
A4	I'm interested in using of a premium version of AI tools with advanced features					
A5	AI tools has a positive impact on my education/learning					
A6	There is a need for specific training on how to use AI tools in order for them to be useful.					
A7	I suggest providing adequate information on establishing ethical guidelines for the use of AI tools.					
A8	I think AI tools should be included in the study curricula					
A9	To improve AI applications in the real world, it is essential to encourage researchers to be honest and transparent about their methods.					
A10	I review and edit the response that generated by AI tools before using them in my work					
A11	AI tools can be listed as an author based on its significant contribution					
A12	I feel comfort with ethical and responsible use of AI-generated content from AI tools.					
A13	AI tools could enhance research (e.g., assisting the researchers in framing the sentences, improving the content drafted by the authors.					
A14	I think the responses generated by AI tools are overall easy and coherent					
A15	I trust the information that I read and see on AI tools?					
	Benefits (6 items)	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
B1	Easy to use					
B2	Save time					
B3	Accessible with low cost					
B4	Help students to ask questions and interact with the material at their own pace					
B5	AI tools are user-friendly					
B6	I know that AI tools are used in education and research					

	Threats (14 items)	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
T1	Lack of human interaction					
T2	Legal issue (e.g. copyright issues, authorship)					
T3	Decrease creativity and critical thinking					
T4	AI tools does not replace practical training					
T5	Security concerns					
T6	Technical issue					
T7	Over-reliance on technology					
T8	Ethical dilemma/concerns such as plagiarism					
T9	Need Internet all the time					
T10	Difficulty in handling complex task in research					
T11	Inaccurate/incorrect or biased information					
T12	Over-detailed, redundant, excessive content					
T13	Using AI tools will reduce skills and abilities of person who use it (e.g., writing skills, critical thinking ….etc)					
T14	I see AI tools as a threat to human ethics					

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Critically reflecting on the use of immersive virtual reality in educational settings: What is known and what has yet to be shown?

Alyssa P. Lawson[A]

Amedee Marchand Martella[B]

A University of California, Santa Barbara; Landmark College, Institute for Research and Training, USA

B University of California, Santa Barbara, USA

Keywords

Immersive virtual reality;
learning environments;
learning technology;
virtual reality.

Correspondence

alyssaplawson@gmail.com [A]

Article Info

Received 12 September 2023
Received in revised form 18 October 2023
Accepted 23 October 2023
Available online 24 October 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.35

Abstract

Interest in the educational applications of immersive virtual reality (IVR) has continued to grow worldwide, particularly in recent years. With the ever-increasing literature base on IVR in educational contexts, two patterns of data have emerged: one focused on the affective component and one focused on the cognitive component of IVR. Research focused on the affective component of this technology has consistently found that it is beneficial in increasing students' motivation to learn. However, there is less of a consensus in the literature on the cognitive benefits of IVR, with results sometimes indicating it (a) is an effective tool for learning, (b) is *not* an effective tool for learning, and (c) is similar to other instructional media in its impact on learning outcomes. As suggested by these inconsistent findings, there is a great deal left to be understood about when and how IVR can be effective for learning. Therefore, the goal of this reflection article is to draw attention to important research gaps that, if filled, may help to explain the inconsistent effects of this technology in the research literature. Additionally, this article highlights areas in need of further research, which we hope will aid in the advancement of knowledge surrounding the effective implementation of IVR in education.

Introduction

To meet the demands of an ever-changing world, it is imperative that students develop critical academic competencies such as reading literacy, critical thinking, digital literacy, and math fluency. At the K-12 levels, the integration of technology in the classroom has been touted as a way to "accelerate, amplify, and expand the impact of effective teaching practices" (Office of Educational Technology, 2017, p. 5). At the college level, improving course instruction has been focused on moving away from the traditional lecture method and towards methods that provide students with opportunities to be more involved in the learning process—termed *active learning* (Association of American Universities, 2017). Immersive virtual reality (IVR) is one tool that has been a topic of interest in the fields of education, educational psychology, and educational technology given that it moves instruction away from more passive forms of learning. IVR does this by fully immersing learners in a new environment, often through a head-mounted display that presents screens in front of each eye. Typical IVR devices include Oculus Quest, PlayStation VR, HTC Vive, and Google Cardboard, but there are many other devices currently available or in development. The immersion afforded by IVR devices allows students to feel as though they are in an environment different from the one in which they are physically present. By immersing students in interactive learning environments, they often increase their sense of presence and agency; these two factors can lead to increases in interest, intrinsic motivation, self-efficacy, embodiment, and self-regulation, which in turn are useful in improving their learning outcomes (see the *Cognitive Affective Model of Immersive Learning* by Makransky & Petersen, 2021).

Despite the increasing popularity of IVR and the appeal of using a technology that, in theory, can benefit students both cognitively and affectively, the results of empirical investigations comparing IVR to other media or to more traditional forms of instruction are mixed, leading to a number of questions as to when and how IVR can be effective for student learning. These mixed results in the literature make it difficult to provide specific recommendations to educators on when and how to effectively implement IVR in the classroom and to provide specific guidelines to VR designers on how to create effective learning environments within an IVR experience.

Therefore, the purpose of this reflection article is to discuss several research gaps that, if filled, may help to explain the inconsistent effects of IVR and to provide subsequent research directions aimed at advancing knowledge surrounding the effective implementation of IVR in education. Before discussing what more needs to be understood about IVR technology, we will first provide a brief overview of research on IVR.

Research on immersive virtual reality

Immersion is one of the prominent aspects of IVR that distinguishes it from desktop virtual reality (DVR; an interactive virtual world presented on a desktop screen) and other types of traditional instructional methods such as lectures/educational videos (Makransky, 2020; Makransky & Petersen, 2021). Immersion can be considered the extent to which the system creates a new virtual world for the learner. When immersion increases—and thus the virtual world is more vivid and realistic—a learner experiences higher presence, which is the subjective experience of being in the environment. Presence is important to develop because an increase in presence can increase certain affective characteristics, like interest and motivation (Makransky, 2020; Makransky & Petersen, 2021). By making students feel as though they are really in the new environment, they likely will enjoy the lesson more, which can increase their motivation to learn and their focus on the task. This may ultimately lead to a deeper understanding of the material.

The motivational and affective benefits of IVR lessons have been well researched and established within the literature. Indeed, empirical research studies and reviews of these studies often demonstrate that IVR is beneficial to the affective experience of learners, with positive effects shown across studies that vary in implementation style, the topic being taught, and the research design used. For example, IVR can positively impact motivation (e.g., Akgün & Atici, 2022; Matovu et al., 2023; Parong & Mayer, 2018 [Experiment 1]; Villena-Taranilla et al., 2022; Yu, 2021), interest (e.g., Akgün & Atici, 2022; Flavia Di Natale et al., 2020; Makransky et al., 2020; Parong & Mayer, 2018 [Experiment 1]; Yu, 2021), and self-efficacy (e.g., Akgün & Atici, 2022; Makransky et al., 2020), with benefits consistently shown across studies that vary in implementation style, the topic being taught, and the experimental design used.

Although the literature on motivation and IVR lessons has been consistent across studies, the research on the benefits of IVR lessons on *learning outcomes* has not demonstrated clear results (Matovu et al., 2023). On one side of the spectrum, many research studies have shown benefits of presenting learning material in an IVR lesson on certain learning outcomes (e.g., Alhalabi, 2016; Kim et al., 2019 [for long-term memory scores]; Kozhevnikov et al., 2013; Makransky et al., 2019a [for behavioral transfer tests]; Webster, 2016; Yang et al., 2022). For example, Makransky et al. (2019a) tasked students with learning lab safety by text, DVR, or IVR. This study demonstrated that students did better on behavioral transfer tests (i.e., lab safety tests conducted in the real world) when they learned in IVR compared to when they learned via text. There have also been a handful of meta-analyses published recently that have demonstrated a positive, albeit small, impact of learning from IVR. For example, in Coban's (2022) review of 49 primary studies on IVR, the overall effect size of using IVR in learning was positive but small (g = .38; Hedge's g is a common metric used to measure the magnitude of difference between two groups), with the effect being strongest in fields like architecture, engineering, geometry, and chemistry. Similarly, Wu et al. (2020) analyzed 35 studies and found that the overall effectiveness of IVR was also positive but small (g = .24).

On the other side of the spectrum, some researchers have found that IVR leads to significantly lower performance on certain learning outcomes as compared to other instructional media/approaches (e.g., Makransky et al., 2019b [for knowledge test]; Makransky et al., 2021 [Experiment 2, for declarative knowledge test]; Moreno & Mayer, 2004 [for retention test]; Parong & Mayer, 2018 [Experiment 1, for factual questions], 2021a [for transfer test], 2021b [for transfer test]). When compared to more traditional or 2D formats, IVR has proved less useful for students' learning in several studies. For example, in multiple studies in which students learned about cells and the blood stream, students who learned using IVR performed significantly lower on factual questions (Parong & Mayer, 2018 [Experiment 1]) and transfer test (Parong & Mayer, 2021a) than those who learned using a slideshow. When compared to non-immersive technologies, IVR also has proved less useful for students' learning; for example, in a game lesson about designing plants for different environments, participants who learned with IVR did not perform as well as those who learned from DVR on both the retention test (Moreno & Mayer, 2004).

In the middle of the spectrum, a large portion of research studies have found no significant difference in the effectiveness of IVR on certain learning outcomes when comparing IVR to other types of instructional media (Ekstrand et al., 2018; Hassenfeldt et al., 2020; Liu et al. 2021; Makransky et al., 2019a [for retention test]; Makransky et al., 2019b [for transfer test]; Makransky et al., 2021 [Experiment 1; Experiment 2, for procedural knowledge and transfer tests]; Parong & Mayer, 2021a [for retention test], 2021b [for retention test]. When looking to systematic reviews for a more comprehensive view of the literature, Luo et al., (2021) meta-analyzed 22 articles from 2000 to 2019 with HMD as a moderator and found that the overall effect of using IVR was not meaningfully different than other instructional media (g = .20, with the 95% confidence interval ranging from an effect size of -.16 to .55).

Given these inconsistent results found within the IVR literature, there is a great deal left to be understood about IVR technology surrounding when and how it can be effectively implemented to promote student learning. Therefore, it is important to discuss research gaps that, if filled, could help to explain the varied results, may help to establish boundary conditions for when IVR is effective versus when more traditional forms of instruction are most useful to students, and could provide insight into how to design effective IVR lessons. Gaps in the literature and subsequent research directions aimed at filling these gaps will be discussed in more detail in the following section.

What more needs to be understood about IVR technology?

To better understand when and how IVR technology is effective for student learning, we (a) consider several possible research gaps that, if filled, may help to explain the inconsistent learning results within the IVR literature and (b) present subsequent research directions that we believe are important for advancing knowledge on how to effectively implement IVR in education. We will focus on three research gaps that we deem to be most pressing, although there are other gaps that exist within the literature. The first research gap relates to whether IVR imposes larger demands on working memory than other forms of instruction. The second research gap relates to how IVR lessons are being designed in experimental studies. And the third research gap relates to what IVR is being compared to in experimental studies. Each research gap is discussed below.

Research gap 1: The cognitive load of the IVR lesson

The impact of the cognitive load of an IVR lesson reflects a research gap within the IVR literature. There are two parts of this research gap that will be discussed: the demands that are imposed on working memory during a lesson and the type of cognitive load that is being increased during a lesson.

Research gap 1A: Demands on working memory

The cognitive load of a lesson can vary from one instructional approach to another. Cognitive load refers to the demands imposed on working memory during learning. Working memory is severely limited in the amount of information that can be processed at one time, which is essential to understand in terms of its role in learning and instruction (Fenesi et al., 2015). With working memory only able to process so much novel information at one time, it is vital for students to be able to deal with all of the novel incoming information. Unfortunately, how IVR lessons impact learners' cognitive load, as compared to other instructional methods, remains an under-researched area.

For those studies that have examined how IVR lessons impact learners' cognitive load (e.g., Baceviciute et al., 2021; Huang et al., 2021; Makransky et al., 2019b, Mayer et al., 2022; Parong & Mayer, 2018, 2021a, 2021b; Petersen et al., 2022), results tend to show that there is an increase in cognitive load when learning from IVR. For example, Makranksy et al. (2019b) had students learn how to conduct a lab procedure in a chemistry course using either IVR or a 2D computer simulation. These learning mediums were then switched in the second part of the lesson. During these lessons, students' cognitive load was measured using EEG. During the first part of the lesson, there were no differences in the amount of workload participants experienced but in the second part of the lesson, those in the IVR condition had higher load than those on the computer. These results demonstrate that IVR can add to students' cognitive load, particularly if they are asked to learn additional material in IVR after already viewing another lesson. Similarly, in a lesson on cells and the bloodstream where students learned via IVR or a slideshow, participants in the IVR condition reported that they were more distracted and/or had a harder time focusing during the lesson as compared to those who viewed the slideshow (Parong & Mayer, 2018).

Not all IVR lessons are designed in the same way, which may lead to variations in the amount of cognitive load that is imposed on learners. Subsequently, certain IVR lessons could impose larger demands on working memory than the instructional methods to which they are being compared. Similarly, because not all instructional methods to which IVR lessons are compared are designed or implemented in the same way, they may also vary in the amount of cognitive load imposed on learners. These possibilities could lead to inconsistent findings in the IVR literature depending on how each mode of instruction has been designed. Being more mindful of the design of the lessons, both within IVR and outside of IVR, and examining the demands that are being placed on working memory during learning would help to identify any potential barriers for learners' processing of relevant information and lend insight into why there might have been differential effects of the compared approaches.

Future research should further investigate the impact of the cognitive load caused by IVR. Cognitive Load Theory (CLT; Paas & Sweller, 2022; Sweller, 1994, 2020) and Cognitive Theory of Multimedia Learning (CTML; Mayer, 2022) are useful theories to consider when specifically designing instruction in IVR. With more traditional forms of instruction (e.g., watching an instructor present information or watching an instructional video on a computer screen), all of the relevant information students need to focus on is presented directly in front of them in one place. With this directly presented information, students can focus their attention and more easily recognize information that is distracting and irrelevant to the lesson. However, in IVR lessons, information that students may be asked to focus on comes from all directions, potentially putting students in a situation where they could easily miss key ideas. Therefore, future research should specifically focus on how the design of the different modes of instruction being compared in a study affect students' cognitive load and learning outcomes. Further, researchers should examine how individual differences in learners' cognitive skills could impact how well they learn with different types of media (Lawson & Mayer, under review) and how these differences could impact the effectiveness of IVR. Given that learning new material through IVR can tax students' limited working memory via the presentation of extraneous material, it is important to examine how this technology affects students who may be able to handle this additional cognitive load (through better inhibition ability or being able to ignore distractions) versus students who are less able to handle this cognitive load (see Albert et al., 2020 and Grenell & Carlson, 2021 for a discussion of individual differences in executive function and academic achievement/learning).

Research gap 1B: Type of cognitive load being increased

Another glaring issue related to cognitive load is the minimal investigation by researchers into what *type* of cognitive load is being increased during a lesson. Researchers have investigated and differentiated three different types of cognitive load that stem from instructional material (Mayer, 2022; Paas & Sweller, 2022; Sweller, 1994, 2020). Extraneous load or extraneous processing occurs when information is presented in a lesson that is not relevant to the lesson itself, such as irrelevant facts or pictures that draw learners' attention away from the important information being conveyed. Intrinsic load or essential processing occurs when learners build a mental representation of the presented material; this load increases as the complexity or the interactivity of the material increases. Lastly, some researchers propose a third kind of processing called germane load or generative processing that occurs when a learner works to make sense of the material presented, develops connections between different parts of the material, and connects the novel information with their prior knowledge.

These different types of cognitive load vary in how they impact learning—having an increase in extraneous processing would likely hurt students' learning because learners are paying attention to information that is not relevant to the main goal of the lesson whereas an increase in generative processing would likely benefit students' learning because it encourages deeper processing of the learning material. For example, in one study conducted by Parong and Mayer (2021a), participants' different levels of cognitive load were measured using self-report Likert scale questions and their reported cognitive state was measured through EEG. Participants reported having higher cognitive load, *specifically extraneous processing*, when they learned in the IVR condition compared to when they learned in the slideshow condition. Furthermore, a mediation analysis of the relationship between learning condition, *extraneous processing*, and retention scores found that there was a significant mediation path wherein those in the IVR condition reported higher extraneous processing which in turn led to worse retention scores as compared to those in the slideshow condition.

Without differentiating the type of processing that a learner is experiencing, as is often an issue in the literature, it is difficult to determine whether increases or decreases in cognitive load will be helpful or detrimental to student learning. Therefore, the question of what type of cognitive load drives the increase in reported cognitive load from IVR technology needs to be further investigated to better to determine why certain patterns of data are emerging from IVR studies.

Future research should be focused on how different components of learning in IVR contribute to different types of cognitive load and what types of strategies can reduce less desirable types of cognitive load (like extraneous processing) and increase more desirable types of load (like generative processing). As part of this research direction, we need to understand what components of an IVR lesson (e.g., interacting with objects in the virtual environment, experiencing a large amount of visual information, trying to understand how to use the device) impact different types of cognitive load. With this understanding in place, we can then better understand how to help remedy the issues that cause increases in extraneous processing and focus our attention on how to encourage generative processing during a lesson.

Research gap 2: The impact of how IVR conditions are designed in IVR studies

The impact of how IVR conditions are designed in IVR studies reflects a research gap within the IVR literature. There are two parts of this research gap that will be discussed: the prioritization of learning theory in the design of an IVR lesson and whether IVR is used to exclusively teach content or is used as a supplemental instructional tool.

Research gap 2A: Learning theory prioritization

Just as much of the literature on learning with IVR lessons does not often consider the demands imposed on working memory during learning, nor does it always prioritize learning theories in the design and implementation of this technology in an educational setting (Lui et al., 2023; Matovu et al., 2023; Radianti et al., 2020). Indeed, a gap in the literature exists for whether the alignment of IVR lessons to effective design guidelines differentially impacts the technology's effectiveness. If certain IVR lessons follow better design guidelines than other IVR lessons, it is highly possible that the outcomes of using these various lessons for learning could impact results, particularly if these lessons lead to differences in how students cognitively interact with the material presented in the lesson. These design considerations could impact the various types of cognitive load present in the IVR lesson, as previously discussed, with learners potentially struggling to keep up with the information presented, thereby hindering learning. These design considerations could also impact the degree of cognitive engagement being cultivated during a lesson.

There have been some studies that have involved the incorporation of learning theory into the design of IVR lessons. More specifically, several researchers have integrated generative learning strategies into IVR to reduce extraneous processing and manage essential processing and/or increase generative processing (e.g., Klingenberg et al., 2020; Makransky et al., 2021; Parong & Mayer, 2018 [Experiment 2]). However, this research area is small and is in need of further investigation. Therefore, there is a need for more focused research on the incorporation of learning theory into the design of IVR lessons to continue as this can provide more insight into how to induce learning more effectively through IVR lessons. It may be the case where IVR lessons that are designed based on learning theories and include effective learning strategies are more effective for student learning than those that do not involve these design considerations, which may be contributing to the variability in findings across the IVR literature.

In future research, investigators should apply learning theories to IVR lessons in the pursuit of recognizing what aspects of the IVR lesson promote better learning outcomes. They should also investigate whether certain learning strategies are more effective during an IVR lesson than other strategies. For example, perhaps adding self-explanations to an IVR lesson helps students learn more than adding retrieval practice activities to the lesson. Or perhaps these strategies work best in tandem—that is, both are needed to maximize student learning. Further, when incorporating effective learning strategies into IVR lessons, researchers should assess whether strategies that have been demonstrated to be effective for learning, such as practice tests, peer teaching, self-explanations, feedback, supplemental instruction, etc., confer the same benefits when embedded in traditional forms of instruction versus when embedded in IVR lessons. In other words, is it the VR technology and immersion themselves that are key in promoting learning, or is it simply the fact that students benefit from embedded strategies that have been demonstrated to be effective for learning, regardless of the particular mode of instruction used during a lesson? It could be that traditional forms of instruction are just as effective as IVR when combined with effective learning strategies. In other words, strong IVR conditions should be compared to strong comparison conditions to determine if IVR provides learning benefits above and beyond the learning strategies themselves (a point elaborated on under Research Gap 3).

Research gap 2B: Exclusive IVR or integrated IVR?

When examining the IVR conditions in IVR studies, one will find that some studies use IVR to directly teach students content (e.g., Lui et al., 2020; Madden et al., 2020; Parong & Mayer, 2018, 2021a, 2021b; Su et al., 2022) whereas other studies use it as an extension activity that is integrated into more traditional modes of instruction (e.g., Campos et al., 2022; King et al., 2022; Liu et al., 2022; Stranger-Johannessen, 2018). These two uses of IVR highlights an important question: Is IVR more effective when used to directly teach content or when used as an active learning activity that is embedded in more traditional forms of instruction? IVR may be most useful as an extension activity to help students further encode the content and promote generalization programming through immersive activities. For example, Lui et al. (2020) taught students about the *lac* operon in two 80-minute lectures. One group of participants received IVR lessons over the following two weeks to help reinforce the ideas from the lecture while the other group did not. Students who participated in the IVR session learned more about the *Lac* Operon Concept Inventory than those who did not participate in this session.

As another example, Makransky & Mayer (2022) studied the impact of a six-lesson intervention that involved teaching students about climate change. The first lesson for both the IVR condition and the video condition involved a fake news article followed by a discussion of controversy surrounding climate change. The second lesson involved instruction on the scientific method and a virtual field trip to Greenland. During this lesson, students in the IVR condition experienced a 360-degree virtual trip to Greenland whereas students in the comparison condition experienced the virtual field trip to Greenland as a class video that was displayed on a projector screen. After this lesson, they took an immediate posttest. Subsequent lessons for both conditions involved learning about experiments and interpreting results as regular class sessions, and a delayed posttest was given after the last session. Results indicated that students learned more when they were able to take the virtual field trip in conjunction with the other course material presented in regular class sessions as compared to those who watched the video in conjunction with the other course material presented in regular class sessions, both on the immediate and delayed posttest. Therefore, using IVR within a more traditional class context seemed to improve student learning.

Within the active learning literature, similar benefits have been demonstrated for the integration of active learning activities into traditional STEM classes (see Freeman et al., 2014). Perhaps when used as a motivational tool and/or an additional encoding tool paired with class lectures or other instructional videos/materials, IVR more consistently improves learning than when it is used to exclusively teach class content. This idea of using IVR as a supplemental tool in education needs to be directly tested, particularly against exclusive VR conditions, before making conclusions about its benefits as a learning tool.

In future research, it is important to assess whether IVR, which is used to directly teach content, is more or less effective than when it is used as an active learning activity embedded in more traditional forms of instruction. Researchers should examine the effects of students learning content directly through an IVR lesson or through an integration of IVR technology and more traditional modes of instruction such as class lectures. Given the potential issues with distracting information in IVR lessons and the subsequent increases in extraneous processing, providing students with foundational knowledge outside of IVR may help to reduce the demands imposed on working memory. As an example, when researching the integration of IVR and other modes of instruction, researchers should also investigate how much class time should be dedicated to IVR experiences and whether this tool should occur after a lecture or should be interspersed throughout the lecture (see Martella & Schneider, in press, for information on lecture and active learning integration). These types of considerations would provide direction to instructors looking to implement the technology in their classroom.

Research gap 3: The impact of how comparison conditions are designed in IVR studies

The impact of how comparison conditions are designed in IVR studies reflects a research gap within the IVR literature. Two parts of this research gap will be discussed: (a) the type and quality of the comparison conditions and (b) the extent to which causal conclusions can be made regarding the efficacy of specific component(s) of the treatment package.

Research gap 3A: Type and quality of non-VR comparison conditions

The impact of how comparison conditions are designed in IVR studies reflects a research gap within the IVR literature. Indeed, the impact of the type and quality of these conditions has not been well studied. One potential issue when designing intervention studies, particularly in the context of instructional comparisons, is the inclusion of "strawman" conditions—conditions that are "easy to knock down" or, in other words, are doomed to fail from the start. For example, inquiry-based conditions that are unguided can be described as a "strawman" in that they are unlikely to be effective and do not serve as a fair comparison to alternative methods, such as direct instruction, given what we know about the importance of guided instruction (see Davis et al., 2017; Hmelo-Silver et al., 2007). Traditional lecture conditions can also serve as a "strawman" if the lecture is fully passive, poorly presented, and the slides are convoluted and/or involve extraneous information. To make the lecture a better comparison condition, strategies such as using mental imagery or increasing the structure of a lecture through outlines, for example, can be incorporated during the lecture design phase (deWinstanley & Bjork, 2002) as can taking into account multimedia design principles such as not presenting the same information in multiple formats simultaneously (Mayer, 2022).

Unfortunately, the extent to which comparison conditions reflect fair comparisons in IVR studies has not been well examined by researchers nor has the impact of different forms of "traditional" instruction. When reading the literature (both published and unpublished studies), the variation in what IVR lessons are being compared to is vast, with conditions involving 2D static images (Porter et al., 2019), class lectures/PowerPoint lessons/recorded videos (Bukoski, 2019; Drake, 2022; Lamb et al., 2018; Parong & Mayer, 2018; Sanzana et al., 2022), textbook/booklet/manual readings (Alrehaili & Al Osman, 2022; Makransky et al., 2019a; Tarng et al., 2022), and hands-on activities (Greenwald et al., 2018; Madden et al., 2020), among a myriad of other modes of instruction. This variation is, unfortunately, collapsed across conditions in meta-analyses, perhaps clouding the impact of different types of comparison conditions on student learning as compared to IVR conditions.

A closer reading of these conditions illuminates the presence of many "strawmen" that are unlikely to be effective as compared to the IVR conditions. For example, in one study, participants in the control group received up to 45 minutes to study a 14-page lab manual on lab safety that was designed according to instructional design principles; participants in the IVR condition were engaged in many activities involving narrative guidance, feedback, practice multiple-choice questions, and lab safety tasks to perform (Makransky et al., 2019a). The retention assessment for both conditions involved multiple-choice questions testing for conceptual and procedural knowledge. The transfer assessment included behavioral tasks—testing experiences that drew on experiences practiced in the IVR condition but not in the control condition. The IVR condition and text condition performed similarly on the retention test but unsurprisingly, the IVR condition resulted in significantly higher scores on the behavioral transfer tasks than the control condition. But one must ask whether the comparison condition served as a fair control. It could be argued that the increase in learning in IVR was not due to the immersive nature of the lesson but instead due to (a) a testing/practice effect, (b) the embedded strategies that the control condition did not receive, (c) more structured learning, and/or (d) the fact that students in the control condition were not explicitly taught the content.

As another illustrative example, learners in a study conducted by Tarng et al. (2022) were asked to learn about physics concepts related to time and space travel. The IVR condition learned this content by playing a pilot and control room staff to complete space exploration missions. During this lesson, they were asked questions that they needed to answer correctly to move forward in the lesson and were also given real-time feedback. The comparison group learned the content by reading a physics textbook. Although the study concludes that the IVR lesson was more effective than the textbook lesson, it is important to discuss whether the driving force for these learning differences was the active learning experiences and practice questions learners received during the lesson or whether it was the immersive nature of the lesson.

Unfortunately, the nature of the conditions compared in IVR studies has not been subjected to critical analysis. Ideally, to determine whether schools should spend money on VR equipment and whether instructors should take the time to adopt and implement the technology in their classrooms, the comparison conditions that serve as control conditions should reflect true "business-as-usual" or "regular instruction" conditions in that what students are asked to do in these studies reflects what they would/could really do during classroom learning. Though some may argue that students are typically asked to read textbooks for class, for example, it is important to ask whether the only exposure students receive to the content in class is through textbook readings. More often, the readings are assigned to provide exposure before coming to class or to solidify learning after class and are not used in isolation, as has been found in IVR studies.

IVR lessons should also be compared to well-designed alternative approaches such as interactive lectures or validated curricular materials that involve effective learning strategies. Designing passive control conditions such as traditional lectures is no longer productive (Freeman et al., 2014). As touted in the active learning literature, incorporating more opportunities for students to participate in the learning process can be beneficial to their learning (see Freeman et al., 2014). Although the active learning literature suffers from many of the same issues as the IVR literature, there is a great deal of research on the benefits of engaging students in learning activities such as retrieval practice, elaborative interrogation, and self-explanations (Dunlosky et al., 2013). Perhaps students do not need to be immersed in a virtual world to experience boosts in motivation and learning—it may simply be the case that we need to design more interactive learning experiences within the real-world class context to aid students in their learning.

By continuously comparing IVR conditions to bad control conditions, we overlook the potential benefits of this technology for real-world instruction and miss out on the opportunity to offer specific suggestions to instructors on how to improve classroom instruction. As an example of a fair comparison condition, Petersen et al. (2022 [Experiment 2]) compared an IVR lesson with an active pedagogical agent who taught a lesson about pipetting in a virtual laboratory setting to a real-life lesson with an instructor who taught a lesson about pipetting in a chemistry laboratory. The setup between conditions was designed to be identical and students in both conditions received active practice with pipetting in addition to explicit instruction. Results indicated that students in the IVR condition made more errors in dexterity with the pipette but performed similarly for serial dilution and safety performance on the real-life transfer test, had lower declarative knowledge scores, experienced significantly higher extraneous cognitive load, and had smaller increases in self-efficacy than the real-life condition. Although the discussion is largely framed to justify the use of VR as a complement to traditional teaching, it is important to highlight that traditional teaching served as an overall better intervention than IVR when designed to be comparable in terms of effective learning experiences (e.g., explicit instruction and active practice).

One critical direction for researchers to take in future research studies is to determine the extent to which comparison conditions reflect fair comparisons in IVR studies, perhaps through a systematic review of studies that have already been conducted. Further, researchers should design more studies that involve true "business-as-usual" conditions or that include rigorous instructional practices (e.g., interactive lectures) and materials as comparison conditions, such as the example above by Petersen et al. (2022 [Experiment 2]). Finally, researchers should investigate the impact of different forms of instruction to determine if specific comparison conditions promote greater student learning than other types of comparison conditions. It may be the case where well-designed lectures serve as a better learning medium than well-designed videos, for example.

Research gap 3B: Causal conclusions about why one condition was more effective

Another important consideration when designing comparison conditions in IVR studies is whether these conditions allow researchers to make causal conclusions about *why*, specifically, one condition outperformed (or did not outperform) the other. To be able to make valid inferences, an experiment should be unconfounded, with only a single contrast occurring between conditions (Klahr, 2013). IVR conditions are often designed as a treatment package with many different instructional components and are frequently compared to a control condition that gets a different, albeit minimal, treatment package (e.g., Alrehaili & Al Osman, 2022; Chittaro & Buttussi, 2015; Makransky et al., 2019a; Tarng et al., 2022). We adopt the use of "treatment package" and "components" from Ward-Horner and Sturmey (2010) in that we use "*component* to refer to variables that comprise a *treatment package* and treatment package to refer to the application of an intervention with all of its components" (p. 686). At first glance, an IVR treatment package versus a control treatment package reflects one contrast—the type of treatment package students receive. However, upon closer examination, the instructional components of these treatment packages are generally different between conditions, leading to more than one contrast between conditions.

As an illustrative example, Alrehaili & Al Osman (2022) assigned students to an IVR condition, a DVR condition, or a booklet condition to assess the impact of immersion. The IVR condition was designed according to multimedia principles. It involved a tutorial video to introduce students to relevant concepts and four different game levels with contextual guidelines in the form of textual messages and with specific tasks to complete. Each level of the game built upon the previous level. The comparison conditions received either the IVR lesson given via a computer monitor rather than a head-mounted display or a small booklet on honeybees that was written to mimic a 7th or 8th grade textbook. This booklet included many pictures from the IVR honeybee game. These conditions reflect three different treatment packages, each with different instructional components such as tasks, videos, and textual messages in IVR and written facts and pictures in the booklet, making it difficult to interpret outcomes on the knowledge test.

When there are multiple differences between or among conditions, only general conclusions can be drawn; for example, "IVR was more effective than the traditional condition." This conclusion might be satisfactory to some researchers who want to know if IVR is more effective than business-as-usual instruction, for example. However, *why* the IVR condition was more effective than the traditional condition cannot be answered from studies that involve multiple differences. It may be the case where immersing students in a virtual environment is the causal factor, or it may be the case where the other instructional components that were embedded in the IVR lesson but not in the comparison condition were responsible for boosting student performance.

Therefore, before we can conclude that immersing students in a virtual world is necessary for improving student learning and motivation beyond what we can give them in a traditional classroom, we need to be sure it is, in fact, the immersion and/or interactivity afforded by the VR environment and not the added instructional components embedded in an IVR lesson that result in greater learning gains than comparison conditions. Without knowing why an IVR condition resulted in greater learning, it is difficult to offer specific guidance and practical advice on effective IVR implementation to instructors (see similar discussion in Martella & Schneider, in press) and to definitively say that IVR should be adopted by instructors as compared to more affordable, real-world instructional interventions. The comparison conditions should thus be designed intentionally to minimize differences between the conditions being compared in the study.

In continuing to do research on IVR, it is vital to isolate and compare key instructional features. There have been a number of research studies that have isolated the immersion and interactivity component of IVR conditions by comparing an IVR lesson to a less immersive DVR lesson (e.g., Alrehaili & Al Osman, 2022; Barnidge et al., 2022; Makransky et al., 2019a) or a more passive 2D video lesson (e.g., Allcoat & von Mühlenen, 2018; Parong & Mayer, 2021a). Although these studies do control variables and isolate the impacts of the IVR technology, they typically afford little insight into how classroom instruction compares to IVR lessons when variables are controlled and contrasts are kept to a minimum (see Petersen et al., 2022 [Experiment 2] for an example of how classroom instruction can be compared to IVR while minimizing differences). Therefore, future research should expand on these prior studies by using component analysis to compare IVR and non-VR conditions, particularly those that incorporate effective learning strategies so as not to fall into the "strawman" trap. One way to identify active elements is to conduct a factorial design where two variables are examined via four conditions, for example. As an illustrative example, Parong and Mayer (2021a) examined whether an IVR lesson was as effective as an equivalent PowerPoint lesson on a desktop computer and whether the generative learning strategy of practice testing boosted performance in either medium. Their study involved a 2 X 2 factorial design with four conditions: IVR, PowerPoint, IVR + practice testing, and PowerPoint + practice testing. Results indicated students who received the IVR lesson performed significantly lower on the transfer test and performed lower, albeit not statistically significantly lower, on the retention test than those who received the PowerPoint lesson, with or without practice questions added to the lessons. This study, therefore, lends insight into the impact of the instructional medium as well as the impact of an embedded generative learning strategy.

Discussion

Recommendations for the future of IVR

In this reflection, we have highlighted three main areas in need of further investigation to advance knowledge surrounding the effective implementation of IVR in education. These include (a) a more thorough investigation of the impact of IVR lessons on different types of cognitive load, (b) a deeper look into how IVR conditions are being designed and whether design differences impact results, and (c) a deeper look into how comparison conditions are being designed and whether design differences impact results. Based on these research gaps, we have put together a set of recommendations for researchers, educators, and VR developers that reflect what we currently know about the field as well as where we believe future research should go next.

Learning and technology researchers

Researchers have the fundamental job of bridging the gap between the development of VR technology and the effective implementation of this technology in the classroom. Although there is a growing literature base on IVR in educational contexts, this reflection demonstrates that there are many gaps remaining that need to be filled in order to determine the usability of this technology for education. As discussed throughout our reflection, there are three major gaps in the literature that need to be addressed in order to determine whether and how best to use this technology in education.

Regarding cognitive load, there are two primary areas in which researchers in technology and learning should further investigate. First, we recommend that researchers investigate how different types of IVR lessons impact learners' cognitive load while learning. For example, researchers could investigate how changes in the interactivity of an IVR lesson can impact the cognitive load learners experience during the lesson. Second, we recommend that researchers keep in mind the variation in types of cognitive load and subsequent impacts on learning when conducting research on the impact of cognitive load in learning with IVR technology. For example, researchers should investigate how learning in a specific IVR lesson can increase or reduce learners' extraneous, essential, and generative processing.

Regarding the integration of IVR into classroom environments, there are two primary areas that should be further investigated. First, we recommend that researchers incorporate learning theories into the designs of lessons being used in research that investigates the use of IVR technology in learning. For example, researchers could design IVR lessons according to Bloom's Taxonomy (Bloom, 1956; Anderson & Krathwohl, 2001) to target different types of knowledge during the lesson. Second, we recommend that researchers investigate the benefits of using IVR as a way for exclusively learning content (i.e., learning content with IVR lessons only) compared to as a way to support learning (i.e., as an active learning tool embedded in a more traditional classroom structure). For example, researchers could directly compare how well students learn content when they are taught the content directly with IVR (including explicit instruction and practice) versus when they are taught the content outside of IVR (e.g., in a more traditional type of learning environment) and use IVR to help cement the new skills through immersive practice activities.

Regarding the comparison conditions used in IVR research, there are two primary areas that are in need of deeper investigation. First, we recommend that researchers investigate the impact of using strawmen conditions versus more well-developed control conditions in IVR research. For example, researchers could investigate the impacts of using IVR for learning when compared to a condition in which students simply read a textbook or when the activities in the control condition match those that are done in the IVR lesson. Second, we recommend that researchers think critically about the research question under investigation when designing the IVR and comparison conditions. For example, if a researcher is interested in understanding whether it is the interactivity of an IVR lesson that impacts learning, they should ensure that only *interactivity* is different between the two conditions. However, if they are interested in understanding whether there is a unique benefit of hands-on learning in IVR, the researcher should present the same hands-on activity in both conditions in order to control variables and draw sound conclusions.

K-16 educators

Based on the current research base for IVR in educational settings, our recommendation to instructors interested in using readily available IVR experiences in their own classrooms is to adopt the technology as a motivational tool rather than as a primary learning tool, at least at this point in time. Educators and students alike should benefit from the aspects of IVR we know work well—that is, IVR is an effective motivational tool for students and can help increase their interest in the learning material. However, more focused research on when and how IVR can benefit learning needs to be conducted before it is adopted as the *primary method* of teaching foundational content. One way to implement IVR to leverage its benefits for motivational and affective components of learning is to have students take a "virtual field trip" (i.e., have students experience a location and/or experience they would otherwise not be able to access by using IVR devices) to spark their interest and then leverage more traditional methods of instruction to teach specific content.

VR developers

As for the developers of VR educational environments, it is important to be aware of and incorporate findings from research on the cognitive processes of learning. As discussed, many IVR lessons are not designed according to theories of learning. As such, the way in which the material is presented to learners is oftentimes inconsistent with how the brain processes information. By integrating findings from research on effective design principles, developers can create better lessons that align with human cognition. We recommend that developers work more closely with educational researchers (and vice versa) to create educational content that can be more effectively implemented in educational spaces.

Conclusion

Immersive virtual reality is quite effective in increasing presence and motivation. These outcomes are a large contribution to learning as an important step in getting students to engage in deeper understanding by motivating them to want to learn (Mayer, 2022). However, when and how IVR is effective for student learning has not been well established, with a mixture of studies showing IVR lessons to be better than, equal to, or worse than other modes of instruction. With these inconsistent findings and design/implementation variation that exist in the literature, providing specific implementation guidance to instructors remains difficult. By outlining research gaps that, if filled, may help to explain inconsistent results in the literature surrounding IVR's effectiveness for learning and by providing future research recommendations for researchers, it is our hope that technological tools will be more effectively and appropriately researched and integrated into K-16 classrooms.

References

Akgün, M., & Atici, B. (2022). The effects of immersive virtual reality environments on students' academic achievement: A meta-analytical and meta-thematic study. *Participatory Educational Research, 9*(3), 111–131. https://doi.org/10.17275/per.22.57.9.3

Albert, W. D., Hanson, J. L., Skinner, A. T., Dodge, K. A., Steinberg, L., Deater-Deckard, K., Bornstein, M. H., & Lansford, J. E. (2020). Individual differences in executive function partially explain the socioeconomic gradient in middle-school academic achievement. *Developmental Science, 23*(5), e12937. https://doi.org/10.1111/desc.12937

Alhalabi, W. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behavior & Information Technology, 35*(11), 919–925. https://doi.org/10.1080/0144929x.2016.1212931

Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology, 26*, 2140. https://doi.org/10.25304/rlt.v26.2140

Alrehaili, E. A., & Al Osman, H. (2022). A virtual reality role-playing serious game for experiential learning. *Interactive Learning Environments, 30*(5), 922–935. https://doi.org/10.1080/10494820.2019.1703008

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational objectives*. Longman.

Association of American Universities (2017). *Progress toward achieving systemic change: A five-year status report on the AAU undergraduate STEM education initiative*. https://www.aau.edu/sites/default/files/AAU-Files/STEM-Education-Initiative/STEM-Status-Report.pdf

Baceviciute, S., Terkildsen, T., & Makransky, G. (2021). Remediating learning from non-immersive to immersive media: Using EEG to investigate the effects of environmental embeddedness on reading in virtual reality. *Computers & Education, 164,* 104122. https://doi.org/10.1016/j.compedu.2020.104122

Barnidge, M., Sherrill, L. A., Kim, B., Cooks, E., Deavours, D., Viehouser, M., Broussard, R., & Zhang, J. (2022). The effects of virtual reality news on learning about climate change. *Mass Communication and Society, 25*(1), 1–24. https://doi.org/10.1080/15205436.2021.1925300

Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. David McKay Co. Inc.

Bukoski, R. J. (2019). *The impact of traditional and video modalities on concept retention*. [Doctoral dissertation, Indiana University of Pennsylvania]. ProQuest Dissertations Publishing.

Campos, E., Hidrogo, I., & Zavala, G. (2022). Impact of virtual reality use on the teaching and learning of vectors. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.965640

Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics, 21*(4), 529–538. https://doi.org/10.1109/TVCG.2015.2391853

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Psychology Review, 36*, 100452. https://doi.org/10.1016/j.edurev.2022.100452

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Pearson.

Davis, A., Goulding, M., & Suggate, J. (2017). *Mathematical knowledge for primary teachers* (5th ed.). Routledge. https://books.google.com/s?hl=en&lr=&id=7T8lDwAAQBAJ&oi=fnd&pg=PP1&dq=Mathematical+knowledge+for+primary+teachers&ots=tJtb9cLwj_&sig=uJC1mIqK756wnnVoFvBTg35kvQA#v=onepage&q=Mathematical%20knowledge%20for%20primary%20teachers&f=false

deWinstanley, P. A., & Bjork, R. A. (2002). Successful lecturing: Presenting information in ways that engage effective processing. *New Directions for Teaching and Learning, 89*, 19–31. https://doi.org/10.1002/tl.44

Drake, E. W. (2022). *The impact of virtual reality and the virtual reality learning environment of student assessment outcomes in a unit of study in middle school life science: A quasi-experimental study.* [Doctoral dissertation, The University of Arizona Global Campus]. ProQuest Dissertations Publishing.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. http://doi.org/10.1177/1529100612453266

Ekstrand, C., Jamal, A., Nguyen, R., Kudryk, A., Mann, J., & Mendez, I. (2018). Immersive and interactive virtual reality to improve learning and retention of neuroanatomy in medical students: A randomized controlled study. *CMAJ Open, 6*(1), E103-E109. https://doi.org/10.9778/cmajo.20170110

Fenesi, B., Sana, F., Kim, J. A., & Shore, D. I. (2015). Reconceptualizing working memory in educational research. *Educational Psychology Review, 27*, 333–351. https://doi.org/10.1007/s10648-014-9286-y

Flavia Di Natale, A., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology, 51*(6), 2006–2033. https://doi.org/10.1111/bjet.13030

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Greenwald, S. W., Corning, W., Funk, M., & Maes, P. (2018). Comparing learning in virtual reality with learning on a 2D screen using electrostatics activities. *Journal of Universal Computer Science, 24*(2), 220–245. https://doi.org/10.3217/jucs-024-02-0220

Grenell, A., & Carlson, S. M. (2021). Individual differences in executive function and learning: The role of knowledge type and conflict with prior knowledge. *Journal of Experimental Child Psychology, 206*, 105079. https://doi.org/10.1016/j.jecp.2020.105079

Hassenfeldt, C., Jacques, J., & Baggili, I. (2020). Exploring the learning efficacy of digital forensics concepts and bagging and tagging of digital devices in immersive virtual reality. *Proceedings of the Twentieth Annual DFRWS USA, 33,* 301011. https://doi.org/10.1016/j.fsidi.2020.301011

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry-learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Huang, C. L., Luo, Y. F., Yang, S. C., Lu, C. M., & Chen, A-S. (2020). Influence of students' learning style, sense of presence, and cognitive load on learning outcomes in an immersive virtual reality learning environment. *Journal of Educational Computing Research, 58*(3), 596–615. https://doi.org/10.1177/0735633119867422

Kim, S., Heo, R., Chung Y., Kim, J. M., Kwon, M. P., Seo, S. C., Park, G., & Kim, M. (2019). Virtual reality visualization model (VRVM) of the Tricarboxylic Acid (TCA) Cycle of carbohydrate metabolism for medical biochemistry education. *Journal of Science Education and Technology, 28*, 602–612. https://doi.org/10.1007/s10956-019-09790-y

King, S., Boyer, J., Bell, T., & Estapa, A. (2022). An automated virtual reality training system for teacher-student interaction: A randomized controlled trial. *JMIR Serious Games, 10*(4), e41097. https://doi.org/10.2196/41097

Klahr, D. (2013). What do we mean? On the importance of not abandoning scientific rigor when talking about science education. *PNAS, 110*(3), 14075–14080. https://doi.org/10.1073/pnas.1212738110

Klingenberg, S., Jørgensen, M. L. M., Dandanell, G., Skriver, K., Mottelson, A., & Makransky, G. (2020). Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. *British Journal of Educational Psychology, 51*(6), 2115-2138. https://doi.org/10.1111/bjet.13029

Kozhevnikov, M., Gurlitt, J., & Kozhevnikov, M. (2013). Learning relative motion concepts in immersive and non-immersive virtual environments. *Journal of Science Education and Technology, 22*(6), 952–962. https://doi.org/10.1007/s10956-013-9441-0

Lamb, R., Antonenko, P., Etopio, E., & Seccia, A. (2018). Comparison of virtual reality and hands on activities in science education via functional near infrared spectroscopy. *Computers & Education, 124,* 14–26. https://doi.org/10.1016/j.compedu.2018.05.014

Lawson, A. P., & Mayer, R. E. (under review). *Individual differences in executive function affect learning with immersive virtual reality.*

Liu, Q., Tang, Q., & Wang, Y. (2021). The effects of pretraining intervention in immersive embodied virtual reality cardiopulmonary resuscitation training. *Behaviour and Information Technology, 40*(12), 1265–1277. https://doi.org/10.1080/0144929X.2021.1960606

Liu, R., Wang, L., Koszalka, T. A., & Wan, K. (2022). Effects of immersive virtual reality classrooms on students' academic achievement, motivation and cognitive load in science lessons. *Journal of Computer Assisted Learning, 38*(5), 1422–1433. https://doi.org/10.1111/jcal.12688

Lui, M., McEwen, R., & Mullally, M. (2020). Immersive virtual reality for supporting complex scientific knowledge: Augmenting our understanding with physiological monitoring. *British Journal of Educational Technology, 51*(6), 2180–2198. https://doi.org/10.1111/bjet.13022

Lui, A. L. C., Not, C., & Wong, G. K. W. (2023). Theory-based learning design with immersive virtual reality in science education: A systematic review. *Journal of Science Education and Technology, 32*(3), 390–432. https://doi.org/10.1007/s10956-023-10035-2

Luo, H., Li, G., Feng, Q., Yang, Y., & Zuo, M. (2021). Virtual reality in K-12 and higher education: A systematic review of the literature from 2000 to 2019. *Journal of Computer Assisted Learning, 37*(3), 887–901. https://doi.org/10.1111/jcal.12538

Madden, J., Pandita, S., Schuldt, J. P., Kim, B., Won, A. S., & Holmes, N. G. (2020). Ready student one: Exploring the predictors of student learning in virtual reality. *PLoS One, 15*(3), e0229788. https://doi.org/10.1371/journal.pone.0229788

Makransky, G. (2020). The immersion principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed). Cambridge University Press.

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology, 113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019a). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning, 35*(6), 691-707. https://doi.org/10.1111/jcal.12375

Makransky, G., & Mayer, R. E. (2022). Benefits of taking a virtual field trip in immersive virtual reality: Evidence for the immersion principle in multimedia learning. *Educational Psychology Review, 34,* 1771–1798. https://doi.org/10.1007/s10648-022-09675-4

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*, 937–958. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Petersen, G. B., & Klingenberg, S. (2020). Can an immersive virtual reality simulation increase students' interest and career aspirations in science? *British Journal of Educational Technology, 51*(6), 2079-2097. https://doi.org/10.1111/bjet.12954

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019b). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Martella, A. M., & Schneider, D. W. (in press). A reflection on the current state of active learning research. *Journal of the Scholarship of Teaching and Learning*.

Matovu, H., Ungu, D. A. K., Won, M., Tsai, C., Treagust, D. F., Mocerino, M., & Tasker, R. (2023). Immersive virtual reality for science learning: Design, implementation, and evaluation. *Studies in Science Education, 59*(2), 205–244. https://doi.org/10.1080/03057267.2022.2082680

Mayer, R. E. (2022). Cognitive theory of multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press. https://doi.org/10.1017/cbo9781139547369.005

Mayer, R. E., Makransky, G., & Parong, J. (2022). The promise and pitfalls of learning in immersive virtual reality. *International Journal of Human–Computer Interaction, 39*(11), 2229–2238. https://doi.org/10.1080/10447318.2022.2108563

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*(1), 165–173. https://doi.org/10.1037/0022-0663.96.1.165

Office of Educational Technology. (2017, January). *Reimaging the role of technology in education: 2017 national education technology plan update.* U.S. Department of Education. https://tech.ed.gov/files/2017/01/NETP17.pdf

Paas, F., & Sweller, J. (2022). Implications of cognitive load theory for multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 73–81). Cambridge University Press.

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*(6), 785–797. https://doi.org/10.1037/edu0000241

Parong, J., & Mayer, R. E. (2021a). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning, 37*(1), 226–241. https://doi.org/10.1111/jcal.12482

Parong, J. & Mayer, R. E. (2021b). Learning about history in immersive virtual reality: Does immersion facilitate learning? *Educational Technology Research and Development, 69,* 1433–1451. https://doi.org/10.1007/s11423-021-09999-y

Petersen, G. B., Klingenberg, S., & Makransky, G. (2022). Pipetting in virtual reality can predict real-life pipetting performance. *Technology, Mind, and Behavior, 3*(3), 1–16. https://doi.org/10.1037/tmb0000076

Porter, C. D., Brown, J. R., Smith, J. R., Stager, E. M., Nieberding, M., Ayers, A., & Orban, C. (2019). A controlled study of virtual reality in first-year magnetostatics. *2019 PERC Proceedings.* https://doi.org/10.111+/perc.2019.pr.Porter

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147,* 103778. https://doi.org/10.1016/j.compedu.2019.103778

Sanzana, M. R., Abdulrazic, M. O. M., Wong, J. Y., Ng, K. H., & Ghazy, S. (2022). Lecture-based, virtual reality game-based and their combination: Which is better for higher education? *Journal of Applied Research in Higher Education, 14*(4), 1286–1302. https://doi.org/10.1108/JARHE-09-2020-0302

Stranger-Johannessen, E. (2018). Exploring math achievement through gamified virtual reality. In V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, & M. Scheffel (Eds.), *Lifelong technology-enhanced learning* (pp. 613–616). https://doi.org/10.1007/978-3-319-98572-5_57

Su, Y-S., Cheng, H-W., & Lai, C-F. (2022). Study of virtual reality immersive technology enhanced mathematics geometry learning. *Frontiers in Psychology, 13.* https://doi.org/10.3389/fpsyg.2022.760418

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68,* 1–16. https://doi.org/10.1007/s11423-019-09701-3

Tarng, W., Liao, Y-C., & Ou, K-L. (2022). Applications of virtual reality in learning the concepts of special relativity and mass-energy equivalence. *Universe, 8*(12), 618. https://doi.org/10.3390/universe8120618

Taranilla, R. V., Cózar-Gutiérrez, R., González-Calero, J. A., & Cirugeda, I. L. (2022). Strolling through a city of the Roman Empire: An analysis of the potential of virtual reality to teach history in primary education. *Interactive Learning Environments, 30*(4), 608–618. https://doi.org/10.1080/10494820.2019.1674886

Ward-Horner, J., & Sturmey, P. (2010). Component analyses using single-subject experimental designs: A review. *Journal of Applied Behavior Analysis, 43*(4), 685–704. https://doi.org/10.1901/jaba.2010.43-685

Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments, 24*(6), 1319–1333. https://doi.org/10.1080/10494820.2014.994533

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness in immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology, 51*(6), 1991–2005. https://doi.org/10.1111/bjet.13023

Yang, Q-F., Lin, H., Hwang, G-J., Su., P-Y., & Zhao, J-H. (2022). An exploration-based SVVR approach to promote students' chemistry learning effectiveness. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2022.2135106

Yu, Z (2021). A meta-analysis of the effect of virtual reality technology use in education. *Interactive Learning Environments.* https://doi.org/10.1080/10494820.2021.1989466

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Contract cheating in higher education: Impacts on academic standards and quality

Kelum A.A. Gamage[A]

Shyama C. P. Dehideniya[B]

Zhiheng Xu[C]

Xiaobin Tang[D]

A Centre for Educational Development and Innovation, James Watt School of Engineering, University of Glasgow, Glasgow, UK

B Department of Education, University of Peradeniya, Peradeniya, Sri Lanka

C Department of Nuclear Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China; Key Laboratory of Nuclear Technology Application and Radiation Protection in Astronautics, Ministry of Industry and Information Technology, Nanjing, China

D Department of Nuclear Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China; Key Laboratory of Nuclear Technology Application and Radiation Protection in Astronautics, Ministry of Industry and Information Technology, Nanjing, China

Keywords

Academic integrity;
academic standards;
assessments;
contract cheating;
Covid-19;
higher education;
quality assurance.

Correspondence

scpkdehideniya@gmail.com [B]

Article Info

Received 27 June 2023
Received in revised form 28 August 2023
Accepted 29 August 2023
Available online 31 August 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.24

Abstract

Contract cheating has become increasingly an issue as universities adapt to online and hybrid teaching, learning, and assessments. Due to the COVID-19 pandemic, higher education institutions began to administer examinations more frequently online, and it led to the emergence of websites and commercial service providers who offer contract cheating services globally. In this paper, we examine the key elements that lead students to turn to contract cheating as well as the elements that deter the students from engaging in such unethical behaviour. We also investigate how assessment design can encourage authentic learning, although assessment design alone cannot eliminate contract cheating. The effects of contract cheating on academic standards and quality assurance are also examined. Mainly the study results show that the act of contract cheating is a result of interrelated internal and external factors in an individual. Although a number of measures, including authentic evaluations and digital tools, have been implemented to discourage students from cheating, no strategy is strong enough to control the issue permanently. Hence, academic integrity is still not assured, highlighting the necessity of a global movement to solve the problem.

Introduction

Academic dishonesty – "cheating or plagiarism that gives a student an illegitimate advantage during an assignment or assessment" (Bleeker, 2008) – is a type of unaccepted behaviour by the academic community that has existed for millennia at every educational institution, from schools to higher education regardless of geographical, economic, or cultural boundaries. Research literature notices various forms of students outsourcing their academic work, and now it has become a globally growing issue in educational contexts. In general terms, types of outsourcing have been categorised into four groups: copying, plagiarism, collusion, and cheating (Guerrero-Dib et al., 2020). Whatever form is used, it leads students to be involved in violations of academic values and standards. The International Center for Academic Integrity (ICAI) defines academic integrity as the "commitment to secure six fundamental values, namely honesty, trust, fairness, respect, responsibility, and courage" (Fishman, 2014, p. 14). These values decide ethical academic behaviour that forms a community committed to learning and honestly exchanging ideas (Holden et al., 2021).

As Harper et al. (2019) mentioned, up to the late 1990s, the subcontracting of assessment was necessarily confined to hidden steps that were limited among peers. With the integration of technology into education to upgrade the quality, new opportunities for "e-cheating" (Holden et al., 2021) were created, increasing the students' tendency for plagiarism, ghost-writing, and contract cheating significantly. In this background, contract cheating emerged as a new challenge since it continued to develop into a large-scale commercial trade. Although the practice of students' paying for assessments began in the 1940s and 1950s, it is evolving at a fast pace (Lancaster, 2019) as education has become a commodity to be acquired (Harper et al., 2019).

Typically, contract cheating involves paid anonymous individuals or a company to complete academic work, and the term has been extensively used globally for more than ten years' time (Lancaster & Clarke, 2007). The phrase 'contract cheating' was first coined in 2006 by Clarke and Lancaster when a student pays someone else to complete their assessment (Rundle et al., 2019; Bretag et al., 2019a; Erguvan, 2021). Accordingly, contract cheating is the submission of work by a student, which contributes to their degree programme, in which they have paid someone unknown to complete their assessment. Eaton (2022) views contract cheating as not an act of individual students in a course making poor choices but as a business, whereas Williamson (2019) interprets it as a particularly insidious method of cheating because it is completely intentional and very challenging to discover. The term 'contract cheating' has now progressed to encompass several practices relating to subcontracting students' academic work to third parties (Bretag et al., 2018). In the recent definition put forward by Newton (2018), contract cheating results in a relationship between a student, their university, and a third party who supports completing the assessments for a fee.

In the act of plagiarism, students intentionally use others' perceptions without acknowledging the original writer. In addition, they lose the chance to learn, practice the skills required, and most importantly, the opportunity to receive valid feedback on their academic performances (Singh & Remenyi, 2016).

Contrarily, the terms ghostwriting: the practice of hiring a writer or writers for the purpose of academic writing (Singh & Remenyi, 2016), and contract cheating (Ali & Alhassan, 2021; Lines, 2016; Tauginien & Jurkeviius, 2017) are interchangeably used to refer to the act of academic cheating (Erguvan, 2021; Ali & Alhassan, 2021). Additionally, according to Tauginien and Jurkeviius (2017), the terms contract cheating, essay mill, paper mill, and unethical tutoring are all interchangeable in the literature. Erguvan (2021) disagrees with this use of the phrases as they do not have the same meaning. In light of the aforementioned information, the term 'contract cheating' is used in this article to describe a practice whereby students hire third parties for scholarly projects, whether they are paid or not.

In higher education, examinations exist as measures of learning, and academic misbehaviour within the process weakens the acceptability of the qualification. When assessment processes cannot provide trusted results, it poses a challenge to the validity of qualifications and the trustworthiness of certificates and degrees (Goff et al., 2020; Martin, 2017). Similarly, there is a joined risk in the trust that society has in educational institutions (Comas-Forgas et al., 2021). Hence, factors such as maintaining high academic standards, academic integrity, and quality assurance have been identified as crucial to minimising contract cheating by higher education institutions.

Most recently, technological improvements in the socio-economic context of tertiary education (Lines, 2016) have led to an increase in cheating behaviours. Noticeably the advent and the expansion of the Internet and its facilities changed how contract cheating occurred globally (Erguvan, 2021; Eaton, 2022). Many higher education institutions have used online or distance learning platforms of instruction for years. Sometimes, students are required to complete assessments in an environment without close supervision, and students have a considerable number of chances to cheat on their work. It is believed that online testing offers additional cheating opportunities as compared to traditional, live-supervised classroom environments (Holden et al., 2021; Lancaster & Clarke, 2014; Slade et al., 2019). This has been worsened by the COVID-19 pandemic, particularly in universities and almost all education institutions that had to transition to online teaching and assessments.

When ethics are concerned, highlighting the outcomes of recent studies Comas-Forgas et al. (2021) suggest that there is a close association between academic dishonesty and professional dishonesty. Further, empirical evidence from research studies has demonstrated that students who engage in dishonest activities in classrooms and/or examinations, particularly undergraduate students, are more likely to establish unsuitable and unacceptable behaviours during their professional life and vice versa (Guerrero-Dib, 2020; Hill et al., 2021). Moreover, Orosz et al. (2018) identified a strong affiliation between academic dishonesty and the level of corruption in a country. In that sense, negligence of cheating behaviours in higher education seems to stimulate corruption and dishonest behaviour. On the other hand, honesty is a highly valued personal quality that begins within the individual, especially as a result of education and extends into the community through practices. Violation of 'honesty' in an academic setting may have a negative impact on society.

Consequently, contract cheating has become increasingly a global issue in the higher education setting, and the focus of the present study is to examine the concept giving importance to the following areas.

- What are the primary factors that influence a minority of students to resort to contract cheating and keep the majority of students away from such dishonest behaviour?

- How can assessment design encourage authentic learning and minimise contract cheating?

- What are the impacts on academic standards and quality assurance due to contract cheating?

Contract cheating

Contract cheating is academically unethical and totally unacceptable. It is considered one of the most serious breaches of academic integrity (Eshet, 2022), which continually increases the suppression of other forms of cheating acts. Hence on a global scale, higher education institutions are trying to tackle the issue with various remedial measures (Erguvan, 2021). Contract cheating behaviours among students may come in many models (Hill et al., 2021) and can appear in any kind of printed or handwritten work (Erguvan, 2021).

Some authors believe that contract cheating necessarily engages a financial transaction (Walker & Townley, 2012) between a customer (student) and a service provider (company), whereas others consider it as a learner outsourcing their work with no money involved for the service they receive (Hill et al., 2021; Eaton & Turner, 2020). In line with Harper et al. (2019) and Lancaster and Clarke (2016), students can use essay writing services or get support from peers, family members or private mentors, and many other outsiders. According to Erguvan's (2021) observations on many occasions of reported contract cheating, colleagues have exchanged their work with each other just as a favour or as a help. According to recent research, students prefer to turn to their close friends and family members for assistance in completing assessments rather than looking for commercial service providers since paying for the work is not always required by close ones (Armond & Varga, 2021). Hence, contract cheating does not depend on money exchange at all times (Hill et al., 2021) and can be funded or not funded (Curtis et al., 2018; QAA, 2020) and the funded contract cheating is referred to as "commercial contract cheating" (Rundle et al., 2019).

Contract cheating is a branch of a massive universal academic business (Lancaster, 2020), and service providers are to be found mainly in the English-speaking Western world (Lines, 2016; Amigud & Dawson, 2019). The popularity of contract cheating services is increasing, and it is effectively involved in advertising to students at all study levels, using advances in digital technology. The ever-growing visibility and highly attractive marketing and advertising techniques of essay mills have made the customer attracted to contract cheating services (Erguvan, 2021). Ease of purchasing at a low cost and quickness (Wallace & Newton, 2014) have increased students' temptation to cheat. When considering the discipline types, Business and Computing studies demonstrate a higher number of contract cheating transactions (Lancaster, 2020). However, it is evident that contract cheating service providers are already deep-rooted in all the subject areas at almost all levels of study. At the same time, students also seek the support of those services to pass the barriers in their academic path created by the socio-economic and cultural contexts.

The role of culture and the internet age in contract cheating

Technological and economic development has dramatically changed the social structures creating a competitive socio-economic environment. The situation has worsened with the emergence of online social networks. These radical changes have indirectly approached the young generation exerting extra pressure on their academic life, forcing them to excel since academic achievements or paper qualifications have become a deciding factor to win the competition in professional life. Hence, the majority attempt to achieve their academic targets by any means at any cost, and as a result, students seek the assistance of commercially available third parties to complete their assignments, essays, and projects, whereas, within current socio-economic contexts, most potential and skilled individuals are always benefited. On the contrary, the commodification of higher education and e-commerce are backing the rapid growth and popularity of a 'sharing economy' (Williamson, 2019; Bretag et al., 2018).

New forms of customer behaviour and sharing economy have aggravated the situation giving rise to academic cheating behaviours globally. Simultaneously, under extreme stress conditions, an increase in contract cheating can be observed (Bretag et al., 2018). The integrity of higher education is affected by a number of factors ranging from a reduction in public funding to increased marketisation and internationalisation, availability of disruptive technologies, and unsecured job markets (Hill et al., 2021). Furthermore, the arrival of digital technologies triggered this speedy decline in academic integrity (Erguvan, 2021; Ison, 2020; Lancaster & Clarke, 2014) as the 'sharing economy' facilitates anyone to outsource any kind of work or goods and services virtually (Bretag et al., 2019a). That has intensified as the millennials have grown up with online environments that encourage sharing information, which progressively encourages cooperative learning approaches allowing students to share information and their work with each other (Lines, 2016).

Several studies have tried to find the root cause for online students' increased engagement in contract cheating. The Internet can cover the separation between honest and dishonest behaviour in academic work; in particular, this comprises the issue of psychological distance, which unfavourably affects interpersonal social relationships and introduces ethical distancing (Ahsan et al., 2021). According to the revealed facts, the current socio-cultural context and the internet collectively affect the students' cheating behaviours. There are many resources, particularly on the internet, that encourage students to engage in dishonest academic behaviour. Hence, culture and the internet age are important in this discussion as the factors that influence academic cheating.

Online/hybrid learning and assessment environments

In response to COVID-19, educational institutions suddenly shifted from in-person mode to emergency remote teaching and learning, removing the academics and students from their usual teaching and learning environments. Students lost their regular contact with peers and teachers, making them feel more vulnerable and isolated. In addition, students had to take their academic evaluations in online mode, and many had made it an opportunity to cheat, aiming for higher grades. Particularly, students supposed that cheating in virtual examinations was more stress-free than the ones held in face-to-face settings (Erguvan, 2021). The data presented in Erguvan (2021) has shown an increase in interest in internet searches for online exam cheating in Spain with the onset of the world pandemic. They emphasise some other research outcomes and suggest that online exams, regardless of the medium of instruction, are vulnerable to breaches of academic integrity.

In addition to that, Comas-Forgas et al. (2021), Erguva (2021), and Ahsan et al. (2021) confirm that the problem of contract cheating has never been as severe as during the Covid-19 pandemic with a speedy growth of many novel methods distinctive to the online learning contexts (Holden et al., 2021). Parallel to that, many third-party contract cheating service providers have aimed students to get the benefit of the uncertain and anxious mindset of students (QAA, 2020). As a whole, Eaton (2022, n.p.) mentions that "in the course of the Covid-19 crisis, we have certainly seen increases in violations of academic integrity", with the increased temptation of students to engage in contract cheating (Ahsan et al., 2021). Relating to that, Eshet's (2020) study results have also shown a substantial decrease in the level of academic integrity during the period to the closure of the first rise of COVID-19 outbreaks.

Undoubtedly, speedy and continuous internet access has altered the students' strategies of learning, engaging with study materials, researching, and producing their own work (Lines, 2016), throughout the pandemic without the direct support and assistance of the teachers or the instructors and lacked chances to develop face-to-face interactions with colleagues (Awdry & Newton, 2019). They all collectively generated dissatisfaction in students towards online practices, and as a side effect of COVID-19, contract cheating became a significant issue in higher education. Studies from various geographical and economic regions revealed that the negative consequences of online education have forced students to cheat. For instance, studies by Mok et al. (2021) and Tran et al. (2021) respectively identified why contract cheating has rapidly increased, taking examples from Bangladesh, Hong Kong, and Vietnam. In accordance with the results, during the pandemic, students emotionally struggled while some showed signs of depression. Further, they were unsatisfied with online education and they were incompetent in technology or with limited access to technology and related hardware to complete online assignments. The growth of contract cheating during the pandemic is a result of a number of factors, for example, academics not setting assessments as appropriate for e-assessments, lack of understanding of students networking through various modes, including social media, students' increased stress levels, and advertising by contract cheating service providers (Eaton & Turner, 2020).

Hence, identifying the negative impacts of contract cheating on academic integrity, and the trust of the general public towards graduates, their professional lives, and education institutes, many approaches have been taken by the respective authorities to minimise it.

Role of technology and social media in contract cheating

Among various factors that stimulate contract cheating, the role of technology and social media is indispensable. The addition of high-quality features to social media has opened up a number of ways and means for students to identify various options for carrying out plagiarism (Bautista & Pentang, 2022) and contract cheating. Social media and intermediary websites work side by side to make a bridge between students and providers who supply contract cheating services (Amigud & Dawson, 2019). Contract cheating has been advanced in two ways as a result of technological advancements (Rigby et al., 2014). First, technical modification has directed cheaters into the contract cheating services because the likelihood of uncovering old-style cut-and-paste and secondhand papers has increased with the use of plagiarism detection software such as Ouriginal, Turnitin, etc. Second, the internet has minimised customer search costs while facilitating quick ordering, payment, and provision.

According to Comas-Forgas et al. (2021), YouTube is one of the best measures of the prevalence of cheating on examinations during the pandemic because a considerable number of videos relating to such experiences can be discovered there. Also, many facility providers are using social media apps such as Twitter to attract students when they are in their weakest states (Amigud & Lancaster, 2019). However, recent studies have found that students using contract cheating services are at risk of being tricked or bribed (Lancaster, 2018). So, it is critically important to improve student awareness of the risks caused by engaging with contract cheating sites (Dawson et al., 2019). Hence, rather than punishing after the mistake, it is wise to take remedial measures at the bottom level. But then again, detecting and proving the cheating act is considered difficult because of the advanced nature of the services. Therefore, determining the part that technology plays in encouraging academic fraud is important. On the other hand, doing so will help in formulating strategies for preventing the practice through the technology itself.

Practices and approaches taken by higher education institutes to minimise contract cheating

Globally, universities and other educational institutions are struggling to control contract cheating. International concerns towards contract cheating are continuously increasing, and they include government actions and legal actions (Lancaster, 2020). Various judicial steps have been taken to deal with contract cheating service agreements. In some states, the act of subcontracting is treated as a crime with legal consequences (Amigud & Dawson, 2019). In the United Kingdom, there have been nationwide requests for action to sanction the supplies and promotions of essay mill cheating service providers (Morris, 2018) and to enact new rules and regulations aimed at contract cheating providers (Draper & Newton, 2017). Since 2022, it has been illegal to use and provide contract cheating services.

Apart from that, a study done at the University of California has emphasised the need to make students aware of the implications of being involved in cheating (Reddin, 2021). In addition, students are asked to sign a special statement before the examinations, and as mentioned in the statement, if a student is accused of cheating, then he or she may lose studentship. Significantly the particular action was a success and has reduced the number of cases. Australia, Europe, and the UK have also made large-scale efforts to combat contract cheating (QAA, 2020). In the meantime, as Eaton (2020) shows, Canadians' attempts to solve the issue of commercial contract cheating have been narrowed to local or regional levels, and quality assurance authorities maintain rather a cold reaction.

As Erguvan (2021) found, the Kuwait Ministry of Commerce has banned businesses that are involved in selling academic papers, projects, and other technical work on a few occasions following objections from the Ministry of Education. and particularly during the pandemic, most of them continued their services through virtual modes. According to the noted facts, it is evident that the evolution of the contract cheating industry is very fast and legal approaches are solely unable to control it since it is always stimulated and supported by technology and social media.

Detecting contract cheating

Generally, contract cheating is difficult to identify and prove. Many researchers have shown that effective proof of contract cheating is largely dependent on the experience of the assessor and the knowledge of the student (Rogerson, 2017) in terms of the student's subject knowledge levels and writing style. In line with the findings of Erguvan's (2021) study in Kuwait, academics have the potential to detect an assessment that has been completed by someone else considering the standards of the completed work with respect to the actual ability of the student (including academic and linguistic abilities) as well as the technical details of the file submitted. In contrast, some expressed that contract cheating is a critical and complex area to identify, and proving such a case is a long and sometimes difficult, time-consuming process (Ali & Alhassan, 2021; Awdry & Newton, 2019).

Effective and efficient solutions for detecting contract cheating are still not being found, but many are in use with their plus and minus points. Among them, technology-based remedies and detection programmes are at the forefront of all. Nevertheless, word-matching detection applications such as Turnitin, PlagScan, AntiPlag, TeSLA, and Urkund could use to recognize subcontracted academic work (Lancaster & Clarke, 2016; Wang & Xu, 2021), they are recognised as unsuccessful in detecting contract cheating (Ahsan, 2019) as work done by those services are normally skillfully written and sufficiently referenced (Lines, 2016). Software tools, such as Cadmus (Lines, 2016) and digital forensic methods – stylometrics and linguistics (Dawson et al., 2019; Ison, 2020) – may help address contract cheating. In particular, as reported in Eshet's (2022) study, although the software was capable of detecting direct copy-paste, tracing a custom-made one is not always possible. Even when using state-of-the-art automated detection methods, contract cheating remains difficult to detect. Besides, Amigud and Dawson (2019) mention that the use of text-matching apps has been found ineffective, as contracted assignments normally cover original content, making it hard to identify cheating. At present, there is no efficient tool or application to recognise any kind of cheating, and technology is not evolving to limit contract cheating (Erguvan, 2021; Hill et al., 2021).

Interviewing the student at the end of the assessed work, introducing remote invigilation using webcams or facial recognition apps, password-protected or sound-recognition applications, online or telephone questioning, or third-party confirmation are some of the approaches suggested to avoid contract cheating during online examinations. Although online supervising of remote examinations is possible through biometric data, eye movement, and keystroke tracking (Hill et al., 2021), it can be detrimental to students' psychological well-being (Eaton & Turner, 2020) due to violation of privacy concerns. However, the financial cost of software and other technology tools, varying policies, or not-so-user-friendly features of these applications (Erguvan, 2021) limit their usage. Furthermore, blocking certain websites on institutional devices, petitioning governments to sanction the supply and marketing of cheating services (Morris, 2018), introducing legal remedies, imposing financial punishments, and banning advertising (Tauginienė & Jurkevičius, 2017), and punishing customers and/or suppliers are some of the strategies proposed by researchers to block service providers. Accordingly, new approaches are wanted to sense subtler potential signs of contract cheating (Eshet, 2022). In addition to that, academics should be made aware of continual developments in the contract cheating industry. As a whole, knowledge of detecting contract cheating is important since it has a direct impact on academic quality and standards.

Methodology

This study is grounded on a systematic review of available work related to contract cheating in higher education and its impacts on academic standards and quality. An extensive literature search was done on online databases, namely: Google Scholar, JSTOR, Taylor and Francis Online, Elsevier and recognized official websites. The pre-decided selection conditions were used during the database search in order to keep the number of resources reasonable and adequate. To ensure the quality of the sources the search was limited to peer-reviewed journal articles, conference papers, and reports found in full text in institutional websites. Blog posts, books and physically available sources were not included. No limitations on the publication time or the geographical areas were considered and the language of selected sources was limited to English. In addition, experimental and theoretical studies were taken into account during the selection process regardless of the type of study methodology (quantitative, qualitative, or mixed).

A significant amount of research has been selected to examine the concepts of contract cheating and its impacts on academic standards. In the search, to ensure quality and the appropriateness of sources for the review, a wide variety of key terms and phrases were used in selecting items. They include mainly "contract cheating", "contract cheating – academic integrity", "contract cheating – higher education", "detecting contract cheating", "contract cheating – technology", "contract cheating – online/hybrid teaching and learning", "contract cheating – Covid-19", "contract cheating – assessment", and "contract cheating – reasons/factors". Furthermore, similar terms related to the examined issue, such as ghostwriting, essay mills, plagiarism and tertiary/university education, that often appeared interchangeably in texts, were also used in finding sources. As the search action resulted in an inadequate number of suitable and reachable sources, the reference sections of the found texts were used in the search for more relevant resources. As a result of the search, 126 documents were downloaded, and their titles, abstracts, and textual contents were studied in detail to extract the most appropriate sources. This caused the removal of duplicate sources and articles with dissimilar content, and the initial sample was reduced to 66 documents for analysis. The content of the selected resources was studied comprehensively and analysed in detail. The review mainly focused on addressing the following specific research questions mentioned in the introduction:

- What are the primary factors that influence a minority of students to resort to contract cheating and keep the majority of students away from such dishonest behaviour?
- How can assessment design encourage authentic learning and minimise contract cheating?
- What are the impacts on academic standards and quality assurance due to contract cheating?

In order to collect the necessary data, each paper was studied in detail, examining the content and extracting any relevant information to support the research questions. The data were coded as "reasons for contract cheating", "minimising contract cheating", and "impact of academic standards" to reduce the risk of missing important information. Then they were organized to build up the answers for each question. As the final step, analysis and the discussion on revised data were done to come to conclusions and to identify further research directions.

Results and analysis

Grounded on the directions provided by the background literature, three key issues identified were:

- why students are involved in contract cheating and the factors that keep students away from contract cheating;
- how assessment design motivates or demotivates contract cheating behaviour; and
- the impacts of contract cheating on academic standards and quality assurance.

These issues were addressed and further discussed with the aid of the available academic literature.

Why are some students more motivated to cheat than others?

A significant number of studies have been completed to understand why learners cheat and why they do not (Bretag et al., 2019a; Amigud & Lancaster, 2019; Ahsan et al., 2021; Harper et al., 2019). Basically, the general theory of crime proposes that the failure of self-control is the foundation for unethical behaviour. On the basis of this, it is also possible to rationally explain the case of contract cheating behaviour. A theoretical foundation is also provided by routine activity theory and the rational choice perspective, which run parallel to the general theory of crime, and they also explain why anyone can engage in unconventional behaviour like contract cheating (Eshet, 2022). However, Curtis et al. (2018) argue that the prevailing theory-based studies of contract cheating have been criminological rather than psychological. As Beckman et al. (2017) suggest, the two principal factors that permit contract cheating to take place are "motivation" and "opportunity". Further, some other studies have proposed "personal, institutional, medium-based, and assessment-specific, contextual, pedagogical, ideological and socio-cultural" as motivational factors for cheating (Holden et al., 2021; Ali & Alhassan, 2021). Generalising the reasons for the choice of contract cheating, Bretag et al. (2019a), Brimble (2016) and Lines (2016) mention the insights that there are lots of chances to cheat, increased availability of contract cheating services, students' misunderstanding that cheating is easy and will not be caught, challenging workloads and assignment difficulties, and lack of inspiration and personal factors: gender, personal temperaments, age, grades or scores and to help friends.

Accordingly, contract cheating appears to be activated by an array of influences ranging from social, economic to cultural, and from educational, academic to personal (Ali & Alhassan, 2021). Hence, particular to this study the primary factors that influence contract cheating are discussed under two main categories, intrinsic: personality factors, and extrinsic: pedagogical, institutional, and socio-cultural factors.

Personality traits or factors are frequently acknowledged in the literature as predictors of why students are involved in academic delinquencies (Rundle et al., 2019). Holden et al.

(2021) highlight three specific conditions referred to as the 'fraud triangle': (1) opportunity, (2) motivation, pressure, or requirements, and (3) rationalisation or attitude, as the personal or individual factors that predict cheating behaviour. Further, students' intentions for learning vary, and they are mostly under pressure with a number of academic projects since it may be the deciding factor of their future. The personal inability to manage time (Rogerson, 2017) brought by life complexities is a critical issue that students undergo. Procrastination, a tension between 'learning for learning's sake' and a 'getting through it adequate to graduate' attitude (Blum, 2016), may lead learners to be involved in cheating behaviours. Erguvan's (2021) study reveals that laziness and the desire to get high scores or grades with little or no effort are some factors linked with contract cheating. High stress created due to a competitive mindset, high self-esteem and fear of losing social respect may also motivate cheating. Slade et al. (2019) have identified student circumstances as one of the central causes of contract cheating, and those supported with time pressures, personal difficulties, and a history of poor academic records may encourage students to use contract cheating services (Amigud & Lancaster, 2019; Eaton, 2020). In addition to that, cheating may be typical personal behaviour or simply a feature of one's personality. Coupled with that, low conscientiousness and no fear or shame of detection of cheating and its consequences may also attract students. Anxiety, a lack of confidence in academic writing and conventions of the subject, and fear of failure also trigger cheating behaviours in students. Misleading expectations that cheating will bring positive results, normalisation of cheating and the idea that others are doing it successfully (Ahsan et al., 2021) may possibly encourage students to subcontract their academic work.

As the commercialisation of education has broken the geographical limits, many students acquire their higher qualifications from foreign countries in which they learn in non-native languages. There is a common idea that students not learning in their mother tongue are more likely to cheat. For example, Bretag et al. (2019a) and Amigud and Lancaster (2019) found that not being a native speaker of the medium of instruction and lack of language proficiency a cause of contract cheating. Several study results show that self-reported commitment to contract cheating was related to disappointment with the learning and teaching and the misperception that opportunities to cheat are there within the project or assignment. Also, a lack of engagement with studies results in a lack of understanding (Curtis & Vardanega, 2016) and makes students incompetent in terms of an assessment's requirements and subject knowledge. Overwhelmingly difficult assessment tasks, decontextualiaed assessments (Ahsan et al., 2021) that involve higher assessment weightings, have limited timeframes, and offer fewer chances for comments (Slade et al., 2019) are other motives for contract cheating. Bretag et al. (2018) confirm the time issue further, stating that too much material is covered in too short a time, and the short turnaround times on assessments may probably increase the tendency for cheating.

Institutional factors and policies related to academic standards and integrity directly impact building a culture of cheating. For example, inadequate sanctions and punishment of academic dishonesty, too simple institutional policies, an inadequate effort made to advise students about these policies, and a lack of understanding of staff members about the policies against academic misconduct provide ample opportunities for students to normalise the cheating behaviours and unconscious promotion of cheating can result. Accordingly, wherever the opportunity is available, students think they can cheat unnoticed (Holden et al., 2021; Bretag et al., 2018). Agreeing with the argument further, Holden et al. (2021) mention that negligent or inadequate penalisation of academic dishonesty, insufficient awareness of policies and standards among students, instructors, and administrators, and unsatisfactory efforts to notify students about these policies and standards motivate students to contract cheating. Husain et al. (2017) also approve that state student perception of staff apathy, knowledge and dedication, and students' awareness regarding the lack of institutional support for academic integrity increase contract cheating. Additionally, the issue becomes even worse when students realise the lenient approaches of educators with regard to cheating and shortcomings in how such behaviours are handled. Empirical research demonstrates that when academic staff or the university expresses little to no concern, students are more likely to justify cheating (Harper et al., 2019). Similar results have been obtained in a study conducted on Iranian ELT students (Husain et al., 2017), and it revealed that having kind and student-friendly academics is a key reason for engaging in different forms of academic cheating, including plagiarism. The cultural and social pressure on students to achieve a higher academic profile has severely affected the occurrence of cheating habits. Currently, competition is a part of the academic system, and parents demand good grades in the examinations. Other than parental pressure, team member issues and influences are attached to outsourcing issues (Ahsan et al., 2021).

As far as the reasons for not being involved in contract cheating are concerned, some researchers highlighted a number of primary reasons that discourage students' involvement in contract cheating: opportunity, fear of detection and punishment, trust, motivation for learning, time management, morals, and norms (Rundle et al., 2019). Studies by Curtis et al. (2018) have empirically found that "higher levels of self-control were protective against student engagement in cheating behaviours". Moreover, students do not engage in contract cheating for a variety of reasons, including their inability to rationalise the actions of outside sources or lack of faith in them (Rundle et al., 2019). Furthermore, the perceived seriousness of unethical behaviour and acceptance of the perceived social norms, especially those of the person's gang or peer group (Curtis et al., 2018; Rundle et al., 2019) might prevent students from looking into shortcuts to complete their work. Rigby et al. (2014) revealed that students who see the benefits of the study are more hesitant to misconduct than those who do not. Contrastingly, Awdry and Newton (2019) found individual factors, discipline, and country do not predict contract cheating and Erguvan (2021) also mentions that the participants in their study have not linked personal factors to the rising numbers of contract cheating.

How can assessment design encourage authentic learning?

Outcomes of assessments and evaluation of students' learning are an important indication of the quality of the instructional process, and the type of assessment likely affects the result depending on the individual's performance. Similarly, the number of cheating acts would be expected to vary according to the assessment type (Holden et al., 2021). Bretag et al.'s (2019a) study report provides strong experimental evidence for conceptualising the interconnection between contract cheating and assessment. Findings show that no assessment type can effectively eliminate the likelihood of being cheated. Lancaster and Cotarlan (2021) highlight that many practices of assessment and examinations are vulnerable to contract cheating, particularly when courses are taught and assessed online. To minimise and prevent such vulnerabilities, new forms of assessment tools and techniques are vital. Although reasonable and practical methods for minimising academic dishonesty have long been taken by educational institutions and policymakers, it continues at higher rates with the advancement of technology. However, Lancaster and Cotarlan (2021) argue that many assessment and examination techniques are open to contract cheating in online learning environments and that necessitates the use of novel assessment technologies. As they further elaborate, although educational institutions and authorities have long adopted reasonable and feasible measures such as authentic and personalised assessment tasks to reduce academic dishonesty, it persists at greater rates. According to Ahsan et al. (2021), the institution, the academic, and the student form the assessment supply chain. When the learner subcontracts the assessment, partially or fully, he/she has broken the contractual relationship. Therefore, it is essential to take necessary measures to safeguard the smooth flow to obtain the true estimation of students' ability levels while keeping academic integrity. Hence, many scholars and researchers have stressed the need for changing teaching pedagogies and assessment designs. As they suggest, assessment should be strong enough to minimise the possibility of cheating while providing space for the learner to show their knowledge and skill levels. Furthermore, Holden et al. (2021) emphasise the need to pay attention to the assessment format and the presentation. Format, content, declaration of academic integrity, alternative forms, and standard design for the number of assessments that count towards final grades are considered under the assessment structure. On the other hand, limited space of availability, time limits, disabled copy/paste functionality in assessment software, preventing referring to previous items, and response option randomisation are factors closely observed during the assessment delivery.

In the discussion of assessment, designing 'authentic assessments' has been recommended by numerous authors, though authentic assessment methods are still vulnerable to contract cheating. As expected, an authentic assessment may limit the impact of cheating since the students will have to actually use their knowledge and skills. Such assessment makes it more challenging to complete relying on contract cheating services. In contrast, the study of Ellis et al. (2019) has provided strong experimental proof to show that authentic assessment tasks do not guarantee academic integrity. For example, a candidate has to face his/her own in practical exams, face-to-face assessments, oral examinations, or presentation of written assignments. In-class tests and invigilated exams (Lines, 2016), designing assessments with specific contextual requirements (Bretag et al., 2019b), and adopting the assessment to the context (Eaton, 2020) are also proposed as remedies for contract cheating issues.

Although time pressures have been found to be a reason for students choosing to use contract cheating services (Wallace & Newton, 2014; Slade et al., 2019; Amigud & Lancaster, 2019), it seems to be acceptable to minimise opportunities for contract cheating by having short turnaround times for assessment submission (Bretag et al., 2019a). Furthermore, regarding online exams, preventing the use of supplementary electronic resources during exams and hindering students from using external websites or using unauthorised applications on the same machine that is used to take the exam probably limits students' engagement in cheating.

Impacts on academic standards and quality assurance

The issue of cheating is not specific only to higher education, but it affects all categories of education institutions regardless of the disciplines and study levels. Higher education providers are responsible for ensuring the quality of their services. On the other hand, the qualification offered by the institute essentially needs to meet nationally and globally accepted standards. Further, the assessment outcome should essentially show the students' true achievement level. However, as a result of contract cheating, students can potentially achieve degree qualifications that do not tally their knowledge and skill set (Bretag et al., 2019b). According to Rigby et al. (2014), contract cheating causes information anomalies, and it has a negative economic impact on graduate attributes by lowering degree grades. In that sense, contract cheating raises an alarming risk towards the reliability of the student's qualification and skill levels and undermines the validity of the student's knowledge evaluation (Jurkevičius & Tauginienė, 2017).

Considerable social mistrust in universities' quality assurance mechanisms (Dawson et al., 2019) in terms of academic quality and standards, assessment system (Slade et al., 2019), trustworthiness and reliability of the institution (Harper et al., 2019; Lancaster, 2019; Slade et al., 2019) create a chain of issues such as destroying community confidence in higher education standards (Jurkevičius & Tauginienė, 2017; Hill et al., 2021) and graduate reputation and credibility related issues (Slade et al., 2019). Prospective employers may not keep faith in universities to generate skilful graduates who are ready to work (Hill et al., 2021). Further, contract cheating leads to the deprivation of fair competition and demotivation to study honestly and is disadvantageous to honest, diligent students due to unmerited academic credits earned by cheating (Jurkevičius & Tauginienė, 2017). It also demotivates staff, adversely affects student equity, undermines employee and employer morale, and presents a serious threat to society as underqualified graduates end up as working professionals (Ahsan et al., 2021; Slade et al., 2019). For example, future doctors, engineers, and social workers who have contracted out their academic work could pose a serious risk to society as states defrauding future employers and career disruption is the result (Jurkevičius & Tauginienė, 2017; Bretag et al., 2019a).

Hill et al. (2021) describe the extent of the influence of contract cheating in detail based on the study related to COVID-19. As they analysed, the students who use the 'services' do not develop essential skills but still receive grades without necessary effort. Another adverse effect highlighted is that the teachers or the instructors who notice and report the cheating acts are losing valuable resources that can be owed to the development of teaching resources, and academics who do not act in the same way are seen as inexperienced by students. Similarly, the universities that act against contract cheating might have lower enrolment, while universities that neglect to act upon cheating might face the issue of letting down academic standards.

Therefore, institutions of higher education need to recognise why students are involved in contract cheating, and then they need to make changes (Williamson, 2019) accordingly to minimise the damage to academic standards and quality. Comparably multifaceted resolutions are vital, including academics and civil society (Hill et al., 2021), to address the global issue.

Discussion and conclusions

The review reveals that academic contract cheating needs to be answered globally due to the high competitiveness in obtaining academic qualifications, the commercialisation of education, and the tendency for e-education. On the other hand, it is a result of the impact of a number of internal and external factors on the student. In addition, there is a noteworthy increase in contract cheating and service providers during the recent pandemic. The influence of each on students' contract cheating habits has been studied adequately, but as Erguvan (2021) highlights, the impact of circumstantial or background features such as society, culture, and religion on cheating behaviours has not been sufficiently studied. However, the presence and extent of cheating depend on the intensity with which the factors influence the student. For example, if institutional parameters such as detection probabilities and penalties are at a low rate, then there is a chance for the student to outsource the academic work to achieve a higher grade. Hence contract cheating can be minimised by blocking the opportunities and motivations.

The research outputs evidence that contract cheating is considerably difficult to identify and confirm (Ahsan et al., 2021). Correspondingly and many forms of assessment and examinations are susceptible to contract cheating, especially where courses are taught online (Lancaster & Cotarlan, 2021). Furthermore, no discipline area is immune from contract cheating (Lancaster, 2020). Therefore, to preserve academic integrity and to give a trustworthy outcome, professional practices in academia need to be updated timely. Henceforth, being a professional, one should possess technical, practical, and/or theoretical competencies to detect cheating behaviours of the students. Other than that, professionals need to be aware of the behavioural patterns that are considered prohibited, and it is better if they can be given training to handle current digital applications to identify cheating. Changing the assessment methods and moving to more authentic assessment types will minimise the issue satisfactorily.

The rise of AI models like ChatGPT has opened another path for contract cheating as it can potentially be used in generating academic content quickly and easily. AI models like ChatGPT are capable of producing human-like texts, and educators and the traditional plagiarism detection tool will find it difficult to differentiate the outcome (Mohammadkarimi, 2023; Hassoulas et al., 2023; Chaka, 2023). Further, students can use the AI-generated text as a guide to modify their answers to appear more original, making them harder to detect. Hence it is essential to establish guidelines and policies regarding the use of AI models for academic purposes (Rudolph et al., 2023a, 2023b). On the other hand, AI technologies can be used to develop advanced detection tools that can better identify instances of contract cheating.

As alternatives for minimising contract cheating, changes in the evaluation model, conveyance, and continuous guided inspection and support, improved institutional resourcing, evidence-based developments in curriculum and pedagogy to foster effective learning and skill development, and working with students in a partnership frequently emerge as main concerns of the educational contexts (Lancaster, 2020). In that lens, students' tendency to hire outsiders to complete their assignments or any other work is reasonable to think of as a cause of inadequate or unsatisfactory teaching, resourcing, or defects in pedagogical practices. Bretag et al. (2019b) confirm the argument and state contract cheating is partially affected by assessment, and therefore, proper course planning, resourcing, and evaluation should be done while adjusting students' perspectives, subjective norms, or their expected personal principles or temperaments (Curtis et al., 2018) to reduce contract cheating intentions. As well, understanding the relationship between the instructional settings and students' fraudulent behaviour is equally important.

Reported literature shows that although technology performs a main function in finding academic cheating, there are certain limitations and sometimes failures in detecting cheating. The use of video summarisation or video abstraction utilises artificial intelligence methods, web video recordings, live online proctoring, or web video conference invigilation (Holden et al., 2021) among the suggested methods of detection.

The experience of professionals about the students and the cheating identification through language, structure, and content is believed to be successful to some extent, yet personal biases and interests might be influential in the decision. Hence, the need for effective improvement in technology-based detection methods or systems is urgent and important. Moreover, it is the common responsibility of officials and academics to clearly define what is meant by academic fraudulence and what behaviours are classified as educationally dishonest in order to guarantee academic trustworthiness and prevent students from contracting cheating in tertiary education.

Furthermore, governments and academic institutions have also adopted a range of legal actions and policy decisions to divert students from contract cheating. Higher education is the main responsible authority to act against contract cheating. However, the lack of commitment from institutions to safeguard or practice those rules and regulations creates ample opportunities for students to follow the wrong path. Therefore, institutional involvement, including authorities, students, and the academic staff, is needed to build up an academically honest culture with both awareness and practice. Apart from that, students need to be made aware of the consequences of being dishonest in their academic journey and how it will affect their future lives. As a whole, contract cheating affects not only one's academic results but also the status of the institutions, educational standards, qualifications, professional conduct, and the safety and security of the general public.

Recommendations

Grounded on the outcomes of the study, recommendations related to contract cheating and academic integrity in higher education are discussed within this section. In that regard, several suggestions to address the issue have been made by the respective authorities, including scholars and educational quality assurance organisations (Lancaster & Clarke, 2016; QAA, 2020). At the very basic level, increasing the attentiveness toward contract cheating among the academic staff and newcomers is of considerable importance since both parties equally experience its unfavourable consequences. Hence, to ensure academic honesty in higher education, respective institutions must clearly define what behaviours are considered academically dishonest and need to convey them to students. Informing learners about the significance of keeping academic integrity at the inauguration and making them practice ethical values, directing students to establish a positive focus on facing academic challenges, supporting students to establish their own strategies in studying, encouraging students to enable their skills in academic writing, using academic resources and researching, emphasising learning goals, and developing their self-control would hopefully keep students away from essay mills and other kinds of commercially available services.

Equally, keeping the academic staff up-to-date with the newest trends in contract cheating and conducting staff professional development programmes, including required training to handle detected unethical conduct, setting effective academic regulations, and fair and transparent practice would be helpful in solving the problem. At the same time, it is important to establish measures to discourage or reduce students' involvement in contract cheating since the continuous practice may normalise unethical behaviours and demotivate students' tendency to achieve their goals with their true potential. Moreover, tertiary education institutions are responsible for establishing a culture of academic integrity. The commitment to safeguarding academic integrity can be conveyed through institutional standards and ethics, policy statements or mission declarations, or even through the student prospectus.

Addressing cheating strategically would be the best way to manage contract cheating. Worldwide, governments have initiated legislation against illegal services that provide contract cheating opportunities (Awdry & Newton, 2019). Further, the international network of contract cheating facilities is evolving rapidly, crossing borders. Therefore, it seemingly requires international collaboration to set up international standards and laws for contract cheating.

Research on contract cheating has been expanded across different dimensions but significantly centred on a few countries, like Australia, the United Kingdom, and Canada. Hence, uncovering the situation of other countries that remain under-researched is important. In addition, motives for students' engagement in contract cheating have been extensively studied, and the causes for not being involved in cheating are yet to be studied. In conclusion, it is obvious that to prevent students from contract cheating and to preserve academic integrity, there is no single reliable solution; instead, we need to go for globally accepted integrated approaches.

This work has certain limitations, which should be acknowledged. For instance, this study's coverage of articles may be constrained by the search terms and electronic databases used. As far as future research is concerned, it can be focused on the impact of novel applications such as ChatGPT on academic cheating behaviours. In addition, designing and assessing educational interventions aimed at preventing contract cheating would be better than imposing laws and punishment. We recommend investigating more on what are the perceptions and motivations of students to engage in cheating and how they can be addressed. Investigating the long-term effects of contract cheating on students' learning outcomes and studying how contract cheating impacts graduates' preparedness for the workforce will reveal the gaps that policymakers need to focus on in future to maintain and safeguard academic quality and standards.

References

Ahsan, K., Akbar, S., & Kam, B. (2021). Contract cheating in higher education: A systematic literature review and future research agenda. *Assessment & Evaluation in Higher Education, 47*(4), 523-539. https://doi.org/10.1080/02602938.2021.1931660

Ahmadi, A. (2014). Plagiarism in the academic context: A study of Iranian EFL learners. *Research Ethics, 10*(3) 151–168.

Amigud, A. (2019). Cheaters on Twitter: An analysis of engagement approaches of contract cheating services. *Studies in Higher Education, 45*(3), 692-705. https://doi.org/10.1080/03075079.2018.1564258

Amigud, A., & Dawson, P. (2019). The law and the outlaw: Is legal prohibition a viable solution to the contract cheating problem?. *Assessment & Evaluation in Higher Education 45*(1), 1-11. DOI: 10.1080/02602938.2019.1612851

Amigud, A., & Lancaster, T. (2019). 246 reasons to cheat: An analysis of students' reasons for seeking to outsource academic work. *Computers & Education, 34*, 98–107. https://doi.org/10.1016/j.compedu.2019.01.017

Armond, A., & Varga, O. (2021). *An alarming facet of ghostwriting: The contract cheating*. https://h2020integrity.eu/an-alarming-facet-of-ghostwriting-the-contract-cheating/

Awdry, R., & Newton, P. M. (2019). Staff views on commercial contract cheating in higher education: A survey study in Australia and the UK. *Higher Education, 78*, 593–610. https://doi.org/10.1007/s10734-019-00360-0

Bautista, R. M., & Pentang, J. T. (2022). Ctrl C + Ctrl V: Plagiarism and knowledge on referencing and citation among pre-service teachers. *International Journal of Multidisciplinary: Applied Business and Education Research, 3*(2), 245–257. https://doi.org/10.11594/ijmaber.03.02.10

Beckman, T., Lam, H., & Khare, A. (2017). Learning assessment must change in a world of digital "cheats". *Phantom Ex Machina: Digital Disruption's Role in Business Model Transformation*, 211-222.

Bleeker, K. (2008). *To be honest: Championing academic integrity in community colleges.* Community College.

Blum, S. D. (2016). "What it means to be a student today." In T. Bretag (Ed.), *Handbook of academic integrity* (pp. 383–406). Springer Singapore. doi: 10.1007/978-981-287-098-857

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., ... & van Haeringen, K. (2018). Contract cheating: A survey of Australian university students. *Studies in Higher Education, 44*(11), 1837-1856.

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., van Haeringen, K., ... & Rozenberg, P. (2019a). Contract cheating and assessment design: Exploring the relationship. *Assessment & Evaluation in Higher Education, 44*(5), 676-691.

Bretag, T., Harper, R., Ellis, C., Van Haeringen, K., Newton, P., Rozenberg, P., & Saddiqui, S. (2019b). *Contract cheating and assessment design: Exploring the connection.* Australia. Department of Education and Training (DET).

Brimble, M. (2016). "Why students cheat: An exploration of the motivators of student academic dishonesty in higher education". In R. Bretag (Ed.), *Handbook of academic integrity* (Chapter 26). Springer Press.

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching, 6*(2), 1-11. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.12

Comas-Forgas, R., Lancaster, T., Calvo-Sastre, A., & Sureda-Negre, J. (2021). Exam cheating and academic integrity breaches during the COVID-19 pandemic: An analysis of internet search activity in Spain. *Heliyon, 7*(10), e08233.

Curtis, G. J., Cowcher, E., Greene, B. R., Rundle, K., Paull, M., & Davis, M. C. (2018). Self-control, injunctive norms, and descriptive norms predict engagement in plagiarism in a theory of planned behavior model. *Journal of Academic Ethics, 16*, 225-239.

Curtis, G. J., & Vardanega, L. (2016). Is plagiarism changing over time? A 10-year time-lag study with three points of measurement. *Higher Education Research & Development, 35*(6), 1167-1179.

Dawson, P., Sutherland-Smith, W., & Ricksen, M. (2020). Can software improve marker accuracy at detecting contract cheating? A pilot study of the Turnitin authorship investigate alpha. *Assessment & Evaluation in Higher Education, 45*(4), 473-482.

Draper, M. J., & Newton, P. M. (2017). A legal approach to tackling contract cheating?. *International Journal for Educational Integrity, 13*(1), 1-16.

Eaton, S. E. (2020). Academic integrity during COVID-19: Reflections from the university of Calgary. *International Studies in Educational Administration, 48*(1), 80-85.

Eaton, S. E., & Turner, K. L. (2020). Exploring academic integrity and mental health during COVID-19: Rapid review. *Journal of Contemporary Education Theory & Research, 4*(2), 35-41. https://zenodo.org/record/4256825#.X6gXr5NKhm8. http://hdl.handle.net/1880/112729

Eaton, S. E. (2022). *Contract cheating in Canada: A comprehensive overview. Academic integrity in Canada: An enduring and essential challenge* (pp. 165-187). Springer.

Ellis, C., Van Haeringen, K., Harper, R., Bretag, T., Zucker, I., McBride, S., ... & Saddiqui, S. (2020). Does authentic assessment assure academic integrity? Evidence from contract cheating data. *Higher Education Research & Development, 39*(3), 454-469.

Erguvan, I. D. (2021). The rise of contract cheating during the COVID-19 pandemic: A qualitative study through the eyes of academics in Kuwait. *Language Testing in Asia, 11*, 34. https://doi.org/10.1186/s40468-021-00149-y

Eshet, Y. (2022). Contract cheating in Israel during the Covid-19 pandemic. *European Conference on Academic Integrity and Plagiarism 2022 (ECAIP 2022).*

Fishman, T. (2014). *The fundamental values of academic integrity. International Center for Academic Integrity* (2nd ed.). Clemson University.

Goff, D., Johnston, J., & Bouboulis, B. S. (2020). Maintaining academic standards and integrity in online business courses. *International Journal of Higher Education, 9*(2), 248-257.

Guerrero-Dib, J. G., Portales, L., & Heredia-Escorza, Y. (2020). Impact of academic integrity on workplace ethical behaviour. *International Journal for Educational Integrity, 16*(1), 1-18.

Harper, R., Bretag, T., Ellis, C., Newton, P., Rozenberg, P., Saddiqui, S., & van Haeringen, K. (2019). Contract cheating: A survey of Australian university staff. *Studies in Higher Education, 44*(11), 1857-1873.

Harper, R., Bretag, T., & Rundle, K. (2021). Detecting contract cheating: Examining the role of assessment type. *Higher Education Research & Development, 40*(2), 263-278.

Hassoulas, A., Powell, N., Roberts, L., Umla-Runge, K., Gray, L., & Coffey, M. (2023). Investigating marker accuracy in differentiating between university scripts written by students and those produced using ChatGPT. *Journal of Applied Learning & Teaching, 6*(2), 1-7. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.13

Hill, G., Mason, J., & Dunn, A. (2021). Contract cheating: an increasing challenge for global academic community arising from COVID-19. *Research and Practice in Technology Enhanced Learning, 16*, 1-20. https://doi.org/10.1186/s41039-021-00166-8.

Holden, O. L., Norris, M. E., & Kuhlmeier, V. A. (2021, July). Academic integrity in online assessment: A research review. In *Frontiers in education* (Vol. 6, p. 639814). Frontiers Media.

Ali, H. I. H., & Alhassan, A. (2021). Fighting contract cheating and ghostwriting in higher education: Moving towards a multidimensional approach. *Cogent Education, 8*(1), 1885837. https://doi.org/10.1080/2331186X.2021.1885837

Husain, F. M., Al-Shaibani, G. K. S., & Mahfoodh, O. H. A. (2017). Perceptions of and attitudes toward plagiarism and factors contributing to plagiarism: A review of studies. *Journal of Academic Ethics, 15*, 167-195.

Ison, D. C. (2020). Detection of online contract cheating through stylometry: A pilot study. *Online Learning, 24*(2), 165–142.

Jurkevičius, V., & Tauginienė, L. (2017). Ethical and legal observations on contract cheating services as an agreement. *International Journal for Educational Integrity 13*(1), 1-10.

Lancaster, T., & Clarke, R. (2007). Assessing contract cheating through auction sites – a computing perspective. In *Proceedings of 8th annual conference for information and computer sciences*. Higher Education Academy.

Lancaster, T., & Clarke, R. (2014). An observational analysis of the range and extent of contract cheating from online courses found on agency websites. In *Proceedings of 2014 eighth international conference on Complex, Intelligent and Software Intensive Systems (CISIS)* (pp. 56–63). IEEE. https://doi.org/10.1109/CISIS.2014.

Lancaster, T., & Clarke, R. (2016). Contract cheating: The outsourcing of assessed student work. In T. Bretag (Ed.), *Handbook of academic integrity*. Springer, Singapore. https://doi.org/10.1007/978-981-287-098-8_17

Lancaster, T. (2018). *Educational blackmail in the world of fake degrees, essay mills and contract cheating*. https://www.linkedin.com/pulse/educational-blackmail-world-fake-degrees-essay-mills-thomas-lancaster.

Lancaster, T. (2019). Profiling the international academic ghostwriters who are providing low-cost essays and assignments for the contract cheating industry. *Journal of Information, Communication and Ethics in Society, 17*(1), 72–86. https://doi.org/10.1108/JICES-04-2018-0040

Lancaster, T. (2020). Academic discipline integration by contract cheating services and essay Mills. *Journal of Academic Ethics, 18*(2), 115-127. https://doi.org/10.1007/s10805-019-09357-x

Lancaster, T., & Cotarlan, C. (2021). Contract cheating by STEM students through a file sharing website: A COVID-19 pandemic perspective. *International Journal for Educational Integrity, 17*(3), 1-16. https://doi.org/10.1007/s40979-021-00070-0

Lines, L. (2016). Ghost-writers guaranteeing grades? The quality of online ghost-writing services available to tertiary students in Australia. *Teaching in Higher Education, 21*(8),889-914. DOI: 10.1080/13562517.2016.1198759

Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *Journal of Applied Learning and Teaching, 6*(2), 1-9. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.10

Morris, E. J. (2018). Academic integrity matters: five considerations for addressing contract cheating. *International Journal for Educational Integrity, 14*(1), 15.

Mok, K. H., Xiong, W., & Bin Aedy Rahman, H. N. (2021). COVID-19 pandemic's disruption on university teaching and learning and competence cultivation: Student evaluation of online learning experiences in Hong Kong. *International Journal of Chinese Education, 10*(1), 221258682110070. https://doi.org/10.1177/22125868211007011

Orosz, G., Tóth-Király, I., Bőthe, B., Paskuj, B., Berkics, M., Fülöp, M., & Roland-Lévy, C. (2018). Linking cheating in school and corruption. *European Review of Applied Psychology, 68*(2), 89-97. https://doi.org/10.1016/j.erap.2018.02.001

Reddin, E. (2021). *A spike in cheating since the move to remote?* https://www.insidehighered.com/news/2021/02/05/study-finds-nearly-200-percent-jump-questions-submitted-chegg-after-start-pandemic

Rigby, D., Burton, M., Balcombe, K., Bateman, I., & Mulatu, A. (2015). Contract cheating and the market in essays. *Journal of Economic Behavior & Organization, 111*, 23-37. doi: 10.1016/j.jebo.2014.12.019.

Rogerson, A. (2017). *Detecting contract cheating in essay and report submissions: Process, patterns, clues and conversations*. https://ro.uow.edu.au/cgi/viewcontenttttpsredir=1&article=2344&context=buspapers

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Rundle, K., Curtis, G. J., & Clare, J. (2019). Why students do not engage in contract cheating. *Frontiers in Psychology, 10,* 2229.

Sharma, P., & Maleyeff, J. (2003). Internet education: Potential problems and solutions. *International Journal of Educational Management, 17,* 19–25. https://doi.org/10.1108/09513540310456365

Singh, S., & Remenyi, D. (2016). Plagiarism and ghostwriting: The rise in academic misconduct. *South African Journal of Science, 112*(5-6), 1-7.

Slade, C., Rowland, S., & McGrath, D. (2019). Talking about contract cheating: Facilitating a forum for collaborative development of assessment practices to combat student dishonesty. *International Journal for Academic Development, 24*(1), 21-34.

Tauginienė, L., & Jurkevičius, V. (2017). Ethical and legal observations on contract cheating services as an agreement. *International Journal for Educational Integrity, 13*(9), 1-10. DOI 10.1007/s40979-017-0020-7

The Quality Assurance Agency for Higher Education (QAA). (2020). *Contracting to cheat in higher education.* https://www.qaa.ac.uk/docs/qaa/guidance/contracting-to-cheat-in-higher-education-2nd-edition.pdf.

Tran, T. K., Dinh, H., Nguyen, H., Le, D. N., Nguyen, D. K., Tran, A. C., ... & Nguyen, T. A. (2021). The impact of the COVID-19 pandemic on college students: An online survey. *Sustainability, 13*(19), 10762. https://doi.org/10.1007/s10805-012-9150-y

Trezise, K., Ryan, T., de Barba, P., & Kennedy, G. (2019). Detecting contract cheating using learning analytics. *Journal of Learning Analytics, 6*(3), 90-104. http://dx.doi.org/10.18608/jla.2019.63.11

Wang, Y., & Xu, Z. (2021). Statistical analysis for contract cheating in Chinese universities. *Mathematics, 9*(14), 1684. https://doi.org/10.3390/math9141684

Wallace, M. J., & Newton, P. M. (2014) Turnaround time and market capacity in contract cheating. *Educational Studies, 40*(2), 233–236, https://doi.org/10.1080/03055698.2014.889597

Williamson, M. (2019). Contract cheating and academic integrity in higher education: What can universities, governments and quality assurance agencies do to understand, prevent and respond to the challenge?. *Kuwait International Law School Journal, Special Supplement, 4*(1), 219-274.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Preservice secondary teachers' beliefs about academic dishonesty: An attribution theory lens to causal search

Lauren D. Goegan[A]

Sierra L. P. Tulloch[B]

Lia M. Daniels[C]

A Assistant Professor, Department of Educational Administration, Foundations and Psychology University of Manitoba, Canada

B Ph.D. student, School and Clinical Child Psychology program, University of Alberta, Canada

C Professor of Educational Psychology, University of Alberta, Canada

Keywords

Academic dishonesty;
attribution theory;
causal search;
cheating;
preservice teachers.

Correspondence

lauren.goegan@umanitoba.ca [A]

Article Info

Received 15 May 2023
Received in revised form 11 July 2023
Accepted 12 July 2023
Available online 13 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.6

Abstract

Academic dishonesty is an area of concern across all levels of education. While previous research has largely focused on what behaviours students engage in and what instructors do in response, little is known about why, and even less incorporates a theoretical framework. To contribute to the existing literature, our aim was to examine preservice secondary teachers' beliefs about academic dishonesty. Moreover, we utilized Attribution Theory as our theoretical framework and examined how preservice teachers engage in causal search when presented with instances of academic dishonesty. Our results demonstrate that preservice teachers have strong beliefs about what is and what is not academic dishonesty; however, context matters. Indeed, when provided with descriptive scenarios compared to discrete behaviours, ratings of academic dishonesty were significantly higher in the former than the latter. Moreover, preservice teachers draw on multiple pieces of information when engaging in the causal search process, identifying not only facts but also embellishments not present in the scenario and highlighting their beliefs around academic dishonesty. Recommendations for educators and administrators for supporting students are provided, as well as limitations and directions for future research.

Introduction

More than 60% of students at university openly admit to cheating (International Academy for Academic Integrity, 2020). Similarly, K-12 teachers are frustrated and concerned about rampant cheating that can near 70% of students (Hasson, 2017). Although informative, these descriptions of prevalence do not capture the complex psychosocial factors at play in both the people who engage in academic dishonesty and the people who detect it and enforce sanctions. Such complexity is made worse as the concept of academic dishonesty itself seems to be changing with the increased use of artificial intelligence (Peritz, 2022; Perkins et al., 2023). During this time of change, one population that may be particularly important to consider regarding academic dishonesty is preservice teachers. As current students, preservice teachers are aware of the increased opportunities to cheat. As future teachers, they represent gatekeepers of academic integrity as they become responsible for educating young people (Fontaine et al., 2020; Romanowski, 2021). Noticing their important role, research examining preservice teachers regarding academic dishonesty has increased in recent years (Bautista & Pentang, 2022; DiPaulo, 2022; Fontaine et al., 2020; Merkle, 2021; Romanowski, 2021), but it is largely descriptive and ignores psychosocial elements of dishonesty. Therefore, the purpose of this paper is to draw on the interpersonal psychosocial elements of Attribution Theory (Weiner, 1985; 2010) to examine preservice secondary teachers' perspectives on, and responses to, hypothetical instances of academic dishonesty.

What is academic dishonesty?

Some scholars define academic dishonesty quite broadly as any act of deception or misrepresentation that violates the fundamental principles of academic integrity (McCabe et al., 2012). While some scholars argue that there are general principles that define academic dishonesty, such as intentional acts of fraud (e.g., submitting someone else's work as your own), others suggest that there is no single definition that encompasses all forms of unethical behaviour in academic settings (Aaron et al., 2011). However, most researchers agree that academic dishonesty should be expansive enough to encompass various behaviours and the contexts in which they occur. The most common behaviours in secondary and post-secondary education usually involve plagiarism, completing individual work with other students, cheating, fabrication, and falsification (Christensen Hughes & McCabe, 2006; Şendağ et al., 2012). The consequences of these actions not only undermine the integrity of the educational system, they raise questions about the qualifications of the individuals engaging in this behaviour (Chibry & Kurz, 2022). For preservice teachers specifically, the impact these actions can have on their roles as future teachers who are integral to the ethical development of students for years to come is important to consider. Perhaps not surprisingly, educational institutions have a renewed sense of urgency in this area with the rapid expansion of Artificial Intelligence, such as ChatGPT (Cotton et al., 2023) and "pay-to-pass" websites (Chibry & Kurz, 2022, p. 203), making it more important than ever to consider how psychosocial theories can help explain academic dishonesty.

Attribution theory

Attribution theory helps explain how individuals understand the causes of their own or others' behaviour through a process known as causal search. Much like it sounds, causal search is the active process of trying to identify the causes of a behaviour. It often involves observing the behaviour and making inferences about the underlying causes. According to attribution theory, individuals typically engage in causal search when outcomes occur that are negative, unexpected, and important (Graham, 1991; Weiner, 1985; 2000; 2010) and the identified "cause," also referred to as causal ascription, in turn, leads to predictable cognitions, emotions, and behaviours. Most teachers would agree that discovering cheating would classify as negative, unexpected, and important, thereby triggering a causal search. During a causal search, people look for information from the current situation, past experiences, personal knowledge and beliefs about the individual, and anything else that may seem relevant.

Teachers may also consider the causal dimensions when evaluating a behaviour as academically dishonest or not. Indeed, according to Weiner (1985), while there are an infinite number of causes for a behaviour, these causes can all be classified according to the dimensions of locus, stability, and controllability. Locus refers to whether the cause of the outcome is internal or external to the individual. Stability refers to how stable or unstable over time the cause is perceived to be. Controllability refers to whether the individual is in control of the cause. Based on these causal dimensions, predictable psychological and behavioural consequences follow (Weiner, 1985; 2018). For example, if a teacher finds evidence of cheating and attributes the cause to the student being lazy, this would be seen as stable, controllable, and internal, and as a result, the teacher would be more likely to feel angry towards the student and offer punishment. Alternatively, if the behaviour was attributed to the student not being taught the rules yet, this could be considered unstable, uncontrollable, and external, wherein the teacher would be more likely to feel sympathetic and offer help. As such, causal search is a critical first step that impacts how behaviours are interpreted and responded to.

The research on academic dishonesty from an attribution theory lens

To date, we found only one study that examined academic dishonesty from an attribution theory lens in terms of preservice teachers. The authors examined how beliefs of controllability related to acts of plagiarism impacted preservice teachers' views on responsibility, emotions, help-giving, and reporting (Goegan & Daniels, 2023). They determined that when scenarios described students who engaged in plagiarism that was controllable, the preservice teachers were most likely to view that student as responsible, feel anger towards them, support student punishment and recommend that the student be reported, compared to acts of plagiarism that were uncontrollable. In other words, the tenets of attribution theory were correct.

Alternatively, researchers have sought to understand academic dishonesty using attribution theory as a framework more broadly. Most of these researchers used attributions to explain students' cheating behaviours, despite knowing or feeling that it is "wrong" (Murdock & Stephens, 2007; Stephens, 2017). Both secondary and postsecondary students often see their academic dishonesty being caused by external or uncontrollable factors, such as pressure from parents to receive good grades or insufficient studying support from others (Murdock & Stephens, 2007). Students who also attribute their or their peers' behaviours to external factors rather than internal factors may be less likely to view their academic dishonesty as a serious violation (Murdock & Stephens, 2007; Stephens, 2017). Seals and colleagues (2014) used attribution theory to provide insight as to why university teaching assistants might consider academic dishonesty to be common in university, but not in their courses.

Research from a K-12 perspective on academic dishonesty rarely incorporates attribution theory. Nevertheless, important links between the findings of previous research and theory can be inferred, particularly in secondary school, where grades have increasing consequences for students (reference). For example, Geddes (2011) found that among high-achieving high school students, the academic reason with the highest agreement among students for cheating was securing a high GPA, while the highest non-academic reason was pressure from parents. While pressure from parents would be considered uncontrollable, the need for a high GPA could be interpreted as either controllable or uncontrollable. Moreover, Galloway (2012) conducted interviews with high-achieving high school students about reasons for cheating, which included feeling forced to cheat, and an academic culture that valued achievement over learning.

How causal search influences dishonesty decisions

Across all levels of schooling, teachers play a crucial role in managing academic integrity among their students and enforcing policy when integrity is compromised. However, policy decisions can also be contingent on the student's intentionality and previous conduct (Amigud & Pell, 2021). Careful consideration of the level of the course, the type of assignment, and the institution or school board policies and procedures also come into play. Regardless of policies, interestingly, it seems that post-secondary faculty members rarely report cases of academic misconduct and instead attempt to resolve cases based on their own judgements (Kwong et al., 2010; Thomas, 2017). One reason for a preference for personal/professional judgment rather than strict policy adherence is that instructors may view plagiarism as a changing concept that requires judgement (Fyfe, 2022). For example, instead of banning all use of AI, Otsuki (2020) suggests training writers how to work with text-generated AI. Nevertheless, there is little empirical evidence on teachers' or preservice secondary teachers' beliefs or decision-making process when it comes to students in K-12 school settings. Regardless of preferences, the decision-making process can be complex, and it is crucial to maintain academic integrity, fairness, and consistency in the academic environment (Gottardello & Karabag, 2022). To our knowledge, no studies have been conducted using attribution theory to examine preservice secondary teachers' causal search when encountering students' dishonest behaviours.

What forms of discipline do instructors and teachers recommend?

Štambuk et al. (2015) found that teachers across elementary, secondary and university levels react fairly similarly to acts of cheating. Faculty members suggest various consequences for academic dishonesty (Pincus & Schmelkin, 2003), but typically agree that consequences should be proportional to the severity of the offence and should educate students on the importance of academic integrity (Keener et al., 2019). Some suggested consequences include verbal warnings, grade reductions, re-submission of assignments, and suspension or expulsion (Keener et al., 2019). Again, there is no empirical work highlighting teachers' or preservice teachers' perceptions and recommendations for academic dishonesty from an attribution theory lens in primary and secondary education settings. Together, these findings hint at the larger need for consideration of what teachers could or should do when faced with challenging circumstances of academic dishonesty.

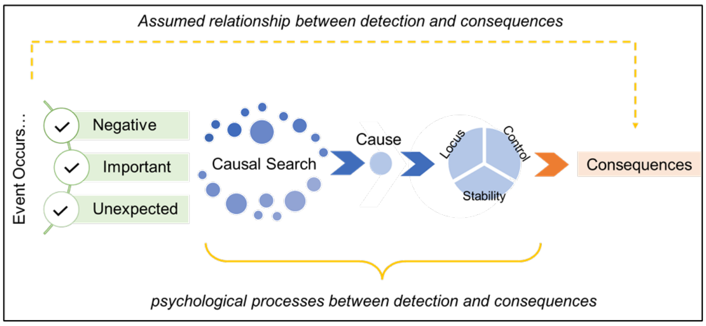


Figure 1. Conceptual model for the present study.

The current study

The purpose of this study was to examine the type of information preservice teachers draw on when considering situations of potential academic dishonesty. Our research questions were: (1) How do preservice teachers rate behaviours as academically dishonest? (2) Does the extent a behaviour is considered dishonest differ when contextual information is presented? (3) What type of information do preservice teachers use when determining if behaviours are dishonest? (4) What forms of consequences do preservice teachers recommend for instances of academic dishonesty?

Method

This correlational study involved two components administered on a single survey embedded in a required education course. Ethics approval was obtained from the Human Ethics Research Office at the researchers' university.

Procedures

Participants were preservice secondary school teachers enrolled in an assessment course that was part of their teacher education program at a mid-size university in Western Canada. The Fall 2021 offering of the course was in an asynchronous format and consisted of several units, one of which specifically addressed academic success and dishonesty in the classroom. This unit required preservice teachers to complete a series of activities online. The specific items within these activities related to this research project are provided below. Once a preservice teacher had completed the activities, they were prompted with the question, "Can we include your responses here for research purposes?" Consent was granted by answering yes. Data were anonymized and analyzed after the completion of the course.

Participants

There were 210 preservice secondary teachers enrolled in the assessment course where potential participants completed the activities included in this study. In total, 166 individuals indicated that "yes" we could use their responses for research purposes. Participants ranged in age from 20 to 48 (M = 24.80) and predominately identified as white (81%). When asked: "*How do you want us to describe your gender?*", 49% of participants identified as women, 46% as men and 5% identified as non-binary or preferred not to disclose. These percentages are consistent with international numbers that find women in the teacher profession average 47% of the total population (OECD, 2019). Data on race could not be located.

Measures

Academic dishonesty discrete behaviours

We asked participants to indicate the extent to which 21 discrete behaviours reflected academic dishonesty. The behaviours included common forms of academic dishonesty identified in the literature, such as submitting someone else's work as your own and peeking at answers during an exam, as well as less obvious examples, such as collaborating on individual work or omitting references. As an attention check, we also included more innocuous behaviours such as studying from available old exams and forming a study group which are generally not considered dishonest. Participants were presented with the stem "To what extent do you consider the following activities as forms of academic dishonesty," and indicated their agreement on a scale from 1 (not at all) to 7 (very much so). A full list of behaviours can be found in Table 1 with means, standard deviations, skew, and kurtosis.

Academic dishonesty expanded scenarios

We expanded six of the discrete behaviours into more elaborate scenarios to examine how preservice teachers used contextual information in their consideration of academic dishonesty. These behaviours included exam situations, such as peeking at another student's answers, sneaking answers into an exam, and communicating answers to another student, as well as situations involving plagiarism, such as submitting someone else's work as their own and taking credit for ideas that are not their own. The scenarios also varied in the weight of the assignment involved in academic dishonesty and the student's acknowledgement of their behaviour. The scenarios were written by the first author and reviewed by the co-authors for clarity and ambiguity. Please see Appendix A for the exact wording of all six scenarios. For each scenario, participants first responded to the Likert scale item: "To what extent do you consider the student's behaviour as academic dishonesty" (1 = not at all to 7 = very much so). Then, participants provided open-ended responses to the following two questions designed to elicit causal search and identify what information in the scenarios influenced their rating: (a) What in the story helped you decide on your response? And (b) What do you think is an appropriate form of discipline (if any) and why?

Plan for analysis

We conducted our analyses in four stages. First, we ran descriptive analyses for the 21 items related to preservice teacher beliefs about behaviours as examples of academic dishonesty or not and ran one-sample t-tests on these items to determine if participants' responses differed from 4 (neutral). This allowed us to answer our first research question: How do preservice teachers rate behaviours as academically dishonest? We hypothesized that participants would have strong agreement with academically dishonest items and strong disagreement with the innocuous behaviours.

Second, we ran descriptive analyses for the Likert scale item associated with each scenario, including means, standard deviations, ranges, skewness, and kurtosis. Then, we used paired samples t-tests to compare mean scores on the discrete behaviours with mean scores for the expanded scenarios. This allowed us to test the extent to which contextual information shifted preservice teachers' convictions that the action was dishonest (Research Question 2). We did not have specific hypotheses for these t-tests as they were exploratory in nature, and no previous research could be located with comparable analyses to inform a hypothesis.

Third, we performed a content analysis (Hsieh & Shannon, 2018) to extract themes from participants' open-ended responses to the prompt, "What in the story helped you decide on your response?" The first and second authors met regularly via video conferencing and, beginning with the first scenario, highlighted meaning units that were then formed into themes across the scenarios from participants' open-ended responses. To aid in the consistency of coding, a codebook was created that contained a description of what each code covered, what was excluded from a code, and examples from the participants that met each criterion. This codebook was reviewed by all authors before incorporating it into the analysis of the remaining scenarios. Any disagreements in coding were discussed until consensus was achieved. Inter-rater reliability for scenarios 2 through 5 was calculated at 96.4%. This process allowed us to answer our third research question: What type of information do preservice teachers use when determining if behaviours are dishonest? Based on attribution theory, we anticipated that participants would identify facts from the scenarios and their own beliefs about academic dishonesty in their responses.

Fourth, we performed an additional inductive analysis on the second open-ended response to the prompt “What do you think is an appropriate form of discipline (if any) and why?” to answer our last research question: What forms of consequences do preservice teachers recommend for instances of academic dishonesty? Consistent with previous research, we hypothesized that participants would identify various forms of discipline across the scenarios (Keener et al., 2019; Pincus & Schmelkin, 2003).

Results

Academic dishonesty discrete behaviours

Descriptive information, including the means, standard deviations, skewness, and kurtosis for the 21 Likert items related to participants’ beliefs about behaviours as academically dishonest or not, are presented in Table 1. Assumptions about the data when performing t-tests were reviewed with one important note, the normality of the distribution. We hypothesized that preservice secondary teachers would have distinct views on these items and anticipated skewness in the data. Nevertheless, we highlight some important findings here. First, participants strongly agreed that most of the items were examples of academic dishonesty, with the top three items being (1) having someone else take your exam for you, (2) submitting someone else’s work as your own, and (3) buying a term paper or essay. Indeed, all of the responses demonstrated large effect sizes apart from the item “re-submitting your own work for a different class,” which only produced a medium effect. This may speak to less understanding of self-plagiarism (to be discussed below). Moreover, four items were endorsed as very strongly not examples of academic dishonesty, being (a) asking for feedback on a draft of an assignment, (b) studying from available old exams, (c) taking a practice exam and (d) forming a study group. Taken together, preservice teachers have very strong beliefs about what constitutes academic dishonesty and what does not.

Table 1: Means for student behaviour items in descending order.

Student Behaviour	Mean	SD	Skew	Kurtosis	*t*-value	Cohen's d
1. Having someone else take your exam for you.	6.95	0.40	-8.45	74.68	94.68***	7.35
2. Submitting someone else’s work as your own.	6.93	0.34	-5.87	40.17	110.77***	8.60
3. Buying a term paper or essay.	6.92	0.38	-5.46	31.66	98.74***	7.66
4. Sneaking answers into a closed book exam (e.g., on water bottles).	6.81	0.55	-3.84	18.48	65.82***	5.11
5. Having someone tell you the answers on an exam.	6.78	0.59	-3.02	9.44	61.01***	4.74
6. Obtaining test answers before the exam.	6.75	0.62	-2.92	8.84	57.46***	4.46
7. Peeking at someone’s answers during the exam.	6.66	0.70	-2.17	4.20	48.72***	3.78
8. Helping someone see an answer during the exam.	6.48	0.84	-1.65	2.34	38.14***	2.96
9. Taking credit for ideas that aren’t yours.	6.39	0.86	-1.42	1.63	35.84***	2.78
10. Looking up answers to an exam online.	6.36	1.19	-2.23	4.84	25.53***	1.98
11. Copying and pasting directly into your assignment.	6.15	1.08	-1.45	2.22	25.71***	2.00
12. Collaborating on a take-home or online exam without permission.	6.12	1.19	-1.61	2.82	23.06***	1.79
13. Asking students who have taken the exam for the questions.	5.87	1.43	-1.35	1.34	16.83***	1.31
14. Including information you know is inaccurate in an assignment.	5.78	1.30	-1.04	0.57	17.69***	1.37
15. Submitting an assignment without citing all resources used.	5.72	1.29	-0.99	0.54	17.18***	1.33
16. Lying to get an extension on a due date.	5.40	1.53	-0.87	0.00	11.81***	0.92
17. Re-submitting your own work for a different class.	4.69	1.77	-0.41	-0.76	5.05***	0.39
18. Asking for feedback on a draft of an assignment.	1.39	0.94	3.23	11.86	-35.78***	-2.78
19. Studying from available old exams.	1.27	0.85	4.25	20.76	-41.61***	-3.23
20. Taking a practice exam.	1.12	0.74	6.98	50.85	-50.35***	-3.91
21. Forming a study group.	1.01	0.08	12.88	166.00	-497.00***	-38.58

Note: * $p < .05$, ** $p < .01$, *** $p < .001$. For the one-sample t-test, the alternative hypothesis specifies that the mean is different from 4.

Academic dishonesty expanded scenarios

Quantitative analyses

Means, standard deviations, and ranges for each of the belief ratings associated with the six scenarios and the eight students involved (A-H) were calculated and are provided in Table 2. Overall, preservice teachers identified the student behaviours in the scenarios as academically dishonest (apart from Student H who is not the central character of Scenario 6). To answer our second research question, we paired participants’ belief ratings from the scenarios with the corresponding Likert-scale items from the 21 items for discrete behaviours and conducted six paired samples t-tests. We excluded scenario 6 from these analyses as, based on further examination of the scenario, it was unclear how the papers of the two students were similar (i.e., in terms of text generated or ideas) and therefore did not connect to the discrete student behaviour items as well as the other scenarios did. Overall, when participants were provided more details via the expanded scenarios, they were more lenient on their rating of academic dishonesty (Table 3).

Table 2: Descriptive Information for item “To what extent do you consider the student’s behaviour as academic dishonesty?” by Scenario.

Scenario	Student	Mean	SD	Min	Max	*t*-value	Cohen's d
1	Student A	5.28	1.36	2.00	7.00	12.15***	0.94
2	Student B	5.94	1.29	1.00	7.00	19.35***	1.50
3	Student C	6.03	1.05	2.00	7.00	24.85***	1.93
3	Student D	6.28	0.98	1.00	7.00	29.90***	2.32
4	Student E	6.43	0.94	3.00	7.00	33.18***	2.58
5	Student F	6.20	1.10	1.00	7.00	25.83***	2.01
6	Student G	5.65	1.36	1.00	7.00	15.61***	1.21
6	Student H	3.78	1.82	1.00	7.00	-1.58	-0.12

Note: * $p < .05$, ** $p < .01$, *** $p < .001$. For the one-sample t-test, the alternative hypothesis specifies that the mean is different from 4.

Table 3: Comparison of Likert scale items.

	Listed Behaviours		Expanded Scenarios			
Comparison	Behaviour	Mean	Student	Mean	*t*-value	Cohen's d
1	Peeking at someone’s answers during the exam.	6.66	Student A	5.28	13.79***	1.07
2	Sneaking answers into a closed book exam.	6.81	Student B	5.94	8.04***	.062
3	Having someone tell you the answers on an exam.	6.78	Student C	6.03	9.56***	.074
3	Having someone tell you the answers on an exam.	6.78	Student D	6.28	6.60***	.051
5	Submitting someone else’s work as your own.	6.93	Student E	6.43	7.16***	.056
6	Taking credit for ideas that aren’t yours.	6.39	Student F	6.20	2.14**	0.17

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Qualitative analyses

Based on the content analysis in response to the question, “What in the story helped you decide on your response?” we identified three themes: (a) facts, (b) embellishments and (c) beliefs. Facts consisted of the participants correctly identifying details within the scenario that they then utilized in their decision of whether the behaviour was an example of academic dishonesty or not. For example, in Scenario 1, participants mentioned: "the student was looking at another student's answer," "the student is squirmy" and "the student has pressure from their parents." Across the six scenarios, 72-83% of the participants identified facts that supported their decision-making about the behaviours.

Embellishments consisted of the participants adding details that were not provided in the scenario. For example, in the case of the student peeking at a peer's exam (Scenario 1), some of the participants inferred that the student is a "good kid" and had "never done this before," neither of which are mentioned in the scenario. Moreover, in the case of the student using a previous assignment of their siblings (Scenario 4), participants said both "the sibling didn't know," and "the sibling willingly gave their assignment to their sibling," neither of which is mentioned. Overall, across the six scenarios, 60-75% of participants identified embellishments that were not provided in the scenarios to make their decisions.

Lastly, the beliefs theme reflected how participants used their pre-existing points of view in determining the extent to which the scenario represented academic dishonesty. For example, in Scenario 2, where the teacher found a sheet of paper that had the answers to the test and suspected Student B, a participant said, "it is unfair to make assumptions; however, sometimes teachers must be academic dishonesty detectives," and "I would have put 7 because cheating in this way on a summative assessment worth 30% of your mark is unacceptable." Moreover, one preservice teacher wrote about the various factors they would weigh as they undertook a causal search while making a decision:

> I would also have to look at my own potential bias towards the student. Why is it that I suspected that the student was cheating? Are there other circumstances in which this paper could have ended up there, or if I am judging this student out of my own preferences?

Moreover, in Scenario 3 where one student communicates the answers to another during a test, the participants identified various beliefs such as "helping out a friend is a natural thing to want to do," "ultimately, cheating is cheating," and, "and to make it worse it was on a summative exam." Examples of facts, embellishments and beliefs for each scenario are provided in Appendix B.

Consequences for acts of academic dishonesty

Within and across scenarios, preservice teachers suggested a wide range of consequences some of which were quite mild such as giving a warning and others of which were highly punitive such as awarding zeros. We identified six common forms of discipline: warnings, re-testing, reducing grades, calling home, detention, and no punishment. Looking at these forms for the first scenario where the student is peeking at answers during a quiz, 83% of participants suggested giving a warning, 31% said re-assess the student, 5% said to change their grade (including giving a zero), 4% said to call home, 4% said detention, 60% said no punishment and 8% said something else (e.g., make the student write a paper why cheating is not acceptable or give the student a choice such as take a zero or redo the test). Note that the percentages add up to greater than 100% for each scenario as participants were able to include more than one form of discipline. A full breakdown of the punishments identify by scenarios is provided in Table 4.

Table 4: Consequences ascribed by the participants by scenario.

Categories	1	2	3	4	5	6
Warning / Conversation	83%	40%	49%	58%	64%	51%
Re-test / Re-assess	31%	45%	63%	70%	52%	73%
Reducing Grade (e.g., give zero, adjust grade, losing marks)	5%	40%	36%	16%	42%	42%
Call home / Talk to Parent	4%	20%	14%	17%	10%	7%
Detention / Suspension (e.g., stay at lunch / after school)	4%	5%	10%	8%	2%	4%
No Punishment	60%	18%	5%	14%	14%	7%
Other (e.g., choice of punishment, write a letter)	8%	10%	10%	8%	11%	10%

Discussion

We examined the beliefs of preservice teachers in terms of academic dishonesty, utilizing attribution theory as our conceptual model. Overall, the participants had strong beliefs in terms of what constituted academic dishonesty both in terms of discrete behaviours, and in response to the scenarios. We discuss the findings from each research question to consider how preservice teachers engage in the causal search process and determine consequences when presented with different behaviours that could be defined as academic dishonesty. In closing, we discuss the limitations and potential avenues for future research.

Teachers have strong beliefs

Based on our first research question, that is, how do preservice secondary teachers rate behaviours as academically dishonest, we found that participants rated behaviours strongly in terms of them being academically dishonest or not. Indeed, behaviours such as having someone else take your exam for you, submitting someone else's work as your own and buying a term paper or essay were all rated strongly as academic dishonesty, while behaviour such as studying from available old exams, taking a practice exam, and forming a study group were rated strongly as not instances of academic dishonesty. However, there was one exception: "resubmitting your own work for a different class." This finding highlights the importance of students understanding self-plagiarism, or what Cajigas and colleagues refer to as "text recycling" (2022, p. 1697). Self-plagiarism has received more attention in recent years (e.g., Rozhkova & Isaeva, 2022), and as such, more information is needed for preservice teachers in their training considering self-plagiarism and how to respond in their future classrooms.

Nevertheless, the extent to which a behaviour was rated as academically dishonest differed when contextual information was present. Indeed, based on our second research question, to what extent does contextual information shift preservice secondary teachers' conviction that an action is dishonest, we see across all paired samples t-tests ratings of academic dishonesty dropped when comparing the Likert scale behaviours and the scenarios. Said differently, context matters to preservice teachers. Based on attribution theory, there are an infinite number of causes that one can perceive when it comes to an outcome (Weiner, 1985), and providing preservice teachers with more details provides more room for speculation and interpretation. Importantly, the theory goes on to suggest that once a cause has been determined, there are only three underlying causal dimensions (locus, stability, and controllability, reviewed above). Research by by Goegan & Daniels (2023) suggests that in terms of plagiarism, when scenarios were deemed within the person's control, preservice teachers were more likely to suggest the student was responsible. Like our results, the largest difference in means was associated with scenario one, wherein the student apologized and said they had been under a lot of pressure to do well at school from their parents and that they had too many other assignments to do; therefore they did not have time to study, perhaps interpreted as uncontrollable and/or not responsible. Alternatively, taking credit for ideas that are not yours (scenario five) had the smallest difference in means which could be interpreted as controllable and responsible. More research is needed to further examine the differences in terms of student actions and teacher consequences for academic dishonesty. Moreover, incorporating scenarios of academic dishonesty into preservice teacher training could provide an important avenue for discussion around responsibility for academic dishonesty and associated consequences before preservice teachers enter the classroom and must make these decisions themselves.

Facts, embellishments, and beliefs

For our third research question, what types of information do preservice secondary teachers use when determining if behaviours are dishonest, we found that participants' causal search extended well beyond the stated facts of the scenarios to include embellishments and beliefs. This was a departure from our hypothesis that suggested the identification of facts and beliefs, but not embellishments. This reinforces the need to consider the psychosocial elements of dishonesty, such as social norms (Daumiller & Janke, 2020). Indeed, during the causal antecedents stage of the theory, there are many causal rules and biases that can impact an individual (Graham & Taylor, 2016; Pintrich & Schunk, 2002). Rudolph and Tscharaktschiew (2014) highlight the difference between the individual interpreting their own behaviour and an interpersonal perspective wherein the individual interprets the behaviour of others; our study focused on the latter. It would be advantageous to explore if the scenarios had been written in the first person and how that might have shifted the results. Indeed, the fundamental attribution error (Graham & Taylor, 2016) in terms of academic dishonesty would suggest that individuals are more likely to attribute their own behaviour to situational factors (e.g., did not know they were plagiarizing), while in the role of observer, are more likely to attribute the behaviour to personal characteristics (e.g., the student was lazy).

Indeed, similar to eyewitness testimony, there can be various biases beyond the fundamental attribution error. For example, Nayak and Khajuria (2019) identified several internal and external factors affecting the accuracy of eyewitness identification, including prejudice, prior experience, cognitive state, degree of certainty, and racial or personal bias, among others. Borrowing further from the eyewitness testimony research field is the idea of the misinformation effect. Here, "a person recollects that they experienced an event in a way that is consistent with false information provided to them after the event" (Puddifoot, 2020, pp. 255-256). In terms of academic dishonesty, it's not just about the biases preservice teachers hold before the behaviour occurs but the information gathered afterwards as well. This may speak to the importance of record-keeping when dealing with instances of academic dishonesty. This may also explain an embellishment in Scenario 2, where the student was accused of sneaking a sheet of paper that had the answers into a test. Based on the student "look[ing] concerned, almost guilty, but deny[ing] that the paper is theirs," many participants adjusted the fact that "the writing looks very similar" to "the writing was a match." Future research could break down the scenarios into discrete events that occurred after the behaviour was detected to determine if ratings or interpretations change over time.

Consequences

For our fourth research question, what forms of consequences do preservice secondary teachers recommend for instances of academic dishonesty, we found that participants suggested a variety of consequences within and across scenarios (see Table 4). On the one hand, this shows consistency across scenarios and suggests these consequences are indeed common. On the other hand, this shows little consistency within scenarios suggesting there is rarely a singularity to consequences for a specific action. Consistency was reduced even more when the scenarios involved more than one student. For example, in Scenario 3 Student C was signaling answers to student D during an exam. Not only do some of the participants interpret the students' actions differently, for example, one individual said, "Student C was only trying to help a friend, Student D was cheating", but also in the severity of the behaviour, as stated by one participant "while student C is not 'cheating' on their exam, they are helping student D cheat, student D is more in the wrong than Student C." The comments by participants may highlight an important element of the definition of academic dishonesty, and that is that the behaviour provides an unfair advantage for the student committing the dishonesty over other students (Hylton et al., 2016). Indeed, another participant said, "Student C knew the material and was wishing to help their pal through [a] tough time. Furthermore, they themselves knew and understood the content. In contrast, Student D was taking the answers and did not know the content." The idea of benefiting from the action is a distinction that preservice teachers made, and it impacted not only the rating of the behaviours but also the consequences. As a result, many of the participants identified more punitive consequences for Student D than C. This was also present in the participants' beliefs that academic dishonesty was less severe in formative rather than summative assessment, presumably because the former is not graded. Therefore, future research should continue to investigate the impact of benefiting from academic dishonesty in relation to how the behaviour is interpreted and the resultant consequences for the individual involved.

Limitations and future directions

The results here need to be interpreted with consideration of three limitations. First, the lessons that the preservice secondary teachers engaged in prior to completing the survey did not include knowledge concerning the policies for how to handle academic dishonesty. This was not included because the preservice teachers would be eventually teaching in various schools across the province, which could have different guidelines. On the one hand, providing preservice teachers with some guidelines to follow may have reduced the range of consequences identified here. On the other hand, previous research has found that instructors rely more on their personal or professional judgements rather than adhering to policy (Kwong et al., 2010; Thomas, 2017). As such, future research could further extend our findings here to investigate how decisions about consequences specifically are determined. Indeed, Keener and colleagues (2019) suggest that faculty members believe that consequences for academic dishonesty should be proportional to the severity of the offence, but the severity of the offence may be subjective. For example, several of the preservice teachers here had the belief that academic dishonesty was more severe when the assignment was summative rather than formative, while others said cheating is cheating. As such, this is an important area for future research.

Second, we asked the participants, "*To what extent do you consider the student's behaviour as academic dishonesty?*" and then "*What in the story helped you decide on your response?*" Based on attribution theory (Weiner, 1985; 2000; 2010), these two questions perhaps should be reversed. In considering what in the story helped them decide, the preservice teachers were engaging in causal search, that is, trying to determine why the student engaged in the act of academic dishonesty. While rating the behaviour would be after causal ascription, where the reason why has been identified, and the preservice teacher is now passing judgement (Weiner, 1985). Future research should consider the ordering of the questions to align with the components of attribution theory more strongly.

Third, this study was conducted during public health restrictions associated with COVID-19. The course was offered fully asynchronously. It has been shown that COVID-19 increased student concerns about academic dishonesty (Dey, 2021) and a perceived increase in cheating due to the shift to online instruction (Ives & Cazan, 2023). These shifting outlooks on academic dishonesty may have played a role in our results. As such, future research should re-examine the items and analyses here with preservice teachers once public health restrictions have ceased to determine if shifting social factors impacted the results. In addition to reflecting on the COVID-19 context, this study was conducted prior to the release of ChatGPT, so it does not consider preservice teachers' perspectives on particularly new elements of academic dishonesty. This will be an important consideration for theory-guided research in the future.

Conclusion

Overall, our study contributes to the growing research examining academic dishonesty and preservice teachers and offers an attribution theory perspective to consider. Indeed, our findings provide valuable information about how teachers engage in causal search when presented with student actions that may be examples of academic dishonesty and the suggested consequences. Essentially, context matters for teachers when it comes to making decisions about situations of academic dishonesty, which can then impact the resultant consequences for students. As such, results from our study provide researchers, educators, and administrations with vital information about the role of facts, embellishments, and beliefs in terms of interpreting academic dishonesty. As concerns regarding academic dishonesty continue in schools, it is important to keep in mind how these behaviours are understood by educators.

References

Aaron, L., Simmons, P., & Graham-Webb, D. (2010). Academic dishonesty and unprofessional beahavior. *Radiologic Technology, 83*(2), 133-140. https://www.researchgate.net/publication/51818062_Academic_dishonesty_and_unprofessional_behavior

Amigud, A., & Pell, D. J. (2020). When academic integrity rules should not apply: a survey of academic staff. *Assessment & Evaluation in Higher Education,* 1-13. https://doi.org/10.1080/02602938.2020.1826900

Bautista, R. M., & Pentang, J. T. (2022). Ctrl C+ Ctrl V: Plagiarism and knowledge on referencing and citation among pre-service teachers. *International Journal of Multidisciplinary: Applied Business and Education Research, 3*(2), 245-257. https://doi.org/10.11594/ijmaber.03.02.10

Cajigas, B. Z. L., Gonzáles, J. L. A., Japura, G. A., Sota, C. Q., Tarazona, J. O. G., Muñoz, G. I. S., & Chuyma, R. C. (2022). Self-plagiarism: reasons and motivations for academic plagiarism or text recycling. *Journal of Positive School Psychology, 6*(2), 1697-1711.

Chibry, N., & Kurz, E. U. (2022). Pay-to-pass: Evolving online systems that undermine the integrity of student work. In *Academic integrity in Canada: An enduring and essential challenge* (pp. 203-216). Springer International Publishing.

Christensen Hughes, J. M., & McCabe, D. L. (2006). Academic misconduct within higher education in Canada. *Canadian Journal of Higher Education, 36*(2), 1-21. https://doi.org/10.47678/cjhe.v36i1.183525

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International,* 1-12.

Daumiller, M., & Janke, S. (2020). Effects of performance goals and social norms on academic dishonesty in a test. *British Journal of Educational Psychology, 90*(2), 537-559. https://doi.org/10.1111/bjep.12310

Dey, S. (2021). *Reports of cheating at colleges soar during the pandemic.* https://www.npr.org/2021/08/27/1031255390/reports-of-cheating-at-colleges-soar-during-the-pandemic

DiPaulo, D. (2022). Do preservice teachers cheat in college, too? A quantitative study of academic integrity among preservice teachers. *International Journal for Educational Integrity, 18*(1), 1-14. https://doi.org/10.1007/s40979-021-00097-3

Fontaine, S., Frenette, E., & Hébert, M. H. (2020). Exam cheating among Quebec's preservice teachers: the influencing factors. *International Journal for Educational Integrity, 16*(1), 1-18. https://doi.org/10.1007/s40979-020-00062-6

Fyfe, P. (2022). How to cheat on your final paper: Assigning AI for student writing. *AI & Society.* https://doi.org/10.1007/s00146-022-01397-z

Galloway, M. K. (2012). Cheating in advantaged high schools: Prevalence, justifications, and possibilities for change. *Ethics & Behavior, 22*(5), 378-399. https://doi-org.uml.idm.oclc.org/10.1080/10508422.2012.679143

Geddes, K. A. (2011). Academic dishonesty among gifted and high-achieving students. *Gifted child today, 34*(2), 50-56. https://doi-org.uml.idm.oclc.org/10.1177/107621751103400214

Goegan L. D., & Daniels L. M. (2023). *An attribution theory lens on plagiarism: Examining the beliefs of preservice teachers*.

Gottardello, D., & Filiz Karabag, S. (2022). Ideal and actual roles of university professors in academic integrity management: a comparative study. *Studies in Higher Education, 47*(30, 526-544. DOI: 10.1080/03075079.2020.1767051

Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review, 3*(1), 5-39. https://doi.org/10.1007/BF01323661

Graham, S. & Taylor, A. Z. (2016). Attribution theory and motivation in school. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school* (pp. 11-33). Routledge.

Hasson, R. (2017). *Cheating runs rampant, teachers search for long-term solutions.* https://thesoutherneronline.com/63861/news/cheating-runs-rampant-teachers-search-for-long-term-solutions/

Ives, B., & Cazan, A. M. (2023). Did the COVID-19 pandemic lead to an increase in academic misconduct in higher education?. *Higher Education,* (2023), 1-19. https://doi.org/10.1007/s10734-023-00996-z

Hsieh, H. F., & Shannon, S. (2018). Content Analysis In B. Frey (Ed.), *The SAGE Encyclopedia of educational research, measurement, and evaluation*, (pp. 393-394). Sage. http://dx.doi.org/10.4135/9781506326139.n149

Hylton, K., Levy, Y., & Dringus, L. P. (2016). Utilizing webcam-based proctoring to deter misconduct in online exams. *Computers & Education,* 92-93, 53-63. https://doi.org/10.1016/j.compedu.2015.10.002

Keener, T. A., Galvez Peralta, M., Smith, M., Swager, L., Ingles, J., Wen, S., & Barbier, M. (2019). Student and faculty perceptions: appropriate consequences of lapses in academic integrity in health sciences education. *BMC Medical Education, 19,* 209. https://doi.org/10.1186/s12909-019-1645-4

Kwong, T., Ng, H. M., Mark, K. P., & Wong, E. (2010). Students' and faculty's perception of academic integrity in Hong Kong. *Campus-Wide Information Systems, 27*(5), 341-355. https://doi-org.uml.idm.oclc.org/10.1108/10650741011087766

McCabe, D. L., Butterfield, K. D., & Trevino, L. K. (2012). *Cheating in college: Why students do it and what educators can do about it.* John Hopkins University Press.

Merkel, W. (2021). Simple, yet complex: Pre-service teachers' conceptions of plagiarism at a Norwegian University. *Scandinavian Journal of Educational Research*, 1-13. https://doi-org.uml.idm.oclc.org/10.1080/00313831.2021.1939778

Murdock, T. B., & Stephens, J. M., (2007). Is cheating wrong? Students' reasoning about academic dishonesty. In E. M. Anderman & T. B. Murdock (Ed.), *Psychology of academic cheating* (pp. 229-251). Academic Press. https://doi.org/10.1016/B978-012372541-7/50014-0

Nayak, B. P., & Khajuria, H. (2019). Eyewitness testimony: probative value in criminal justice system. *Egyptian Journal of Forensic Sciences, 9*(1), 1-2. https://doi.org/10.1186/s41935-018-0109-z

OECD. (2019). *TALIS 2018 results (volume i): Teachers and school leaders as lifelong learners*. TALIS, OECD Publishing. https://doi.org/10.1787/1d0bc92a-en

Otsuki, G. J. (2020). OK computer: to prevent students cheating with AI text-generators, we should bring them into the classroom. *The Conversation*. https://theconversation.com/ok-computer-to-prevent-students-cheating-with-ai-text-generators-we-should-bring-them-into-the-classroom-129905.

Peritz, A., (2022). *A.I. Is making it easier than ever for students to cheat.* https://slate.com/technology/2022/09/ai-students-writing-cheating-sudowrite.html

Perkins, M., Roe, J., Postma, D., McGaughran, J., & Hickerson, D. (2023). *Game of tones: Faculty detection of GPT-4 generated content in university assessments,* 1-12. https://doi.org/10.48550/arXiv.2305.18081

Puddifoot, K. (2020). Re-evaluating the credibility of eyewitness testimony: The misinformation effect and the overcritical juror. *Episteme, 17*(2), 255-279. doi:10.1017/epi.2018.42

International Academy for Academic Integrity, (2020). *Facts and Statistics.* https://academicintegrity.org/resources/facts-and-statistics

Pincus, H. S., & Schmelkin, L. P. (2003). Faculty perceptions of academic dishonesty. *The Journal of Higher Education, 74*(2), 196-209. https://doi.org/10.1080/00221546.2003.11777196

Romanowski, M. H. (2021). Preservice teachers' perception of plagiarism: A case from a college of education. *Journal of Academic Ethics, 20*(3), 289-309. https://doi.org/10.1007/s10805-021-09395-4

Rozhkova, M. A., & Isaeva, O. V. (2022). Plagiarism and "self-plagiarism" in scientific works in the age of digital technologies. *Digital Law Journal, 3*(2), 25-35. https://doi.org/10.38044/2686-9136-2022-3-2-25-35

Rudolph, U. & Tscharaktschiew, N. (2014). An attributional analysis of moral emotions: Naïve scientists and everyday judges. *Emotion Review, 6*(4), 344-352. 10.1177/1754073914534507

Seals, M., Hammons, J. O., & Mamiseishvili, K. (2014). Teaching assistants' preparation for, attitudes towards, and experience with academic dishonesty: Lessons learned. *International Journal of Teaching and Learning in Higher Education, 26*(1), 26-36.

Şendağ, S., Duran, M., & Fraser, R. M. (2012). Surveying the extent of involvement in online academic dishonesty (e-dishonesty) related practices among university students and the rationale students provide: One university's experience. *Computers in Human Behavior, 28*(3), 849-860. https://doi.org/10.1016/j.chb.2011.12.004

Štambuk, M., Maričić, A., & Hanzec, I. (2015). Cheating is Unacceptable, but... Teachers' perceptions of and reactions to students' cheating at schools and universities. *Croatian Journal of Education: Hrvatski časopis za odgoj i obrazovanje, 17*(Sp. Ed. 4), 259-288. https://hrcak.srce.hr/153248

Stephens, J. M. (2017). How to cheat and not feel guilty: Cognitive dissonance and its amelioration in the domain of academic dishonesty. *Theory Into Practice, 56*(2), 11-120. https://doi.org/10.1080/00405841.2017.1283571

Stoklasa, J., Talášek, T., & Stoklasová, J. (2019). Semantic differential for the twenty-first century: Scale relevance and uncertainty entering the semantic space. *Quality & Quantity, 53*(1), 435-448. https://doi.org/10.1007/s11135-018-0762-1

Thomas, A. (2017). Faculty reluctance to report student plagiarism: A case study. *African Journal of Business Ethics, 11*(1), 103-119.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548-573. https://doi.org/0033-295X/85/S00.75

Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attribution perspective. *Educational Psychology Review, 12*(1), 1-14. https://doi.org/10.1007/978-1-4615-1273-8_2

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educational Psychologist, 45*(1), 28-36. https://doi.org/10.1080/00461520903433596

Weiner, B. (2018). The legacy of an attribution approach to motivation and emotion: A no-crisis zone. *Motivation Science, 4*(1), 4–14. https://doi.org/10.1037/mot0000082

Appendices

Appendix A: Descriptions of scenarios

Scenario 1: *As part of your assessment in a Chemistry course, you have students complete weekly quizzes. While these quizzes are not for marks, but rather a form of formative assessment so you can see how students are doing. This week you notice* Student A *is very squirmy in their seat. You keep an eye on them and notice them peeking at another student's answers. At the end of class, you ask the student to stay behind and ask them about their behaviour. Student A starts apologizing and tells you how much pressure they have been under to do well at school from their parents, and that they had too many other assignments to do, so they didn't have time to study.*

Scenario 2: *You just finished a big unit in your Math class. At the end of the unit, you have decided to administer a final exam as part of the summative assessment that will be worth 30% of their final grade. You administered the final exam, and everything seemed to be fine, but after the students leave the room, you notice a sheet of paper that has the answers to the test. Based on where you found the paper, you think it belongs to Student B. The next day you ask Student B about the paper. They look concerned, almost guilty, but deny that it is their paper. After they leave, you compare the writing style to another assignment that they completed, and the writing looks very similar.*

Scenario 3: *Your class is nearing the end of a unit in Social Studies; it is time for a summative assessment test on the material. The test is worth 10% of their final grade. On exam day, you notice Student C acting odd. They keep moving their eraser around their desk in what appears to be four different positions. Your test is multiple choice, and you are concerned they are communicating the answers to another student. You talk to Student C about the eraser after class, and they admit that they were helping their friend Student D. Student D has been really sick lately and missed a lot of classes. Student D has been having a hard time keeping up with class material.*

Scenario 4: *You are an English teacher, and your class is beginning the poetry unit. As a diagnostic assessment, you ask students to write a short poem. The poems are for you to get a sense of where students are at before you begin your instruction for the unit. Student E is a student in your English class, and they submitted a wonderful poem. The problem is you think you've read it before. In fact, their older sibling may have written it last year. You go through your records and find that your suspicions are correct; Student E took their sibling's work and submitted it as their own. When you ask Student E about it, they admit to taking their sibling's work and tell you they were having writer's block and couldn't think of anything to write.*

Scenario 5: *Your class just handed in their reports on hereditary traits in your Biology course. The reports count 10% towards students' summative assessment in the course. You were reading Student F's assignment, and something seemed off. Parts of the writing seem consistent with the writing you have come to see from them, and other parts sound too sophisticated. You enter the questionable parts into a plagiarism checker online, and these sections get flagged by the program. When you ask Student F about it, they say they copied and pasted into their assignment, but it's not a big deal because everyone does it, and they were just using resources they found online.*

Scenario 6: *Your class has been busy for weeks writing their research papers in your Social Studies class. As part of this assignment, students are required to link the ideas of Nationalism to a topic that is personally meaningful. This paper is part of your summative assessment and worth 30% of their final grade. Once the papers are submitted, you notice two very similar papers from Student G and Student H. You show the two students the papers and ask them about the similarities. Student G admits that they asked Student H to send them the paper they were writing to help Student G get started as Student G was struggling to come up with an idea. Student G apologizes and says they thought they changed the paper enough to make it reflect their own ideas.*

Appendix B: Examples of facts, embellishments and beliefs from each scenarios

	Facts	Embellishment	Beliefs
1	- Weekly quizzes, formative assessment. - Peaking at another student's answers. - Apologizes and tells you how much pressure they have been under.	- They are a good kid. - They have never done this before. - Had a good reason (e.g., doing this out of fear).	- I think this it is important to apply the consequences of academic dishonestly in all scenarios. - This behaviour is minor. - The circumstances around it make it less severe.
2	- Unit final exam (summative assessment), worth 30% of their final grade. - You find sheet of paper with answers. - Student B denies paper is theirs - The writing is similar.	- Student B intentionally brought in written answers. - student refuses to take responsibility for what they have done, - The writing is the same / matches.	- Because this is such a high stakes exam, cheating has larger implications, than if it was small quiz or formative assessment. - I would also have to look at my own potential bias towards the student. - Sometimes you have to go with your gut.
3	- Test (summative assessment) worth 10% of their final grade. - Student C is acting odd and moving eraser around their desk. - Student C admits they were helping Student D who has been sick lately.	- Student C had good intentions. - Student C may have felt a lot of peer pressure. - Student D wanted a good mark on his/her exam and had no way out	- Helping out a friend is a natural thing to want to do. - The fact that this plan was premeditated between the two students suggests genuine academic dishonesty. - While student C is not "cheating" on their exam, they are helping student D cheat, student D is more in the wrong than Student C
4	- Short poem (diagnostic assessment), not for grades. - You think you've read Student E's poem before - Sibling wrote it last year. - Student E submitted it as their own, they had writers block and couldn't think of anything to write.	- Their sibling didn't know. - Their sibling willingly gave it. - Seems lazy, but they could also have issues with their confidence in writing.	- I don't think having writer's block is ever an excuse to claim someone else's work as your own. - Since this is a diagnostic assessment, I don't believe it is as severe as other situations. - It is natural to want to help a sibling. - Cheating is cheating regardless.
5	- Report (summative assessment), 10% of their final grade. - Reading Student F's assignment and something seems off, parts are entered in a plagiarism checker, and it gets flagged. - Student F says that they copy and pasted but it's not a big deal.	- They did not know how to cite resources. - Did it thinking they would not get caught. - The student made a mistake. - They have not been educated in what plagiarism is and why it is of utmost importance to avoid.	- I think this would be a serious form of academic dishonesty because of student's response. - There are no excuses for this behaviour. - Sometimes it can feel like a fine line between quoting, citing, and plagiarizing. - I think the amount of plagiarized work also plays a role.
6	- Research paper (summative assessment), 30% of their grade. - You notice two very similar papers from Students G and Student H - Student G admits that they asked Student H to send them their paper, apologizes, thought they changed the paper to reflect their own ideas.	- Student H was trying to help a friend. - Student H didn't know they were going to steal their ideas. - Student G didn't cite their source. - The papers were the same. - There is a lack of understanding by both students	- It doesn't matter the intention that these two students had, they cheated. - I believe that collaboration is important for learning. - I don't think it is necessarily a bad thing to show a friend what your assignment looks like if they are struggling. - Maybe I'm too soft, but I want to believe that students aren't actively trying to cheat and get away with it.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Construction and standardisation of an instrument measuring lecturers' persistence to publish in Scopus-indexed journals

Valentine Joseph Owan[A]
Bassey Asuquo Bassey[B]
Isaac Ofem Ubi[C]

A *Department of Educational Foundations, University of Calabar, Calabar, Nigeria; Ultimate Research Network (URN), Calabar, Cross River State, Nigeria*

B *Department of Educational Foundations, University of Calabar, Calabar, Nigeria*

C *Department of Educational Foundations, University of Calabar, Calabar, Nigeria*

Keywords

Academic publishing;
construct validity;
questionnaire development;
research productivity;
validation study.

Correspondence

owanvalentine@gmail.com [A]

Article Info

Received 4 September 2023
Received in revised form 14 November 2023
Accepted 14 November 2023
Available online 23 November 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.37

Abstract

Academic publishing is a critical aspect of research, contributing to knowledge dissemination and career advancement. However, there is a paucity of standardised instruments for assessing academics' persistence in publishing. This study developed and validated the Persistence to Publish Questionnaire (PPQ) as a valid and reliable tool for evaluating academics' persistence to publishing in Scopus-indexed journals. The PPQ was developed through a rigorous process, including item generation, content validity assessment, pretesting, and pilot testing of items. A sample of academics (n = 262) from various disciplines across two public universities in Cross River State participated in the validation process. Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were conducted to determine the instrument's factor structure and evaluate its fit. The results from the analysis revealed that the PPQ is a multidimensional instrument with five underlying factors – persistence in manuscript preparation, manuscript submission, handling revisions, dealing with rejections, and publication delays. The PPQ exhibited strong reliability in terms of internal consistency, with Cronbach's α values ranging from .89 to .99. McDonald's ω and split-half reliability corrected with the Spearman-Brown prophecy formula (rtt) results further supported its reliability. Construct validity evidence showed both convergent and discriminant validity, confirming that the PPQ effectively measures persistence to publish. The PPQ represents a valuable contribution to the field of academic publishing. It offers an opportunity for researchers and institutions to assess the degree to which academics are willing to publish, empowering researchers and institutions to identify areas of improvement and provide targeted support. This tool holds promise for enhancing research productivity and quality within the global academic community.

Introduction

The concept of persistence to publish is relatively new and has not been extensively defined in the literature. However, to understand the phrase "persistence to publish", it is important to get an overview of the word "persistence." Persistence is the quality of consistently maintaining one's determination and motivation to achieve goals, refusing to give up even in the face of potential challenges or obstacles (Quintana et al., 2022). It involves a resolute commitment to continue striving towards desired outcomes and a willingness to overcome difficulties that may arise along the way. In a more recent definition, Styk et al. (2023) conceive persistence as the capacity to embrace and persevere in the face of challenges and overcome obstacles to attain predetermined or self-established objectives.

Various related concepts, such as perseverance, grit, and tenacity, underscore individuals' unwavering commitment to achieving their goals (Duckworth et al., 2007; Williams & DeSteno, 2008). These concepts highlight the depth of individuals' engagement in pursuing their objectives, with some considering it an inherent aspect of their character (Constantin et al., 2011). The cited authors argued that persistence should be distinguished based on whether the goal is short-term or long-term, as this distinction can impact the level of effort required. When pursuing short-term goals, individuals must focus on sustaining their attention, enduring boredom, stress, and setbacks, as well as overcoming distractions or obstacles that may arise. In contrast, pursuing long-term goals demands a sustained commitment, necessitating significant resources and an extended investment of time.

The concept of persistence has been the subject of numerous studies, leading to various conceptualisations and associated terms, such as goal pursuit, commitment, self-control, courage, drive, diligence, and conscientiousness (Khindri & Rangnekar, 2022; Styk & Klinkosz, 2020; White et al., 2017). In the past, one common method used to assess adult persistence involved placing individuals in challenging situations requiring endurance (Lufi & Cohen, 1987). This was achieved through physical endurance tests (Cleeton & Knight, 1924) or by assigning them lengthy and almost unsolvable intellectual tasks (Morgan & Hall, 1926). Another approach, widely employed in educational settings, involved observing individuals in real-life situations requiring persistence and comparing dropouts to graduates in specific activities such as schools or educational programs (Wood, 1968).

Furthermore, questionnaires have been utilised as a method of measuring persistence. For instance, Wang (1932) developed a self-appraisal schedule, a 111-item questionnaire to assess persistence. Mukherjee (1974) created the Persistent Disposition Questionnaire, which he claimed could be valuable in studying achievement-oriented personality. Lufi (1979) devised a 67-item scale to evaluate persistence in the academic domain. However, these scales have not gained significant popularity, potentially due to inadequate validation. The lack of existing instruments urged Hart (2014) to develop and validate an instrument with acceptable psychometric properties that could measure persistence among higher education students. Since then, a few instruments have been developed to measure persistence in different populations (see examples in De Luca et al., 2016; Porter et al., 2020; Thalib et al., 2019).

Similarly, Kozlowski and Fouad (2022) developed a scale to measure academic persistence among college students following psychometric procedures. Additionally, Lockhart et al. (2022) constructed and established the validity and reliability of a questionnaire to measure persistence among students in Science, Technology, Engineering and Mathematics (STEM) programmes. In the same year, Quintana et al. (2022) validated the Spanish version of the "motivational persistence scale"; a scale previously developed in English by Constantin et al. (2011). Although these scales were all developed to measure persistence across different populations and contexts, none was developed to measure the concept of "persistence to publish", and none of the existing scales was developed in Africa. For these reasons, there was a need for a scale to be developed to address the gaps. Thus, the "Persistence to Publish Questionnaire (PPQ)" was developed in this study. A detailed description of the PPQ is provided in subsequent sections of this article.

Measuring lecturers' persistence to publish in Scopus-indexed journals holds significant pertinence, warranting the creation of an instrument designed to assess and evaluate this critical aspect of academic scholarship. Firstly, publishing research in Scopus-indexed journals signifies the quality and impact of an academic institution's research output. It serves as a visible marker of academic prestige, which is instrumental in attracting top talent and fostering valuable research collaborations. Secondly, Scopus-indexed journals have a vast international readership and are recognised worldwide. Thus, measuring lecturers' persistence to publish in these journals ensures that their research findings reach a broad global audience, facilitating the dissemination of knowledge on a global scale.

Furthermore, funding agencies and institutions often consider lecturers' publication records when allocating research grants and resources. Measuring this persistence improves the likelihood of securing research funding, which is essential for advancing meaningful research projects and supporting academic programmes. Additionally, a strong publication record is frequently a prerequisite for academic progression through promotions. By measuring persistence in publishing, lecturers can effectively demonstrate their commitment to scholarly contributions, which are central to career development. Moreover, measuring lecturers' persistence to publish serves as a quality assurance mechanism for universities and institutions. It ensures that faculty members consistently uphold a high standard of research and scholarship, reinforcing the institution's commitment to research excellence.

With the concept of persistence clarified; it is important to attempt to define and conceptualise "persistence to publish" by deriving ideas from the meaning of persistence. "Persistence to publish" can be defined as the sustained and determined effort of an academic staff or researcher to pursue the publication of their scholarly work consistently. It involves the commitment and dedication to overcome challenges, setbacks, and obstacles throughout the publication process. Persistence to publish reflects the continuous drive to contribute to the body of knowledge in a specific field or discipline by submitting research reports or scholarly articles to reputable journals or publishing outlets. This concept encompasses the resilience, perseverance, and tenacity required to navigate the rigorous and competitive publishing landscape, including manuscript preparation, submission, peer review, revisions, delays, rejections or eventual acceptance and dissemination of the research. To persist in the publication process, strong motivation and belief are required in the process of sharing research findings with the broader academic community and society at large. Researchers may encounter multiple rejections from journals or face challenges during peer review. However, individuals who possess the persistence to publish remain undeterred by these obstacles and view them as opportunities for improvement and growth.

Styk et al. (2023) considered persistence to be a multidimensional construct and developed a scale to measure persistence, with perseverance and perfectionism as sub-dimensions. Similarly, other researchers have also approached persistence as a multidimensional construct. For instance, one dimension of persistence can be understood as the ability to persevere despite challenges, another as the ability to persist in the face of fear, and yet another as the capacity to maintain persistence despite inadequate circumstances (Howard & Crayne, 2019). Thus, in this study, persistence to publish is viewed as a multidimensional concept that encapsulates the determination, passion, resilience, commitment to quality, and effective time management required to successfully navigate the process of sharing research findings with the wider academic community. Furthermore, due to the series of activities and potential setbacks researchers face in the publication process, persistence is required at every stage, further contributing to the multidimensionality of the concept. For instance, researchers must demonstrate persistence from initial idea and conceptualisation of their research project to the final publication and dissemination of the research outcome. At the outset, persistence is needed to formulate a research question, design a study, and obtain ethical approvals and funding. Researchers must overcome challenges in recruiting participants, collecting data, and ensuring the quality and validity of their research results,

Persistence is crucial in crafting an adequately developed and coherent manuscript during the writing phase. Researchers must invest time and effort in conducting thorough literature reviews, analysing and interpreting data, and effectively communicating their findings. This process may involve numerous revisions, addressing feedback from co-authors, mentors, and reviewers. The peer review process often presents additional hurdles that require persistence. Researchers may face rejection or receive critical feedback on their work. Persistence is essential in responding to reviewer comments, revising the manuscript, and resubmitting it for further consideration. It may take multiple rounds of revision and re-submission before achieving publication. Persistence is necessary to deal with potential delays, waiting periods, and uncertainties inherent in the publication process. Researchers may experience extended review timelines, unexpected editorial decisions, or changes in journal requirements. In line with this thinking, the conceptual model in Figure 1 was developed to show these processes with persistence at the centre of the activities.

Figure 1 shows that persistence to publish can be demonstrated across five crucial activities, including: manuscript preparation, submission, handling revisions, dealing with rejections and publication delays. In each of these activities, there are specific challenges that academic staff will face; requiring only persistence to overcome them. These challenges are presented as a bulleted list in the bigger boxes in the model. As shown in the conceptual model, the challenges vary with each activity. Single-headed arrows are used in the model to show the next activity/challenge that an academic staff will face after completing the previous activity. On the other hand, double-headed arrows show two-way activities, implying that fulfilling one and moving to the next activity could return you to the previous activity.

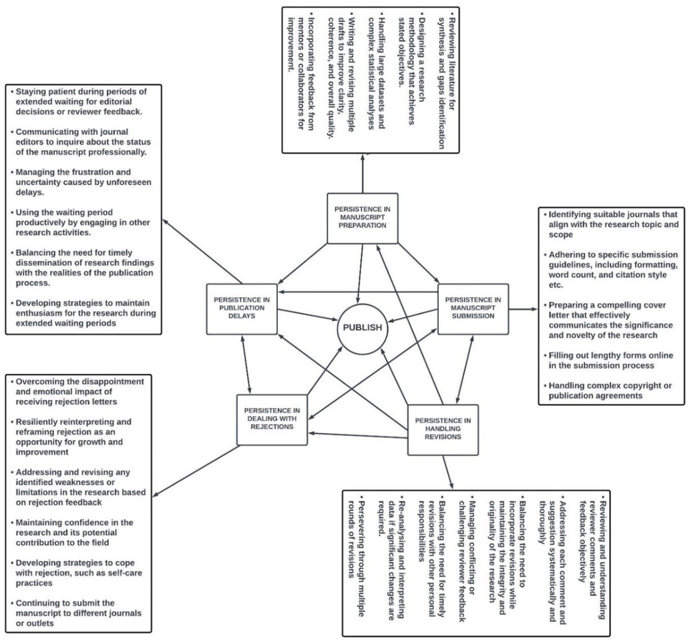


Figure 1: Conceptual model of academic staff persistence to publish.

Purpose of the study

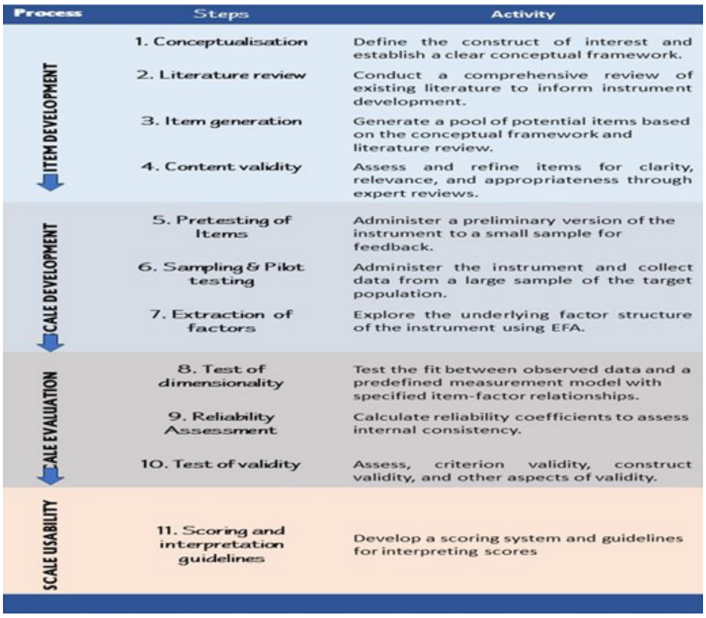
The main purpose of this study was to develop and validate the Persistence to Publish Questionnaire (PPQ). The specific objectives of this study are to:

1. Explore the factor structure of the PPQ through exploratory factor analysis (EFA).
2. Validate the factor structure using confirmatory factor analysis (CFA).
3. Assess the internal consistency of the PPQ items through reliability analysis.
4. Test the content, criterion, and construct validity of the PPQ.
5. Establish scoring procedures and guidelines for interpreting PPQ scores

Methods

Research design

The study adopted the cross-sectional survey research design. The choice of a cross-sectional survey design is justified as it efficiently captures data at a single point in time, aligning with the study's goal of developing and validating the Persistence to Publish Questionnaire (PPQ). This approach allows for the collection of diverse responses, assessment of psychometric properties, and immediate application of the PPQ. The validity process of this study will follow the framework provided in Figure 2.



Process	Steps	Activity
ITEM DEVELOPMENT	1. Conceptualisation	Define the construct of interest and establish a clear conceptual framework.
	2. Literature review	Conduct a comprehensive review of existing literature to inform instrument development.
	3. Item generation	Generate a pool of potential items based on the conceptual framework and literature review.
	4. Content validity	Assess and refine items for clarity, relevance, and appropriateness through expert reviews.
CALE DEVELOPMENT	5. Pretesting of Items	Administer a preliminary version of the instrument to a small sample for feedback.
	6. Sampling & Pilot testing	Administer the instrument and collect data from a large sample of the target population.
	7. Extraction of factors	Explore the underlying factor structure of the instrument using EFA.
CALE EVALUATION	8. Test of dimensionality	Test the fit between observed data and a predefined measurement model with specified item-factor relationships.
	9. Reliability Assessment	Calculate reliability coefficients to assess internal consistency.
	10. Test of validity	Assess, criterion validity, construct validity, and other aspects of validity.
SCALE USABILITY	11. Scoring and interpretation guidelines	Develop a scoring system and guidelines for interpreting scores

Figure 2: Framework showing the validity process of this study.

Purpose of the instrument and conceptualisation

The PPQ was developed to measure lecturers' persistence to publish in Scopus-indexed journals. The concept of persistence to publish is defined as the unwavering commitment and sustained effort of academic staff to consistently pursue the publication of their scholarly work, overcoming challenges and obstacles throughout the process. The researchers considered developing the instrument due to the lack of a previously developed instrument measuring the construct globally and in the study context.

Item generation

Some of the items included in the PPQ were adapted and modified from the "Self-Appraisal Schedule" (Wang, 1932), "Persistent Disposition Questionnaire" (Mukherjee, 1974), and Persistence in the Academic Domain Questionnaire (Lufi, 1979). However, most of the items in the PPQ were based on the researchers' experiences with the publication processes and dynamics in Scopus-indexed journals. The initial version of the PPQ comprised 40 items arranged on a six-point Likert-type scale format, with response options such as "Very Strongly Disagree," "Strongly Disagree," "Disagree," "Agree," "Strongly Agree," and "Very Strongly Agree."

Content validity

The Persistence to Publish Questionnaire (PPQ) underwent face and content validity assessment by nine independent experts, with feedback received from seven of them. These experts, primarily in Research, Measurement and Evaluation, and Educational Psychology, had extensive academic and research backgrounds. Most were aged 50 to 59, with two aged 60 or older, and one below 40. They held doctorate degrees and had over a decade of teaching and research experience. The assessment involved rating each questionnaire item for relevance, clarity, simplicity, and lack of ambiguity on a four-point scale. Higher ratings (three and four) indicated items were relevant and clear in measuring the construct, while lower ratings (one and two) suggested irrelevance or ambiguity. Their independent ratings were collated and scored, following the quantitative approach to content validity suggested by some scholars (Hadi et al., 2020; Lawshe, 1975; Zamanzadeh et al., 2014). In deciding which items should be retained, revised or deleted, the average proportion of experts' agreement was computed to determine the Item Content Validity Index (I-CVI) and Scale Content Validity Index (S-CVI), as shown in Table 1.

The results in Table 1 suggest that the I-CVIs for most variables across relevance, clarity, simplicity and ambiguity were acceptable. However, the researchers examined the I-CVIs of individual items to identify those with values lower than 0.70. Items, such as: MPR1, MPR2, MPR3, MPR4, MPR5, MSU4, MSU5, and MSU8 had an I-CV value of 0.67, respectively for clarity. Again, items, such as: MPR1, MPR2, MPR6, MSU1 and HRE1 had an I-CVI value of 0.67 for simplicity, respectively. Furthermore, items, such as: MSU1 and MSU5 had an I-CVI value of .67, respectively for ambiguity. All the items listed above were revised for improved relevance, clarity, simplicity and unambiguity, where applicable, following the experts' suggestions.

Table 1: Item- and Scale-content validity indices for persistence to publish variables.

	Basis	Manuscript preparation	Manuscript submission	Handling revisions	Dealing with rejections	Publication delays
I-CVIs	Relevance	.83 (all items)	.83 – .99	.99 (all items)	.99 (all items)	.83 – .99
	Clarity	.67 – .83	.67 – .83	.99 (all items)	99 (all items)	.83 – .99
	Simplicity	.67 – .83	.67 – .99	.67 – .99	.99 (all items)	.83 – .99
	Ambiguity	.83 (all items)	.67 – .99	.67 – .99	.83 – .99	.83 – .99
S-CVIs	Relevance	.83	.85	.99	.99	.98
	Clarity	.73	.79	.99	.99	.98
	Simplicity	.77	.83	.96	.99	.98
	Ambiguity	.83	.81	.85	.90	.92

Note: I-CVIs of .70 or above suggest that the item has an acceptable rating (with 70% or more of the experts' agreement). Values between 0.50 to .69 suggest that the items need revisions; values below .50 indicate that less than 50% of experts agree, suggesting that such items should be discarded.

Pretesting the questions

A focus group session was conducted with 10 university lecturers, six from the University of Calabar and four from the University of Cross River State, all of whom had previously published in Scopus-indexed journals. The session aimed to gather qualitative input on a survey. Participants were given physical copies of the survey to review and discuss, providing feedback on item relevance, clarity, and comprehensibility. The session, lasting about an hour, was audio-recorded for transcription. Analysis of lecturer opinions and suggestions led to survey revisions, including refining item wording and addressing ambiguity. These insights, from lecturers not involved in the main study or expert validation, improved the survey's content validity and relevance.

Sampling and pilot testing

Before conducting the pilot study, careful consideration was given to the sample size required to ensure the reliability of results, particularly in the context of structural equation modelling (SEM) techniques like confirmatory factor analysis (CFA). SEM generally demands large sample sizes for robust results (Boateng et al., 2018; Hadi et al., 2020; Owan et al., 2022b). Determining the exact sample size is complex, relying on factors like model complexity, latent variables, statistical power, and effect size. While there's no universal consensus, several guidelines exist. Some recommend a minimum of 300 respondents (Clark & Watson, 2016; Tabachnick & Fidell, 2013), while others suggest ratios like 20 to 1 (Kline, 2015), 10 to 1 (Schreiber et al., 2006), or 5 to 1 (Bentler & Chou, 1987).

In this context, a sample of 330 lecturers was targeted for the pilot study, all of whom had previously published in Scopus-indexed journals. Ultimately, 285 responses were obtained, with 45 lecturers not participating. Despite the attrition, the sample size was deemed suitable for factor analysis or SEM, considering it was close to the recommended 300. Additionally, Comrey and Lee's scale suggests 300 as a "good" sample size for SEM (Comrey & Lee, 1992).

Results

Demographic characteristics of the respondents

The demographic profile of the 285 respondents in the pilot sample revealed a balanced gender distribution, with 51.6% males and 48.4% females. Regarding age, 26.7% were under 40, 24.2% between 40 and 49, 21.4% between 50 and 59, and 27.7% were 60 or older. In terms of education, 54.4% held master's degrees, while 45.6% were doctorate holders. The rank distribution among participants showed that 18.9% were Assistant Lecturers, 14.7% were Lecturer II, 19.3% were Lecturer I, 13.7% were Senior Lecturers, 20.4% were Associate Professors, and 13.0% were Professors. These demographic details provide a comprehensive overview of the pilot study's participant characteristics.

Exploratory Data Analysis

To evaluate data normality, multiple tests, including histograms, Shapiro-Wilk's, Kolmogorov-Smirnov, and Q-Q plots, were employed. While minor deviations from normality were observed in the histograms, with some items exhibiting bell-like shapes, most Shapiro-Wilk and Kolmogorov tests yielded insignificant results. Efforts to detect outliers included scrutinising the dataset for out-of-range values introduced during data imputation, but none were found. Boxplots were also utilised to identify potential outliers across all items, yielding no outliers. Data were assessed for multivariate outliers using a Mahalanobis Distance Test (Tabachnick & Fidell, 2013), resulting in the identification and removal of 23 such outliers. This process reduced the number of cases from 285 to 262. Descriptive statistics were computed and Table 2 shows that the mean values range from 3.48 to 3.70. These values are all acceptable for a six-point Likert scale instrument. The standard deviations, ranging from 1.61 to 1.80, indicate some variability or dispersion in the responses around the mean. The skewness values range from -0.15 to 0.23, and kurtosis values range from -1.41 to -1.06. These results provided further evidence that the data possess some normal distribution properties.

Table 2: Descriptive atatistics of the items in the PPQ.

Dimensions	Items	M	SD	Skew.	Kurt.
Persistence in Manuscript preparation	MPR1	3.70	1.71	-0.15	-1.25
	MPR2	3.70	1.71	-0.15	-1.25
	MPR3	3.70	1.71	-0.15	-1.25
	MPR4	3.37	1.65	0.14	-1.22
	MPR5	3.70	1.71	-0.15	-1.25
	MPR6	3.70	1.71	-0.15	-1.25
	MPR7	3.51	1.66	-0.04	-1.23
	MPR8	3.70	1.71	-0.15	-1.25
Persistence in Manuscript Submission	MSU1	3.60	1.68	-0.09	-1.26
	MSU2	3.55	1.64	0.01	-1.21
	MSU3	3.48	1.72	0.04	-1.26
	MSU4	3.62	1.79	-0.10	-1.37
	MSU5	3.65	1.71	0.01	-1.32
	MSU6	3.51	1.70	-0.04	-1.25
	MSU7	3.66	1.69	-0.09	-1.28
	MSU8	3.58	1.73	-0.07	-1.30
Persistence in Handling Revisions	HRE1	3.66	1.74	-0.15	-1.25
	HRE2	3.41	1.72	0.05	-1.30
	HRE3	3.56	1.73	-0.05	-1.28
	HRE4	3.58	1.75	-0.15	-1.33
	HRE5	3.53	1.69	-0.04	-1.21
	HRE6	3.57	1.77	-0.10	-1.32
	HRE7	3.58	1.70	-0.08	-1.20
	HRE8	3.55	1.66	-0.09	-1.20
Persistence in Dealing with Rejections	DWR1	3.44	1.66	0.13	-1.19
	DWR2	3.39	1.69	0.09	-1.25
	DWR3	3.29	1.64	0.23	-1.09
	DWR4	3.60	1.69	-0.10	-1.18
	DWR5	3.52	1.69	-0.02	-1.28
	DWR6	3.32	1.61	0.16	-1.06
	DWR7	3.32	1.64	0.18	-1.13
	DWR8	3.47	1.67	0.03	-1.21
Persistence in Publication Delays	PDE1	3.30	1.71	0.23	-1.19
	PDE2	3.35	1.73	0.18	-1.24
	PDE3	3.29	1.72	0.20	-1.21
	PDE4	3.34	1.73	0.16	-1.25
	PDE5	3.37	1.72	0.15	-1.24
	PDE6	3.36	1.72	-0.02	-1.28
	PDE7	3.45	1.80	0.04	-1.41
	PDE8	3.39	1.72	0.14	-1.24

Extraction of factors

Exploratory Factor Analysis (EFA) was performed on the pilot data obtained for the items in PPQ. Principal Axis Factoring (PAF) was the extraction method, with a varimax rotation, used to identify the factorial structure of the scale. The analysis was set to extract factors with Eigenvalues greater than one, while items with loadings below .50 were suppressed. It initially yielded an 11-factor outcome. Sampling accuracy was acceptable (KMO = 0.813), and Bartlett's test yielded a significant value, $\chi 2(780) = 8502.19$, $p< .001$. The 11 factors cumulatively explained 72.93% of the total variance. Nevertheless, examining the rotated factor matrix revealed several problematic and dysfunctional items. For instance, several items did not load unto any factor, such as PDE7, MPR4, DWR8, DWR5, MPR7, HRE8, MSU4, HRE4, DWR4 and MSU5. Thus, they were deleted. Two items (HRE2 and PDE6) were deleted because they did not correlate with other items in the analysis. Furthermore, two items (MSU3 and MSU6) loaded to factor 6. However, a minimum of three items are needed to retain a factor. As a result, the two items were also deleted.

The analysis was re-performed without the problematic items using the same settings. The result extracted five factors with Eigenvalues greater than one. The five factors jointly explained 79.70% of the total variance. The Scree plot in Figure 3 also shows that five factors have Eigenvalues greater than one. Relatively, factors 1, 2, 3, 4, and 5 explained 21.95, 19.73, 16.24, 11.49 and 10.29% of the total variance, respectively. The rotated factor matrix was examined for naming purposes. The factors were named "persistence in manuscript preparation (factor 1)", "persistence in publication delays" (factor 2), "persistence in handling revisions" (factor 3), "persistence in dealing with rejections" (factor 4) and "persistence in manuscript submission" (factor 5). The KMO value of sampling adequacy was 0.87, while Bartlett's test of sphericity was statistically significant, $\chi2(325) = 8311.95$, $p < .001$. The summarised results can be found in Table 3.

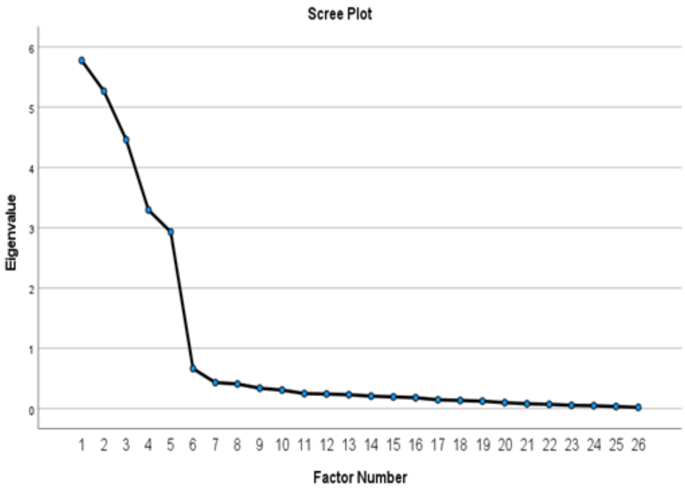


Figure 3: Scree plot showing the factors in the PPQ with their Eigenvalues.

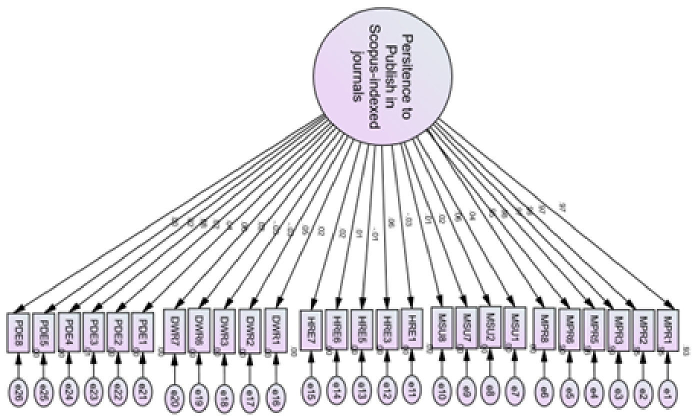
Table 3: Loadings of Exploratory Analysis for the PPQ.

Factors	Item Label	EFA			
		λ	λ^2	ε	z
Persistence in Manuscript Preparation	MPR6	.98	.96	.04	26.06
	MPR3	.98	.95	.05	20.58
	MPR2	.97	.95	.05	18.27
	MPR5	.97	.94	.06	14.90
	MPR1	.97	.93	.07	14.03
	MPR8	.95	.89	.11	9.00
	SUM	5.81	5.62	0.38	102.84
Persistence in Publication Delay	PDE1	.94	.88	.12	7.81
	PDE3	.94	.88	.12	7.68
	PDE5	.93	.87	.13	7.09
	PDE2	.93	.86	.14	6.41
	PDE8	.92	.84	.16	5.91
	PDE4	.90	.81	.19	4.84
	SUM	5.55	5.14	0.86	39.74
Persistence in Handling Revisions	HRE5	.91	.83	.17	5.36
	HRE7	.91	.83	.17	5.23
	HRE3	.89	.79	.21	4.32
	HRE6	.87	.76	.24	3.55
	HRE1	.86	.74	.26	3.35
	SUM	4.44	3.95	1.05	21.82
Persistence in Dealing with Rejections	DWR3	.81	.65	.35	2.31
	DWR1	.80	.64	.36	2.23
	DWR6	.79	.63	.37	2.11
	DWR2	.78	.61	.39	2.03
	DWR7	.77	.60	.40	1.92
	SUM	3.96	3.13	1.87	10.62
Persistence in Manuscript Submission	MSU7	.94	.89	.11	8.67
	MSU2	.90	.81	.19	4.74
	MSU1	.88	.78	.22	3.97
	MSU8	.54	.30	.70	0.77
	SUM	3.27	2.77	1.23	18.15

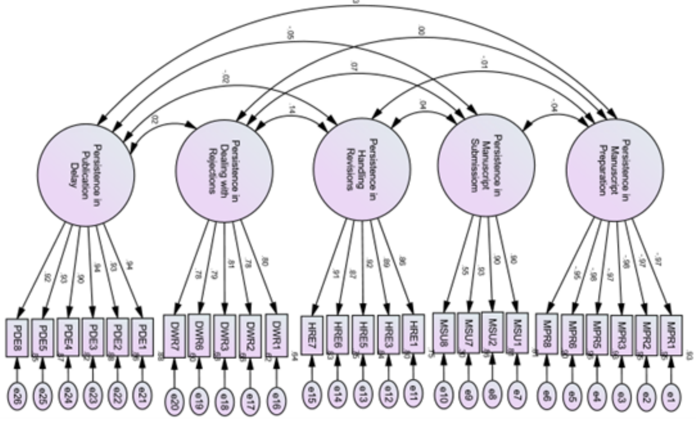
Test of dimensionality

A dimensionality test was performed through Confirmatory Factor Analysis (CFA). This study used four competing CFA models to determine the best-fitting model. These models include: the single-factor model (Model 1), the oblique or correlated factor model (Model 2), the hierarchical or second-order factor model (Model 3) and the Bifactor model (Model 4). Table 4 depicts the single-factor, oblique, second-order factor and bifactor CFA models. Table 5 shows that the single-factor model does not fit the data well. The $\chi2$ value is 5394.10 with 299 degrees of freedom and a significant p-value. The RMSEA value of 0.256 and SRMR value of 0.246 exceeded the recommended benchmark, indicating a poor fit. The CFI value of 0.386 and TLI value of 0.333 fell below the desired criteria, further supporting a poor fit for this model. The oblique model demonstrates a better fit compared to the single-factor model. The $\chi2$ value is 537.23, with 289 degrees of freedom and a significant p-value. The RMSEA and SRMR values of value .057 and .027 met the requirements for acceptability, suggesting a good fit. The CFI value of .970 and TLI value of .966 exceeded the desired thresholds, further supporting the acceptability of this model.

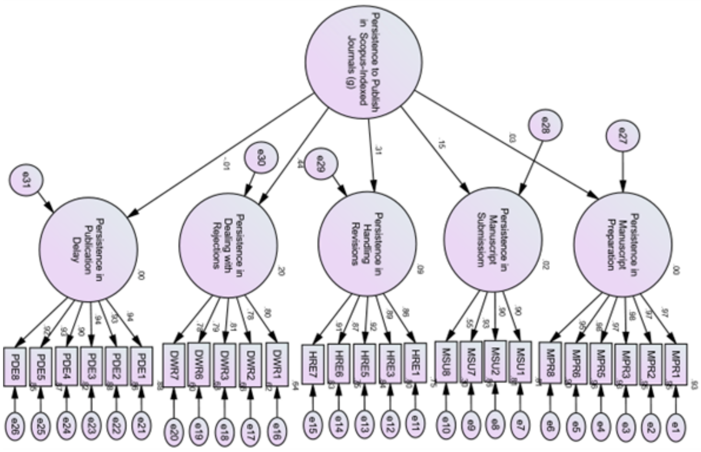
The second-order model shows an even better fit compared to the single factor and oblique models. Although the Chi-square test is significant $\chi2(294) = 538.64$, $p < .05$, the RMSEA and SRMR values of .056 and 0.032 met the requirements for retaining the model. Furthermore, the CFI and TLI values of .971 and .967 exceeded the desired thresholds, further supporting a better fit of this model. Lastly, the bifactor model demonstrates the best fit among the considered models. Even though the Chi-square test is significant, $\chi2(273) = 417.72$, $p < .05$, the RMSEA and SRMR values of .045 and .013 met the recommended benchmark thresholds, indicating the best fit. The CFI value of .983 and TLI value of .979 exceeded the desired thresholds, further supporting the superior fit of this model. The single-factor model had a poor fit, while the oblique, second-order, and bifactor models showed progressively better fit. The bifactor model displayed the best fit among the models considered, with the lowest RMSEA and SRMR values and the highest CFI and TLI values.



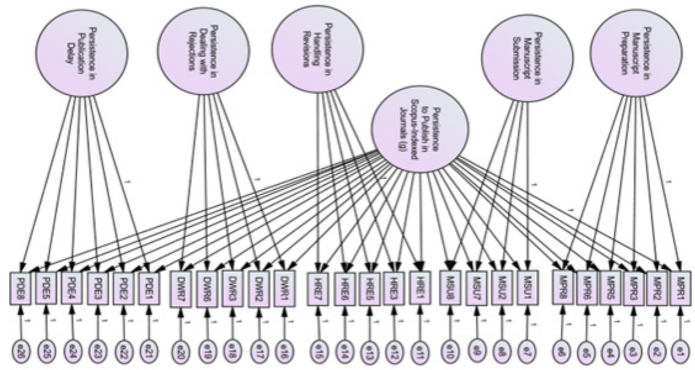
Model 1: Single-factor or unidimensional CFA model.



Model 2: Oblique or correlated factors CFA model.



Model 3: Nested or higher-order CFA model.



Model 4: Bi-factor CFA model.

Table 4: Standardised confirmatory factor analysis loadings for the single, oblique, second-order and bifactor models of the PPQ.

	Single Factor Model	Oblique or Correlated Factor Model					(Schmid-Leiman transformation) Second-order Model						Bi-Factor or Nested Model					
	G	1	2	3	4	5	g	1	2	3	4	5	g	1	2	3	4	5
MPR1	.97	-.97					.03	.97					-.05	.97				
MPR2	.97	-.97					.03	.97					-.06	.97				
MPR3	.98	-.98					.03	.98					-.08	.98				
MPR5	.97	-.97					.03	.97					1.0	1.0				
MPR6	.98	-.98					.03	.98					-.06	.98				
MPR8	.04	-.95					.03	.95					-.04		.94			
MSU1	.06		.90				.13		.90				.04		.90			
MSU2	.02		.90				.13		.90				.02		.90			
MSU7	.01		.93				.14		.93				.01		.93			
MSU8	-.03		.55				.08		.55				-.01		.55			
HRE1	.06			.86			.27			.78			-.01			.86		
HRE3	-.01			.89			.27			.81			-.01			.89		
HRE5	.01			.92			.28			.83			.00			.92		
HRE6	.02			.87			.27			.79			.02			.87		
HRE7	.02			.91			.28			.82			.00			.91		
DWR1	.05				.80		.35				.64		.01				.80	
DWR2	-.03				.79		.36				.63		-.01				.79	
DWR3	-.03				.81		.36				.65		-.02				.81	
DWR6	.02				.79		.36				.64		-.02				.79	
DWR7	.06				.78		.36				.62		-.01				.78	
PDE1	.95					.94	-.01					.92	.01					.94
PDE2	.02					.93	-.01					.91	.02					.93
PDE3	.01					.94	-.01					.92	.02					.94
PDE4	.00					.90	-.01					.89	.02					.90
PDE5	.02					.93	-.01					.91	.01					.93
PDE8	.08					.92	-.01					.90	.02					.92
Latent S^2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.00	.02	.09	.20	.00	1.0	1.0	1.0	1.0	1.0	1.0
λ (unique)								1.00	1.0	.91	.80	.98						
λ (Second order)								.03	.15	.31	.44	-.01						

Notes: 1 = Manuscript preparation; 2 = Manuscript submission; 3 = Handling revisions; 4 = Dealing with rejections; 5 = Persistence in publication delays.

Table 5: Comparison of the four competing models.

Model	χ^2(df)	RMSEA	SRMR	CFI	TLI
Single factor	5394.10(299), p < .05	.256	.246	.386	.333
Oblique	537.23(289), p < .05	.057	.027	.970	.966
Second-order	538.64(294), p < .05	.056	.032	.971	.967
Bi-factor	417.72(273), p < .05	.045	.013	.983	.979
Recommended Benchmarks	p > .05	< .08	< .08	≥.95	≥.95

Bifactor Model Test

The bifactor model proved the best-fitting model among the four competing CFA models using traditional fit indices, such as RMSEA, Chi-Square, SRMR, TLI, and CFI. However, there has been much criticism of using traditional fit indices to evaluate the bifactor model. Relying solely on traditional goodness-of-fit indices, such as CFI and RMSEA, when evaluating bifactor models using SEM techniques can result in false positives (Sellbom & Tellegen, 2019; Ventura-León et al., 2021). This is because these indices do not adequately consider the influence of the general factor and specific factors on the individual items (Bonifay et al., 2017; Flores-Kanter et al., 2018). Research suggests that traditional goodness-of-fit indices may statistically favour bifactor models (Morgan et al., 2015). This may explain why the model outperformed all other models across the three instruments in the current study. Therefore, it is important to employ alternative methods and indices that explicitly assess the impact of the general and specific factors in bifactor models to evaluate their fit and validity comprehensively. Therefore, some auxiliary measures were used to evaluate the bi-factor model for increased reliability and acceptability.

The Excel package "BifactorIndicesCalculator" developed by Dueber (2017) was used to generate the auxiliary fit indices based on the results of general and specific factors obtained from the AMOS program and earlier reported in Tables 4. These include Omega Coefficients, Explained Common Variance (ECV), Percentage of Uncontaminated Correlations (PUC), Factor Determinacy (FD), Construct Replicability (H), and Average Relative Parameter Bias (ARPB). Omega coefficients assess the internal reliability of multidimensional composites in various forms: Total Omega (ω), Subscale Omega (ωS), Hierarchical Omega (ωH), and Hierarchical Omega for Subscale (ωHS). These coefficients aid in assessing the reliability, dimensionality, and validity of measurement models. Based on these parameters, the bifactor model was re-evaluated, with results in Table 6.

Table 6 shows that the ω for the general factor is .96. At the sub-scale levels, the values of ωS are greater than .70, suggesting high internal consistency across the sub-scales and instruments. For ωH, values of .80 or higher are needed to provide unidimensionality (Rodriguez et al., 2015). An examination of Table 6 shows that unidimensionality was not achieved for the general factor of the instrument. This is because the value of the ωH was .01, well below the required value of ≥ .80. The results provide support that the instrument is multidimensional. However, at the subscale level, all the ωH values are above the .80 threshold, suggesting that the items within each factor (subscale) are measuring, to a large extent, a dominant trait/factor.

A look at the ECV values revealed a value of .048. This value is well below the recommended value of 0.60 or higher needed to justify the unidimensionality of the instrument. Therefore, the instrument can be considered multidimensional based on the ECV grounds at the scale level. IECV values of .85 at the item level will yield a unidimensional model (Stucky & Edelen, 2015). Table 6 shows that the IECV range of values for the PPQ is .00 to .49. Again, these values are well below the .85 threshold. These results provide strong support for the bifactor model across the three instruments.

In using the PUC criteria, a value less than .80 is needed to validate the results of the ECV. It has been suggested that when the PUC is greater than .80, the ECV is irrelevant (Ventura-León et al., 2021). As shown in Table 6, the PUC value of the instrument is marginally greater than .80, meaning that the ECV results reported earlier should be reconsidered. However, the PUC being greater than .80 is not a sufficient reason to consider the instrument unidimensional since, in addition to being greater than .80, the ECV must be greater than .60 for unidimensionality to be established. From another perspective, PUC and ECV must be greater than .70 to achieve unidimensionality (Rodriguez et al., 2015). Therefore, the results in Table 6 do not meet these conditions, suggesting that the instrument is multidimensional.

For the FD, values of .80 or above (Gorsuch, 1983) or greater than .90 (Grice, 2001; Rodriguez et al., 2015) are required to allow an estimate of the general factor score. Table 6 shows that the FD coefficient met the recommended benchmark. This further supports the choice of the bifactor model, where the general factor can be estimated alongside specific factors. Furthermore, evidence was gathered for all the subscales regarding the H values. The H values of three scales and all subscales are greater than .70, as recommended by experts. This means that the instrument and its corresponding subscales are well-defined by the number of retained items measuring them and are more likely to be stable in other studies (Ventura-Léon et al., 2021).

The ARPB measures the difference between the factor loadings of the bifactor facto and general factor model. Scholars have recommended values in the range of .12 and .15 as ideal for retaining the general factor model; otherwise, the bifactor model would be favoured. Table 6 shows that the ARPB value is 0.22, outside the range of permissible values. This provides support for the multidimensional bifactor model for the PPQ.

Table 6: Auxiliary fit assessment of the dimensionality of the bifactor CFA model of the PPQ.

Dimensions	ECV (S&E)	ECV (NEW)	ω/ω_S	ω_H/ω_{HS}	Relative ω	H	FD
General Factor	.05	.05	.96	.01	.01	1.0	.99
Manuscript Preparation	.22	.82	1.0	1.0	.98	.98	.99
Manuscript Submission	.17	1.0	.93	.93	1.0	.96	.98
Handling revisions	.18	1.0	.95	.95	1.0	.95	.98
Dealing with rejections	.14	1.0	.89	.89	1.0	.89	.95
Publication delays	.24	1.0	.97	.97	1.0	.97	.99
IECV = .000 to 0.492; PUC = 0.831; ECV = .048; ARPB = .22							

Notes: ω = Omega; ω_H = Omega Hierarchical; ω_S = Omega for Specific factors; ω_{HS} = Omega hierarchical for the subscales; H = Construct Replicability; FD = Factor Determinacy; IECV = Item Explained Common Variance; PUC = Percent of Uncontaminated Correlations; ECV = Explained Common Variance (ECV), Average Relative Parameter Bias (ARPB). For unidimensionality of the general or subscales: $\omega > .80$; $\omega_H > .80$; $\omega_S > .80$; $\omega_{HS} > .80$; H > .80; FD > .90; IECV > .80; PUC > .70; ECV > .60; ARPB = .12 to .15

Reliability assessment

The reliability evidence for the instruments was gathered using three measures of internal consistency – Cronbach alpha (α), McDonald's omega (ω) and split-half reliability corrected with the Spearman-Brown prophecy formula (rtt). Using multiple measures of internal consistency is informed by their overlapping strengths and weaknesses, and using all three allowed the researchers to triangulate their results and obtain a more robust understanding of internal consistency. For instance, relying solely on Cronbach's alpha as a measure of internal consistency has several weaknesses. First, Cronbach's alpha assumes that all items in a scale are essentially measuring the same underlying construct (Dunn et al., 2014; McNeish, 2018), which may not always be the case. Second, Cronbach's alpha is sensitive to the number of items in a scale (Flora, 2020), meaning that longer scales tend to yield higher alpha values, even if some items are weakly related to the overall construct. Conversely, shorter scales may have lower alpha values, even if they are highly internally consistent. To address these weaknesses, researchers often use additional measures of internal consistency, such as McDonald's omega and split-half reliability, to obtain a more comprehensive assessment of the reliability of measurement instruments.

Table 7 shows that the questionnaire demonstrates strong internal consistency, with Cronbach's α values ranging from .89 to .99, indicating high internal consistency across all factors (such as persistence in manuscript preparation, publication delays, handling revisions, dealing with rejections, and manuscript submission). Similarly, MacDonald's ω reliability estimates range from .89 to .99, indicating good reliability. The split-half reliability corrected with the Spearman-Brown prophecy formula (rtt) values ranges from .87 to .98, suggesting strong internal consistency for all factors.

Table 7: Cronbach's alpha and McDonald's omega and Split-half reliability estimates of the PPQ.

Sub-scales	K	M	SD	α	ω	r_{tt}
Persistence in Manuscript Preparation	6	22.00	9.94	.99	.99	.98
Persistence in Manuscript Submission	4	14.33	5.81	.89	.97	.98
Persistence in Handling Revisions	5	17.91	7.88	.95	.95	.95
Persistence in Dealing with Rejections	5	16.75	6.91	.89	.89	.90
Persistence in Publication Delays	6	20.04	9.69	.97	.90	.87

K = Number of items in a subscale; M = Mean; SD = Standard deviation; α = Cronbach alpha; ω = McDonald's omega. r_{tt} = Split-half reliability corrected with Spearman-Brown prophecy formula. For all reliability measures, values > .70 indicate acceptable internal consistency.

Item level reliability analysis was performed to assess the quality and consistency of individual items within each sub-scale. It is useful for identifying weak or problematic items and understanding how each item contributes to measuring the underlying construct. For persistence in manuscript preparation sub-scale, Table 8 shows that items exhibit robust internal consistency, with Cronbach's alpha (α) and McDonald's Omega (ω) values of .99 across all items. This suggests that these items effectively measure the same underlying construct. The corrected item-total correlations (ITC) for these items are also notably high, indicating strong item-scale relationships. Additionally, the squared multiple correlations (SMC) suggest that a significant proportion of each item's variance is accounted for by the Manuscript Preparation scale. Consequently, removing any of these items is unlikely to enhance the internal consistency of the scale.

The persistence in manuscript submission sub-scale demonstrates good internal consistency for all items, with α and ω values ranging from .83 to .94. However, item MSU8 displays a relatively lower α value compared to the others. The corrected item-total correlations (ITC) for MSU1, MSU2, MSU7, and MSU8 are moderately high, indicating reasonably strong item-scale relationships. The squared multiple correlations (SMC) suggest that a substantial proportion of each item's variance is explained by the persistence in the manuscript submission sub-scale. However, removing item MSU8 might have a marginal positive impact on the scale's internal consistency.

The persistence in handling revisions sub-scale showcases high internal consistency, with item α and ω values consistently at .94. This indicates that the items collectively measure the same construct effectively. The corrected item-total correlations (ITC) for these items are also notably high, denoting strong item-scale relationships. Furthermore, the squared multiple correlations (SMC) suggest that a significant proportion of each item's variance is accounted for by the persistence in handling revisions sub-scale. Consequently, removing any of these items is unlikely to improve the internal consistency of the scale.

Table 8: Item-level reliability estimates for the PPQ.

Scale	Items	M	SD	SM	SS^2	ITC	SMC	α	ω
Manuscript preparation	MPR1	3.70	1.71	18.30	68.75	.96	.93	.99	.99
	MPR2	3.62	1.70	18.37	68.81	.97	.94	.99	.99
	MPR3	3.71	1.69	18.29	68.78	.97	.95	.99	.99
	MPR5	3.66	1.71	18.34	68.75	.96	.94	.99	.99
	MPR6	3.62	1.70	18.37	68.49	.98	.96	.99	.99
	MPR8	3.69	1.71	18.31	69.13	.94	.89	.99	.99
Manuscript submission	MSU1	3.60	1.68	10.73	19.19	.80	.75	.83	.85
	MSU2	3.55	1.64	10.77	19.16	.83	.75	.82	.85
	MSU7	3.66	1.69	10.67	18.33	.87	.78	.80	.83
	MSU8	3.51	1.73	10.82	22.26	.53	.32	.94	.94
Handling Revisions	HRE1	3.66	1.74	14.25	40.43	.84	.71	.94	.94
	HRE3	3.56	1.73	14.35	40.08	.87	.76	.94	.94
	HRE5	3.53	1.69	14.38	40.26	.88	.80	.94	.94
	HRE6	3.57	1.77	14.34	40.06	.85	.72	.94	.94
	HRE7	3.58	1.70	14.33	40.25	.88	.78	.94	.94
Dealing with rejections	DWR1	3.44	1.66	13.31	31.11	.75	.57	.87	.87
	DWR2	3.39	1.69	13.37	31.05	.73	.55	.87	.87
	DWR3	3.29	1.64	13.47	31.22	.76	.58	.87	.87
	DWR6	3.32	1.61	13.43	31.80	.74	.56	.87	.87
	DWR7	3.32	1.64	13.44	31.70	.73	.53	.88	.87
Publication delays	PDE1	3.30	1.71	16.74	65.50	.92	.86	.97	.97
	PDE2	3.35	1.73	16.69	65.45	.91	.85	.97	.97
	PDE3	3.29	1.72	16.75	65.28	.92	.86	.97	.97
	PDE4	3.34	1.73	16.70	65.88	.89	.80	.97	.97
	PDE5	3.37	1.72	16.67	65.42	.92	.86	.97	.97
	PDE8	3.39	1.72	16.65	65.76	.90	.84	.97	.97

M = Item mean; SD = Item Standard Deviation; SM = Scale Mean if Item Deleted; SS^2 = Scale Variance if Item Deleted; ITC = Corrected Item-Total Correlation; SMC = Squared Multiple Correlation; α = Cronbach's Alpha if Item Deleted; ω = McDonald's Omega if Item Deleted

For persistence in dealing with rejections sub-scale exhibit good internal consistency, with α and ω values consistently at .87. This suggests that the items collectively measure the intended construct reasonably well. The corrected item-total correlations (ITC) for these items are moderately high, indicating reasonably strong item-scale relationships. While the squared multiple correlations (SMC) suggest that a moderate proportion of each item's variance is explained by the Dealing with Rejections scale, removing any of these items might slightly enhance the scale's internal consistency.

Regarding persistence in publication delays sub-scale, the items demonstrate high internal consistency, with α and ω values consistently at .97. This indicates that the items effectively measure the same underlying construct. The corrected item-total correlations (ITC) for these items are also notably high, indicating strong item-scale relationships. Additionally, the squared multiple correlations (SMC) suggest that a substantial proportion of each item's variance is accounted for by the persistence in publication delays sub-scale. As a result, removing any of these items is unlikely to improve the internal consistency of the scale.

Convergent and discriminant validity tests

The result of the construct validity of the instrument is presented in Table 9. The Average Variance Extracted approach was used, with values above .50 providing evidence of convergent validity (Owan, et al., 2022a; Rönkkö & Cho, 2022). The PPQ achieved convergent validity since the range of AVE values is .63 to .94, above the cut-off value of .50.

The instrument was also assessed for discriminant validity using the Fornell-Larcker approach (Fornell & Larcker, 1981). In this approach, the Average Variance Extracted (AVE) square root is computed for each factor, and these values are compared with the correlation estimates off the diagonal. For discriminant validity to be achieved, the square root of the AVE for each factor should be greater than the correlation estimates between that factor and other factors (off-diagonal correlations). This indicates that each factor shares more variance with its measures than with measures of other factors (Owan et al., 2022a). As shown in Table 9, all the bolded values are greater than the correlation coefficients, suggesting that discriminant validity is achieved for all the factors in the instrument.

Table 9: Construct validity evidence for the PPQ.

S/N.	Factors	AVE	CR	1	2	3	4	5
1	Persistence in manuscript preparation	.94	.99	**.97**				
2	Persistence in publication delays	.86	.97	.03	**.93**			
3	Persistence in handling revisions	.79	.95	.01	-.03	**.89**		
4	Persistence in dealing with rejections	.63	.89	.00	.02	.14	**.79**	
5	Persistence in manuscript submission	.69	.90	.04	-.05	.04	.07	**.83**

Notes: AVE = Average variance extracted (Values > .50 indicate convergent validity); CR = Composite reliability estimates (Values > .70 are acceptable); Bolded values are square roots of AVE. The square root of AVE > Correlation estimates off-diagonal for discriminant validity.

Scoring and interpretation guidelines

Scoring the Persistence in Publishing Questionnaire (PPQ) involves several key steps to effectively measure an individual's level of persistence in the academic publishing process. These steps are designed to provide a comprehensive assessment of an individual's attitudes and behaviours related to academic publishing, and the scoring guidelines ensure consistency and reliability in data interpretation. Firstly, the PPQ utilises a 6-point Likert scale for item responses, ranging from 1 (Strongly Disagree) to 6 (Strongly Agree). Each item on the questionnaire corresponds to a specific aspect of persistence in academic publishing, and respondents provide their level of agreement or disagreement with these statements. The PPQ is structured into five distinct factors, each representing a unique dimension of persistence in the publishing process. These factors include "Persistence in Manuscript Preparation," "Persistence in Publication Delays," "Persistence in Handling Revisions," "Persistence in Dealing with Rejections," and "Persistence in Manuscript Submission."

To calculate factor scores, researchers should sum the scores of the individual items belonging to each factor. For example, to determine the "Persistence in Manuscript Preparation" factor score, sum the scores of items MPR1, MPR2, MPR3, MPR5, MPR6, and MPR8. Repeat this process for each factor to obtain factor-specific scores. Additionally, a total score for the PPQ can be computed by summing all the item scores across all factors. This overall score provides a comprehensive measure of an individual's persistence in academic publishing. Higher total scores indicate a greater level of persistence, while lower scores suggest lower persistence. Researchers should consider interpreting subscale scores individually to gain insights into specific aspects of persistence. Each subscale reflects a different dimension of the publishing process, enabling a more nuanced analysis of an individual's publishing persistence. Factor-level analysis can also be valuable, allowing researchers to examine patterns of persistence in each specific area. This approach can help identify strengths and weaknesses in different aspects of the academic publishing process.

While specific score thresholds can be established for various purposes, researchers should base these thresholds on their research objectives and the distribution of scores within their sample. To ensure the reliability and validity of the PPQ scores, it is crucial to follow these scoring guidelines consistently across different samples and studies. Additionally, considering context and research objectives when interpreting scores is essential for drawing meaningful conclusions based on PPQ results.

Discussion

The current study on the development and validation of the Persistence to Publish Questionnaire (PPQ) is firmly anchored in the existing body of research on the persistence of academics in the realm of scholarly publishing. It builds upon and extends prior research in several ways, contributing to the ongoing discourse on the factors that shape academics' unwavering commitment to publishing their scholarly work. First and foremost, the study addresses a critical gap in the literature by providing a comprehensive instrument, the PPQ, designed to measure the construct of persistence to publish. While previous studies have explored various aspects of academic publishing, such as barriers, motivations, and publication productivity (e.g., Andriani et al., 2020; Lambovska, 2022; Lambovska & Todorova, 2021), there has been a notable absence of a standardised tool to assess the overarching concept of persistence in this context. The PPQ fills this void and offers researchers a reliable and validated instrument for measuring academics' persistence in publishing.

Moreover, the study aligns with prior research that emphasises the significance of understanding the challenges and obstacles academics face in the publishing process. The concept of persistence to publish is rooted in the recognition that scholars often encounter a multitude of hurdles (See Cleeton & Knight, 1924; Lufi & Cohen, 1987; Morgan & Hall, 1926). These challenges have been explored individually in past research, but the PPQ synthesises them into a coherent framework, acknowledging their interconnectedness and cumulative impact on academics' publication persistence.

In terms of item generation, the study draws on both established scales (such as those developed by Lufi, 1979; Mukherjee, 1974; Wang, 1932) and the researchers' experiences, a methodological approach that echoes previous research efforts to develop contextually relevant measurement tools. This fusion of existing scales with experiential insights reflects a commitment to building upon the strengths of prior research while tailoring the instrument to the unique dynamics of publishing in Scopus-indexed journals.

The study also contributes to the ongoing discussion on the psychometric properties of measurement instruments. It supports previous studies attempting to measure the concept of academic persistence (Constantin et al., 2011; De Luca et al., 2016; Kozlowski & Fouad, 2022; Lockhart et al., 2022; Porter et al., 2020; Quintana et al., 2022; Thalib et al., 2019), even though the focus and contexts are different. Nevertheless, by employing a rigorous process of content validation, including expert assessments and pretesting with experienced lecturers, the study aligns with previous research emphasising the importance of face and content validity in instrument development (e.g., Boateng et al., 2018; Owan et al., 2022a; Owan et al., 2022d). Additionally, the use of exploratory and confirmatory factor analysis mirrors the methodological choices made in earlier studies that sought to establish the dimensionality and construct validity of measurement instruments (Ekpenyong et al., 2022; Owan et al., 2022c). The multidimensions of the PPQ, including manuscript preparation, submission delays, revisions, rejections, and publication delays, support previous research, which reveals that persistence is a multidimensional variable (Howard & Crayne, 2019; Styk et al., 2023).

Furthermore, the study's exploration of the bifactor model, including the use of auxiliary measures to comprehensively assess its fit and validity, is in line with emerging research that highlights the limitations of traditional goodness-of-fit indices for bifactor models (e.g., Bonifay et al., 2017; Flores-Kanter et al., 2018; Morgan et al., 2015; Sellbom & Tellegen, 2019; Ventura-León et al., 2021). This methodological refinement underscores the researchers' commitment to advancing the field of instrument construction and validation with global best practices in psychometric analysis. Thus, developing the PPQ not only bridges a significant gap in the literature but also aligns with and extends the existing body of research on persistence in academic publishing. It draws on established research traditions and methodological approaches while introducing innovative elements that enhance our understanding of the complexities surrounding scholars' persistence to publishing their work in reputable journals, such as those in Scopus. Ultimately, this instrument can be used to support research that contributes to a broader understanding of the factors that drive and sustain academics' persistence in the face of publishing challenges.

Limitations and future research focus

The current study represents a significant step in instrument development for measuring persistence to publish. However, it is important to acknowledge its limitations and offer future research directions to further enhance the instrument's validity and applicability across diverse academic contexts. First, the findings may have limited generalisability since the study primarily focused on lecturers with experience in publishing in Scopus-indexed journals, which could restrict the applicability of the developed instrument to this specific academic population. Future research should consider extending the validation process to encompass a more diverse sample of academics from various disciplines and career stages.

Additionally, while content validity was assessed through expert ratings and feedback, the study did not explore other forms of validity evidence, such as criterion and predictive validity. To enhance the instrument's robustness, future research should consider evaluating the PPQ's validity in predicting actual publishing behaviour and outcomes, thereby establishing its predictive validity. This entails investigating the extent to which the PPQ can predict actual publication rates, submission frequencies, or the quality of publications in Scopus-indexed journals. Future research should assess the criterion validity of the PPQ by examining how closely the instrument developed in this study relates to other instrument measuring similar constructs. Furthermore, measurement invariance is a critical consideration that was not tested in the present study. Given potential cultural and contextual variations in the publishing process, future research should examine the measurement invariance of the PPQ across different groups to ensure its validity and comparability.

Although the study employed multiple reliability measures, it primarily focused on internal consistency measures. Future research could explore other aspects of reliability, such as test-retest reliability and inter-rater reliability. Test-retest reliability would assess the instrument's stability over time, while inter-rater reliability would examine consistency among different raters or observers, particularly in cases where multiple perspectives contribute to the assessment. Future research could employ alternative validation techniques, such as item response theory (IRT) or generalizability theory (G theory) on the PPQ.

Conclusion

This study has successfully developed and validated the Persistence to Publish Questionnaire (PPQ), an invaluable instrument for assessing academics' commitment to publishing in Scopus-indexed journals. The PPQ underwent a rigorous development process, including item generation, content validity assessment, pretesting, and pilot testing. Both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) corroborated the five-factor structure, with the bifactor model emerging as the best-fitting model for the PPQ. The instrument demonstrated robust internal consistency. Construct validity evidence strongly supported both convergent and discriminant validity, affirming that the PPQ effectively measures persistence to publish while distinguishing it from related constructs. Thus, the PPQ offers a powerful tool for researchers and institutions to assess academics' persistence in publishing in Scopus-indexed journals. This instrument can be used to identify individuals or departments with lower levels of persistence, enabling the development of targeted interventions and support mechanisms. Researchers can employ this instrument to explore the antecedents and consequences of publishing commitment, deepening our understanding of the academic publishing process.

This study, therefore, presents the PPQ as a standardised tool for evaluating academics' persistence to publishing in Scopus-indexed journals. The results of the study have significant implications for informing research policy, shaping institutional support, and guiding interventions to enhance researchers' persistence in publishing. Institutions can use the PPQ as an assessment tool to gauge the publishing persistence of their faculty members. This data can inform the development of interventions, such as workshops, seminars, or mentorship programmes, aimed at enhancing researchers' skills and resilience in the face of common publishing challenges.

The results of the study may guide institutional strategies for fostering a culture of continuous improvement in publishing skills. For instance, institutions and policymakers can use the PPQ to identify academics who may need additional support in manuscript preparation, manuscript submission, handling revisions, dealing with rejections, and publication delays. This information can be used to tailor policies that address these challenges, whether through targeted training programmes, mentorship initiatives, or the provision of resources to support academics in handling various stages of the publication process. Integrating these programmes into graduate training, faculty development initiatives, or ongoing professional development opportunities can contribute to a culture of continuous improvement in publishing skills.

Furthermore, the PPQ can be used by funding agencies to evaluate the persistence of researchers to publish in Scopus-indexed journals. This can help funding agencies to identify researchers who are committed to publishing and support their research activities. The PPQ can also be used by publishers to evaluate the persistence of authors to publish in their journals. This can help publishers to identify authors who are committed to publishing and provide them with necessary support to improve the quality of their manuscripts. The PPQ can also play a role in performance evaluation processes for researchers. Recognising and rewarding persistence in publishing can incentivise academics to invest more effort in this aspect of their work. Institutions may consider incorporating publishing persistence as one of the criteria for tenure and promotion decisions, thereby reinforcing the importance of sustained commitment to scholarly dissemination.

Given the standardised nature of the PPQ, it allows for potential benchmarking across institutions and on a global scale. Researchers and institutions can compare their scores to national or international averages, fostering healthy competition and collaboration. Collaborative efforts can be initiated to share best practices in addressing common challenges identified by the PPQ, promoting a collective approach to enhancing research productivity. Overall, the PPQ is a valuable tool for shaping interventions that address specific challenges faced by researchers, ultimately contributing to the advancement of knowledge dissemination and research quality within the academic community.

References

Andriani, D., Purwana, D., & Susita, D. (2020). Analysis of factors that effect lecturer productivity producing international scientific article in private university: Motivation as a moderating variable. *International Journal of Human Capital Management, 4*(1), 87-107. https://doi.org/10.21009/IJHCM.04.01.08

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78-117. https://doi.org/10.1177/0049124187016001004

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science, 5*(1), 184-186. https://doi.org/10.1177/2167702616657069

Clark, L. A., & Watson, D. (2016). Constructing validity: Basic issues in objective scale development. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 187–203). American Psychological Association. https://doi.org/10.1037/14805-012

Cleeton, G. V., & Knight, F. B. (1924). Validity of character judgments based on external criteria. *Journal of Applied Psychology, 8*, 215-231. https://doi.org/10.1037/h0072525

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis.* Lawrence Erlbaum Associates, Inc., Publishers. https://doi.org/10.4324/9781315827506

Constantin, T., Holman, A., & Hojbotă, M. A. (2011). Development and validation of a motivational persistence scale. *Psihologija, 45*(2), 99-120. https://doi.org/10.2298/PSI1202099C

De Luca, M. L., Grossi, G., Zaccarello, G., Greco, R., Tineri, M., Slavic, E., & Palummieri, A. (2016). Adaptation and validation of the "continuing bond scale" in an Italian context. An instrument for studying the persistence of the bond with the deceased in normal and abnormal grief. *International Journal of Psychoanalysis and Education, 2*, 37-52. https://bit.ly/3OgUIQa

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087. https://doi.org/10.1037/0022-3514.92.6.1087

Dueber, D. M. (2017). *Bifactor indices calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models.* https://dx.doi.org/10.13023/edp.tool.01

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412. https://doi.org/10.1111/bjop.12046

Ekpenyong, J. A., Owan, V. J., Ogar, J. O., & Undie, J. A. (2022). Hierarchical linear modelling of educational outcomes in secondary schools: What matters – teachers' or administrators' input? *Cogent Education, 9*(1), 2133491. https://doi.org/10.1080/2331186X.2022.2133491

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science, 3*(4), 484-501. https://doi.org/10.1177/2515245920951747

Flores-Kanter, P. E., Dominguez-Lara, S., Trógolo, M. A., & Medrano, L. A. (2018). Best practices in the use of bifactor models: Conceptual grounds, fit indices and complementary indicators. *Revista Evaluar (Evaluate Magazine), 18*(3), https://doi.org/10.35670/1667545.v18.n3.22221

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research, 18*(3), 382–388. https://doi.org/10.1177/002224378101800313

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Psychology Press. https://doi.org/10.4324/9780203781098

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*(4), 430-450. https://doi.org/10.1037/1082-989x.6.4.430

Hadi, H. J., Omar, M. A., Osman, W. R. S., Ibrahim, M. F., & Hussaini, M. (2020). Performing a content validity: Establishing a reliable instrument to measure the intention to adopt cloud computing software as a service in public organisation. *Journal of Theoretical and Applied Information Technology, 98*(22), 3531–3540. https://bit.ly/3rASySv

Hart, C. (2014). Development of a persistence scale for online education in nursing. *Nursing Education Perspectives, 35*(3), 150–156. https://doi.org/10.5480/12-993.1

Howard, M. C., & Crayne, M. P. (2019). Persistence: Defining the multidimensional construct and creating a measure. *Personality and Individual Differences, 139*, 77-89. https://doi.org/10.1016/j.paid.2018.11.005

Khindri, A., & Rangnekar, S. (2022). The trait of perseverance: A literature review and future research directions. *The American Journal of Psychology, 135*(3), 271-285. https://doi.org/10.5406/19398298.135.3.02

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press. https://bit.ly/3DC39zv

Kozlowski, M. B., & Fouad, N. A. (2022). Development and validation of the academic persistence outcome expectations scale. *Journal of Career Assessment, 31*(3), 555–570. https://doi.org/10.1177/10690727221126145

Lambovska, M. (2022). A fuzzy logic model for evaluating the motivation for high-quality publications: Evidence from a Bulgarian university. *Management: Journal of Contemporary Management Issues, 27*(2), 87-108. https://doi.org/10.30924/mjcmi.27.2.6

Lambovska, M., & Todorova, D. (2021). 'Publish and flourish' instead of 'publish or perish': A motivation model for top-quality publications. *Journal of Language and Education, 7*(1), 141-155. https://doi.org/10.17323/jle.2021.11522

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lockhart, M. E., Kwok, O. M., Yoon, M., & Wong, R. (2022). An important component to investigating STEM persistence: The development and validation of the science identity (SciID) scale. *International Journal of STEM Education, 9*(1), 34. https://doi.org/10.1186/s40594-022-00351-1

Lufi, D. (1979). *Development of scale for determining persistency in academic field* (Doctoral Dissertation), Washington State University. https://bit.ly/3XYLnPW

Lufi, D., & Cohen, A. (1987). A scale for measuring persistence in children. *Journal of Personality Assessment, 51*(2), 178-185. https://doi.org/10.1207/s15327752jpa5102_2

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412-433. https://doi.org/10.1037/met0000144

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence, 3*(1), 2-20. https://doi.org/10.3390/jintelligence3010002

Mukherjee, B. N. (1974). A questionnaire measure of persistence disposition. *Indian Journal of Psychology, 49*, 263-278. https://doi.org/10.1037/t52330-000

Owan, V. J., Akah, L. U., Ekpo, A. J., Ubi, I. O., Abeng, F. A., & Akah, G. T. (2022a). Socioeconomic factors and the evaluation of HIV/AIDS prevention programs: A psychometric analysis of an instrument. *Electronic Journal of General Medicine, 19*(6), em405. https://doi.org/10.29333/ejgm/12320

Owan, V. J., Ekpenyong, J. A., Chuktu, O., Asuquo, M. E., Ogar, J. O., Owan, M. V., & Okon, S. (2022b). Innate ability, health, motivation, and social capital as predictors of students' cognitive, affective and psychomotor learning outcomes in secondary schools. *Frontiers in Psychology, 13*, 1024017. https://doi.org/10.3389/fpsyg.2022.1024017

Owan, V. J., Odigwe, F. N., Okon, A. E., Duruamaku-dim, J. U., Ubi, I. O., Emanghe, E. E., Owan, M. V., & Bassey, B. A. (2022c). Contributions of placement, retraining and motivation to teachers' job commitment: Structural equation modelling of the linkages. *Heliyon, 8*(4), e09334. https://doi.org/10.1016/j.heliyon.2022.e09334

Owan, V. J., Owan, M. V., & Lata, N. (2022d). Discharge of pedagogic duties: A bootstrapped structural equation modelling of teachers' use of research materials in school libraries. *Journal of Applied Learning & Teaching, 5*(2), 116–131. https://doi.org/10.37074/jalt.2022.5.2.13

Porter, T., Molina, D. C., Blackwell, L., Roberts, S., Quirk, A., Duckworth, A. L., & Trzesniewski, K. (2020). Measuring mastery behaviours at scale: The Persistence, Effort, Resilience, and Challenge-Seeking (PERC) task. *Journal of Learning Analytics, 7*(1), 5-18. https://doi.org/10.18608/jla.2020.71.2

Quintana, I., Barriopedro, M. I., & Ruiz Perez, L. M. (2022). Achievement goals across persistence—validation of the Spanish version of the motivational persistence scale. *International Journal of Environmental Research and Public Health, 19*(14), 8474. https://doi.org/10.3390/ijerph19148474

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2015). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223-237. https://doi.org/10.1080/00223891.2015.1089249

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods, 25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338. https://doi.org/10.3200/JOER.99.6.323-338

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment, 31*(12), 1428–1441. https://doi.org/10.1037/pas0000623

Stucky, B. D. & Edelen, M. O. (2015). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 183-206). Routledge. https://psycnet.apa.org/record/2015-00285-009

Styk, W., & Klinkosz, W. (2020). The maze test: A computer tool for testing perseverance. *Psychology Research and Behavior Management, 2020*(13), 1277-1288. https://doi.org/10.2147/PRBM.S271558

Styk, W., Zmorzynski, S., & Samardakiewicz, M. (2023). Persistence is multi-trait: Persistence scale development and persistence perseveration and perfectionism questionnaire into polish translation. *Brain Sciences, 13*(6), 864. https://doi.org/10.3390/brainsci13060864

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Allyn and Bacon. https://bit.ly/3pWWIn6

Thalib, T., Hanafi, S. P., Aufar, F., Irbah, S., & Eduardus, J. S. (2019). The academic persistence scale. In *Proceeding of* *the 3rd International seminar on education* (pp. 122-130). Presented at the 3rd International Seminar on Education, November 07th, 2018. https://bit.ly/3K2Vo9d

Ventura-León, J., Quiroz-Burga, L., Caycho-Rodríguez, T., & Valencia, P. (2021). BifactorCalc: An online calculator for ancillary measures of bifactor models. *Revista Evaluar (Evaluate Magazine), 21*(3), 1-14. https://doi.org/10.35670/1667-4545.v21.n3.36272

Wang, C. K. A. (1932). A scale for measuring persistence. *Journal of Social Psychology, 3*, 79-89. https://doi.org/10.1080/00224545.1932.9919133

White, R. E., Prager, E. O., Schaefer, C., Kross, E., Duckworth, A. L., & Carlson, S. M. (2017). The "batman effect": Improving perseverance in young children. *Child Development, 88*(5), 1563-1571. https://doi.org/10.1111/cdev.12695

Williams, L. A., & DeSteno, D. (2008). Pride and perseverance: The motivational role of pride. *Journal of Personality and Social Psychology, 94*(6), 1007. https://doi.org/10.1037/0022-3514.94.6.1007

Wood, S. E. (1968). *Relationship of certain non-intellective factors to persistence.* St. Louis: Missouri State University. https://eric.ed.gov/?id=ED023374

Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Majd, H. A., Nikanfar, A., & Ghahramanian, A. (2014). Details of content validity and objectifying it in instrument development. *Nursing Practice Today, 1*(3), 163–171. https://npt.tums.ac.ir/index.php/npt/article/view/24

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Learning in a Disrupted Environment: Exploring higher education student resilience using the Dynamic Interactive Model of Resilience

Adeela Ahmed-Shafi[A]
Tristan Middleton[B]
Richard Millican[C]
Sian Templeton[D]
Jenny Hill[E]
Chris Jones[F]

[A] *School of Education & Humanities, University of Gloucestershire, Cheltenham*
[B] *School of Education & Humanities, University of Gloucestershire, Cheltenham*
[C] *School of Education & Humanities, University of Gloucestershire, Cheltenham*
[D] *School of Education & Humanities, University of Gloucestershire, Cheltenham*
[E] *Academic Development Unit, University of Gloucestershire, Cheltenham*
[F] *School of Education & Humanities, University of Gloucestershire, Cheltenham*

Keywords

Disruption;
Dynamic Interactive Model of Resilience (DIMoR);
higher education;
online learning;
pandemic;
resilience;
UK.

Correspondence

rmillican@glos.ac.uk [C]

Article Info

Received 30 May 2023
Received in revised form 31 July 2023
Accepted 1 August 2023
Available online 4 August 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.18

Abstract

The coronavirus pandemic brought unprecedented circumstances, providing insights into how systems (people, institutions and societies) cope during a disruption. This paper reports research undertaken at one university in the South West of England, which adopted a mixed-methods approach to investigate how students responded to and coped with the impact of Covid-19 disruption and what they perceived as influencing their resilience.

Data were gathered from 434 students (undergraduate and postgraduate) using an online survey. Twenty of these students were subsequently interviewed individually. Data analysis used the lens provided by the Dynamic Interactive Model of Resilience (DIMoR) to explore the complexity of resilience and how it is shaped and impacted by internal and surrounding environments for any given system.

The research revealed the value of DIMoR as a tool for analysis and highlighted the dynamic, interactive and multifaceted nature of resilience as something that is influenced by multiple other systems rather than being a static quality within a system. A range of impacting risk/protective factors and vulnerabilities/invulnerabilities were identified, which are not either/or but fluctuate and exist to a greater or lesser degree depending on context and influences. The research also showed the shifting nature of surrounding systems that can become more or less proximal and influential depending on circumstance. Additionally, the study provided insight into the overriding importance of proximal relationships and the role lecturers/tutors can play in helping students to access university support services. Wider implications of the findings are discussed in relation to university processes and practices.

Introduction

The year 2020 saw major disruptions to life across the globe. Unexpectedly, societies were confronted by a new virus, Covid-19, with the World Health Organization (WHO) declaring a pandemic on 11 March 2020 (WHO, 2020). The pandemic brought with it unprecedented circumstances as societies wrestled with rising death tolls and over-stretched hospitals and health services, with leaders having to decide how to respond to keep populations safe.

A common approach, adopted by the UK, was to impose various levels of restriction on social activity ranging from physical distancing, to wearing masks, to complete lockdown, whereby individuals were expected to stay at home and refrain from any form of in-person social contact beyond familial support 'bubbles' (UK Gov, 2020). The restrictions had major ramifications for the economy (UK Parliament, 2021) as businesses or activities that involved direct personal contact, such as restaurants, hotels and gyms, were forced to close.

Educational institutions, being social systems, were not exempt from restrictions and found themselves having to adapt to changing circumstances and, for a substantial period, rapidly move all teaching online (World Economic Forum, 2020). This situation had the potential to threaten resilience at all levels – individual and collective. Systems are in a constant state of flux and accustomed to change, evolution and adaptation, but the pace of change on this occasion was unparalleled and brought new risks, challenges and threats (e.g., Laborde et al., 2020).

Many educators and students found the change and its pace difficult and experienced a need for support; however, others seemed to manage and thrive despite the challenges. Why some staff and students thrived, and others did not, and what sources of support were accessed, is important to comprehend if we are to learn from this experience and create environments that allow systems to develop resilience and the ability to withstand future shocks and disruptions.

To investigate this issue, we explore the impact of the disruption caused by the pandemic on the resilience of a university in England. The research was designed in three parts focusing on i) the student body, ii) the staff and iii) the university as a whole system. This particular paper reports on the research with reference to student resilience.

The research is framed by, and analysed using, the Dynamic Interactive Model of Resilience (DIMoR) (Ahmed Shafi et al., 2020a, 2020b). This model views resilience not as a fixed, static, within-system quality but as something that changes according to context and circumstance as systems interact with others encountered. It considers protective/risk factors and vulnerabilities/invulnerabilities and acknowledges surrounding ecological systems. The overarching purpose of the research was to explore how and to what extent student resilience was impacted during the disruptions caused by the pandemic, what helped students to cope, and to discover lessons for future practice.

The paper reviews literature around the themes of disruption, change, resilience, relationships and pedagogy in higher education (HE) and overviews the structure of and rationale behind DIMoR. It presents and discusses key findings and lessons learned and ends with a consideration of ramifications for future practice.

Literature review

Study selection

To support a systematic approach to reviewing the literature, inclusive of a wide scope of literature whilst ensuring pertinent studies were identified for inclusion, a Boolean database search was conducted on 20 May 2021 using PsycINFO, ERIC and Web of Science. The search focused on HE student response to the Coronavirus pandemic, HE student resilience and their mental health and well-being (MHWB) (for search strategy, see Appendix A). A total of 2,017 articles were initially identified, which was reduced to 145 articles through duplicate filtering and title screening. Article abstract screening using inclusion and exclusion criteria (see Appendix B) left 50 articles. Full-text screening using the criteria outlined in Appendix 2 eliminated a further 25 articles, and an additional article was added as a result of a paper-based search, leaving 26 articles. These articles were selected due to their identification of features (such as risk/protective factors, internal/external resources etc.) identified as pertinent within the DIMOR discussed below.

Disruption, change and adversity

Whilst recognising the unusual situation caused by the pandemic, Camfield et al. (2021) point out that disruptions and setbacks are a regular occurrence for HE students and, as such, it is the responsibility of universities to help mitigate the potentially damaging impact on student mental health and well-being (MHWB). The link between living through a crisis and the subsequent increase in feelings of stress and insecurity is well-recognised across a range of cultures (Gonzalez-Ramirez et al., 2021; Quintiliani et al., 2021; Wen et al., 2021) and is identified by Browning et al. (2021) as being particularly acute amongst HE students. The swift move to online-only learning has been identified as affecting student MHWB, leading to feelings of vulnerability, reduced confidence, self-regulation difficulties and a subsequent detrimental impact on resilience and ability to engage cognitively with challenges (Camfield et al., 2021; Conrad et al., 2021; Gonzalez-Ramirez et al., 2021; Quintiliani et al., 2021). The sudden change in typical day-to-day routines altered students' ability to seek and access support from their course teams (Camfield et al., 2021; Hagedorn et al., 2021), resulting in a negative impact on feelings of security (Conrad et al., 2021; Copeland et al., 2021) alongside self-efficacy and agency (Bourion-Bédès et al., 2021; Camfield et al., 2021). Compounding these factors, the typical age of undergraduate students is suggested by Wen et al. (2021) to render them more developmentally vulnerable to mental health difficulties.

The literature shows that there are strategies and approaches that can help mitigate sudden disruptive change, such as: engendering a sense of belonging (Camfield et al., 2021), creation and maintenance of routines (Rodgers et al., 2020) and a proactive approach to putting supportive systems in place prior to any significant disruption (Hagedorn et al., 2021). These approaches (amongst others) can be usefully conceptualised within the construct of a resilience theoretical framework.

Resilience in the face of adversity – the Dynamic Interactive Model of Resilience (DIMoR)

Resilience has recently been understood as dynamic, emerging as a result of reciprocal interactions between systems whilst also being based on features within any given system, be that human, institutional or organisational (Ahmed Shafi et al., 2020). The importance of interpersonal relationships in helping students to manage change and adversity in a resilient way is well recognised (Conrad et al., 2021; Rodgers et al., 2020; Sun et al., 2020; Ye et al., 2020; Zhang et al., 2021). These aspects of resilience have been conceptualised within the Dynamic and Interactive Model of Resilience (DIMoR) (Ahmed Shafi et al., 2020a, 2020b) (see Figure 1 below). Within this model, reciprocal interactions are indicated not only on an interpersonal level but also within and between wider systems surrounding the individual, which is highlighted as key for developing resilience. Important within DIMoR is the concept of individual agency and its impact on interactions with other individuals and also within and between the systems in which these interactions take place, the reciprocity of all these interactions being fundamental. DIMoR echoes the perspective of Schlesselman et al. (2020), noting that individuals come from unique contexts and respond to stress and adversity in different ways. DIMoR recognises the dynamic and fluctuating nature of various factors that might influence resilience and highlights the role of protective factors in mitigating risk and adversity.

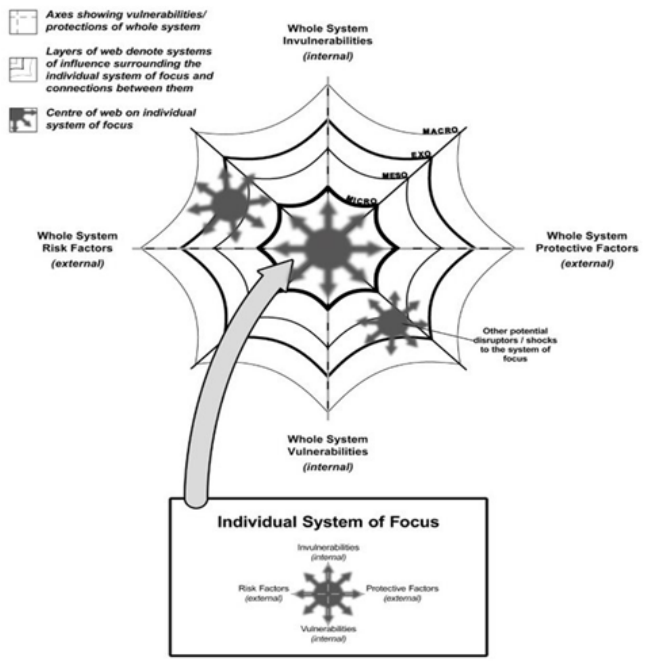


Figure 1. Using the DIMoR as a lens to help analyse optimum conditions for the emergence of resilience.

The DIMoR can help educators to interpret how a system, such as a university or its students, may respond to adversity by enabling us to 'see' the system as a dynamic, multiple and complex set of interactions of its different elements. For example, if Figure 1 was used to depict a university system around students, the web itself would represent the university to include its systems (e.g., cohort, course team, university support services, senior leadership team) and structures (e.g., policies, procedures, timetabling, online platforms etc). The risk-protective axes would refer to external risks such as the pandemic or finances or policy changes. Protective factors would refer to factors that act as mitigators to the risks, e.g., strong leadership, sound finances or robust policies and procedures. The vulnerabilities would include, for example, high staff turnover or low student retention, whereas invulnerabilities could be the university's strong identity or specialism. The students (and staff) are the orbs (or actors) within this system who both impact on it (e.g., staff illness) and are impacted by it due to being in the system. The individual system of focus, in this case, would be the students. Invulnerabilities might be robust physical and mental health and vulnerabilities existing health issues, whereas risk factors could be separation from family and friends and protective factors strong relationships, hobbies and exercise routines. Resilience is emergent and dependent on the range of dynamic interactions between the components of the system. As such, creating and fostering conditions for resilience to emerge is key to being resilient to adversities.

An example of the interaction between systems and the impact on individuals is portrayed by Rodgers et al. (2020), who argued that policy measures taken in light of Covid-19 became a barrier to accessing social support, resulting in students experiencing increased vulnerability in their self-regulation and ability to manage the changes that ensued. Quintiliani et al. (2021) also acknowledge the increased vulnerabilities of students as a result of the reduction in social support and suggest that the development of skills to support resilience can improve MHWB, thus acting as a protective factor against future challenges.

Recent research with a specific focus on HE students has established the importance of developing a range of protective factors to support the development of resilience (Holdsworth et al., 2018), including: support and intervention to promote positive thinking (Yang et al., 2020); social support (Zhang et al., 2021); clear and timely communication; a calm and safe learning environment (Holdsworth et al., 2018); experiencing a sense of community and contact with the course team (Hagedorn et al., 2021); adaptive coping strategies (Ye et al., 2020); exercise and hobbies (Bourion-Bédès et al., 2021) and opportunities to visit the outdoors (Browning et al., 2021). The importance of not only recognising these protective factors but also identifying potential risk factors is emphasised by Bourion-Bédès et al. (2021), suggesting that it is essential to develop targeted interventions and support based around a knowledge and understanding of protection and risks. Some of the possible risk factors identified for HE students include substance misuse (Bourion-Bédès et al., 2021) and other maladaptive coping strategies (Browning et al., 2021; Ye et al., 2020); demographic factors, such as being female

(Alemany-Arrebola et al., 2020; Bourion-Bédès et al., 2021) and in the 18-24-year-old age group (Browning et al., 2021); financial difficulties (Hagedorn et al., 2021); use of social media (Browning et al., 2021); and specifically in response to Covid-19, increased responsibilities within the home context (Wallace et al., 2021).

The themes above are not unique to our understanding of HE student resilience linked to the specific challenges presented by the Covid-19 pandemic. In a qualitative study conducted prior to the pandemic, Holdsworth et al. (2018) found that students perceived universities as having a role to play in nurturing their resilience, identifying the importance of relationships between peers and lecturers alongside a challenging and stimulating curriculum as key to developing student resilience. When considering these themes in light of HE student resilience more generally, Hagedorn et al. (2021) indicated that they often feature in the profile of students who drop out of HE. This finding adds substance to the need for a more proactive approach to supporting the development of student resilience to help mitigate future adversity.

Engagement with learning

One of the primary roles of a university is to provide opportunities for learning and development. However, there is an integral link between emotion and learning (Camfield et al., 2021) which has received comparatively little consideration within the HE literature (Gonzalez-Ramirez et al., 2021). The emotional distress caused by disruption and change, alongside a lack of opportunity for social interaction, has been found to lead to mental health disorders and difficulties in attending to learning (Copeland et al., 2021; Quintiliani et al., 2021). The relationship between skills in emotional regulation and both self-efficacy and engagement with learning is inexorable (Panayiotou et al., 2021), with a flexible, hopeful and optimistic mindset being key to coping and engaging (Browning et al., 2021; Copeland et al., 2021; Wallace et al., 2021; Wen et al., 2021; Yang et al., 2020; Ye et al., 2020). There is emerging evidence that first-year students are particularly vulnerable emotionally to the challenges brought about by disruption and change (Copeland et al., 2021; Gonzalez-Ramirez et al., 2021; Quintiliani et al., 2021). That said, engagement with learning grounded in positive relationships with course tutors and peers is protective for this student group and consequently has a positive impact on academic outcomes (Camfield et al., 2021). It is important to acknowledge that this need for belonging and connection is not unique to first-year students but a fundamental human need that permeates across the student population (Ye et al., 2020; Camfield et al., 2021), and there is evidence from Sun et al. (2021) that the movement to online studies as a result of the Covid-19 pandemic has had a negative impact on the sense of belonging amongst the student population.

Impact of online studies

In their exploration of the move to online study in response to Covid-19, Gonzalez-Ramirez et al. (2021) found that the sense of belonging to a student's university community and course team diminished significantly, negatively impacting their MHWB. This finding was echoed within the study by Quintiliani et al. (2021), who established that the perceived stress of learning online had a detrimental impact on students' engagement and completion of their studies. The speed of transition to online learning also appeared to have a disruptive influence on the relational aspects of learning, leading to feelings likened to grief and loss (Camfield et al., 2021; Wallace et al., 2021), particularly concerning the reduction in opportunities for spontaneous interactions and questions during class.

The rapid shift to online delivery caused additional pedagogical issues for both lecturers and students alike, with students experiencing significant difficulties in terms of pace and efficiency of learning (Camfield et al., 2021); self-regulation (Gonzalez-Ramirez et al., 2021); attention and concentration (Quintiliani et al., 2021; Wallace et al., 2021) and access to technology (Gonzalez-Ramirez et al., 2021; Wallace et al., 2021), which all had a deleterious impact on their motivation and attainment (Gonzalez-Ramirez et al., 2021; Wallace et al., 2021). In addition, Millican et al. (2023) explained that many HE lecturers were insufficiently prepared for online learning and course materials had not been designed for this mode of delivery. Lecturers also had to reconsider and possibly adapt their pedagogy – a strong appreciation of which is, as Chu et al. (2021) note, essential to optimising learning and engagement. These pedagogical difficulties were also noted by students to create additional barriers to learning (Wallace et al., 2021). However, Gonzalez-Ramirez et al. (2021) offer us a note of caution in drawing overly causal attributions from these findings, suggesting instead that individual student characteristics (e.g., age and gender), and their immediate surrounding systems (e.g., paid employment and familial responsibilities), had more of an impact on their access to learning rather than the mode of teaching delivery. The literature has also pointed to positives for some students as a result of the movement to online learning, such as an increase in creativity and problem-solving abilities (Wallace et al., 2021) and a flexibility to maintain a healthy work-life balance (Schlesselman et al., 2020; Wallace et al., 2021). Despite these positives for some members of the student population, Wallace et al. (2021) still found that online learning created additional stress, and although some stress is needed to support motivation, too much can have a detrimental impact on self-efficacy and attainment.

Conditions to support developing resilience

To help protect against the emotional response to challenges arising from disruption and change, some university-wide approaches and support systems have been explored within the literature. Copeland et al. (2021) found that those students who were already accessing MHWB services from the university appeared less emotionally impacted by Covid-19 and consequently found engagement with learning easier.

The need for more formal support systems to help students develop the psychological flexibility and development of regulation skills to respond to future challenges is advocated by Panayiotou et al. (2021). Alongside this, there is also recognition of the significant positive impact of nurturing relationships between lecturers and students and of peer-to-peer support (Holdsworth et al., 2019; Sun et al., 2020; Millican et al., 2023). A relational approach to teaching and learning appears to have the strongest protective impact, particularly in light of the finding from Browning et al. (2021) that students typically do not take full advantage of university MHWB services.

It would therefore seem that there are a number of protective and risk factors alongside individual vulnerabilities and invulnerabilities influencing HE student resilience in the face of significant disruption and change. Some of these factors are internal to the student such as their mindset, self-regulatory skills and self-efficacy, whilst others are more external, such as their relationships with the course team and peers, and wider university systems and communication. The DIMoR offers us a resilience-based theoretical lens from which to explore the various influences that might create conditions to support the emergence of resilience in times of shock and disruption.

The specific research objectives for this study were to:

(1) Explore what university students perceived as influencing their resilience as they responded to the disruptions caused by the Covid-19 pandemic;

(2) Consider how the identified influences help in understanding student resilience in the context of the pandemic;

(3) Use this understanding to identify ways to create conditions to support the emergence of resilience in times of shock and disruption.

Methodology

We adopted an interpretivist research approach (Burbules et al., 2015), analysing data using the DIMoR framework. 'Epistemological vigilance' (Bourdieu et al., cited in Guzman-Valenzuela, 2016) was maintained by the research team through repeated reference to our position and perspectives. We focussed on a single case study higher education institution, a post-1992 university in the South West of England with a student cohort of approximately 7,915, comprising a gender split of 59.9% female and 39.9% male. We used mixed-methods sequential design, where the quantitative and qualitative data hold equal status (Leech & Onwuegbuzie, 2009) and were collected in two phases (Teddlie & Tashakkori, 2006). In addition to triangulation (Biesta, 2017), this approach provided complementary and developmental data in which the quantitative data informed the qualitative (Mertens, 2015). At each stage of the research, collaboration took place between the team members to ensure consistency and rigour and to enhance insight (Ciuhan & Iliescu, 2020).

Phase 1 of the research consisted of an online survey of students generating quantitative and qualitative data. The survey design was shaped by the research objectives, findings from the literature review and the lens of DIMoR, with the aim of identifying key factors to be followed up in the Phase 2 interviews. The majority of the survey questions used a closed-ended format to ascertain the frequency of key factors. These questions were supplemented by some open-ended responses to capture any missing factors. The number and complexity of questions were minimised following piloting with two students who referred to 'online fatigue'. At the end of the survey, a request was made for volunteers to take part in a follow-up individual online interview (Phase 2).

The survey was communicated to the entire student body using a banner placed on the university student login webpage. This was supplemented by prompts from individual Course Leaders, from course administrators and from Postgraduate Research Leads, and by using our own student and staff networks. Responses were incentivised through a random draw to win three £20 shopping vouchers. In total, 434 survey responses were received in the last six weeks of 2021. Demographics of the survey respondents are identified in Tables 1a to 1e.

Table 1a. Survey respondent demographics: Response to "What is your gender?"

Option	Count	%
Male	85	19.6%
Female	332	76.7%
Other	7	1.6%
Prefer not to say	9	2.1%

Table 1b. Survey respondent demographics: Response to "Are you a UK student or an international student?"

Option	Count	%
UK	413	96.3%
International	16	3.7%

Table 1c. Survey respondent demographics: Response to "What level course are you on?"

Option	Count	%
Foundation Year	4	0.9%
1st Year Undergraduate (Level 4)	16	3.7%
2nd Year Undergraduate (Level 5)	98	22.6%
3rd Year Undergraduate (Level 6)	105	24.2%
Post-graduate Master's (Level 7)	74	17.1%
Post-graduate Doctorate (Level 8)	4	0.9%

Table 1d. Survey respondent demographics: Response to "What is your age range?"

Option	Count	%
18-21	229	52.5%
22-29	121	27.8%
30+	86	19.7%

Table 1e. Survey respondent demographics: Response to "How would you describe your ethnicity?"

Option	Count	%
White	400	92%
Asian	8	1.8%
Black	7	1.6%
Mixed race	12	2.8%
Other	8	1.8%

Following an initial review of the survey responses, interview questions and prompts were generated for the Phase 2 qualitative interviews. The aim of the interviews was to provide richer understanding of the areas identified within Phase 1. Piloting of the semi-structured interview schedule led to the adaptation of wording and prompts to ensure clarity and establish reliability across interviewers.

Via the online survey, 171 students volunteered for a follow-up interview. These students were sampled using a stratified demographic approach (Mertens, 2015) according to gender, ethnicity, age and level of study to reflect the University population as a whole and 20 participants were interviewed. The interviews were conducted by all members of the research team, using Microsoft Teams, and were recorded and transcribed.

The ethical approach was informed by BERA (2018) guidance, and ethical approval was provided through the researchers' University Research Ethics Panel. Further ethical concerns relating to student wellbeing were addressed through extensive signposting of University and wider support services at the end of the survey and interviews.

Data analysis

A case-oriented analysis approach (Onwuegbuzie et al., 2009) was used as a way of focusing on meanings using the lens of DIMoR. Quantitative data from the survey provided descriptive statistics, with a six-point Likert scale condensed to four responses for clarity of reporting (Table 2).

Table 2. Example of condensed Likert scale responses.

(4) Survey Likert scale options	Condensed responses
Helped a lot	Positive impact
Helped a bit	
Made no difference	Made no difference
Did not help much	Did not help
Made things worse	
Not applicable	Not applicable

The descriptive data were organised as column charts, representing the condensed responses. After an initial inspection of the data, patterns of interest were subject to inferential statistical analysis to test for the significance of apparent differences. Significance tests were conducted using non-parametric methods, namely a Kruskal-Wallis Rank Sum Test for an initial assessment of the significance of between-group differences and, wherever significance was found, this was followed by a Wilcoxon Rank Sum Test to locate significant pair-wise contrasts.

Analysis of the Phase 2 interview data used a constant comparative approach (Onwuegbuzie et al., 2011), with four stages of thematic analysis undertaken using NVivo to enable effective collaboration across the research team. The first stage of analysis consisted of early theme development (Braun & Clarke, 2021) through an inductive approach. In stage two, the team cross-checked and refined the codebook, ensuring reliability through intercoder agreement (O'Connor & Joffe, 2020). The refined codebook was then used to deductively code the qualitative responses from the Phase 1 survey as a way of triangulating with the rich interview data. Stage three refined the codebook into broad themes and subthemes. In stage four, the DIMoR framework (Figure 1) was adopted as a lens to analyse the combined coded data to identify resilience factors, including dynamic risk-protective factors and vulnerabilities/invulnerabilities.

Results

Survey data

These results are organised according to the DIMoR's lens of protective and risk factors, vulnerability and invulnerability, across macro, exo, meso and micro systems.

Protective factors

From the survey prompts, the majority of students selected friends and family as the factors that most helped them cope with day-to-day life during the pandemic (Figure 2). Faith/religion and government support were the factors chosen by the smallest number of respondents. Having more time, spending time alone and accessing social media prompted polarised views. Open-ended 'other' responses included having a job and undertaking University studies, which delivered focus and routine and afforded students a sense of self-worth. Spending time outdoors and/or exercising and adopting mindful activities also helped students to cope.

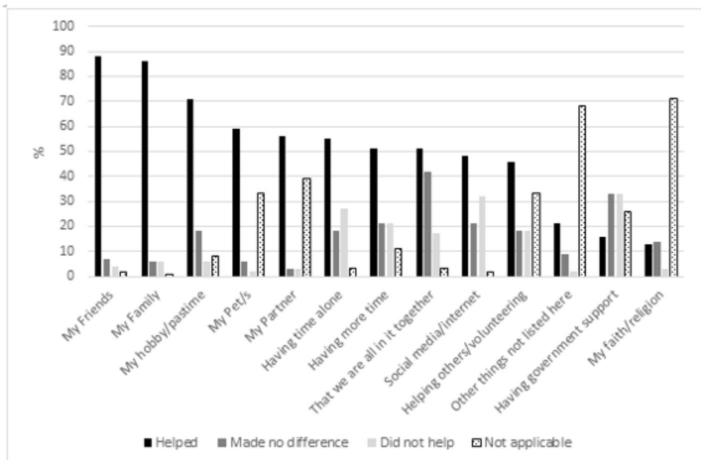


Figure 2. Responses to the question 'What things have helped you cope with your day-to-day life since the pandemic started?'

When prompted about who or what had helped them to get on with university study during the pandemic, most students selected lecturers and friends/family, followed by having face-to-face teaching when possible and contact with Personal Tutors and course mates (Figure 3).

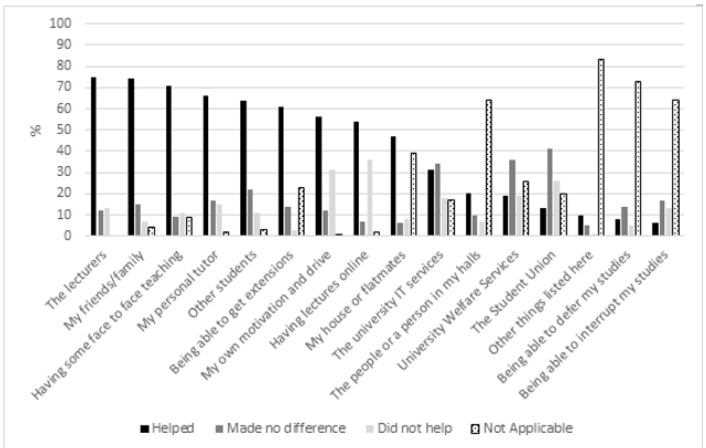


Figure 3. Responses to the question 'Thinking more specifically about University, who or what has helped you to get on with your studies during the pandemic?'

A Kruskal-Wallis Rank Sum Test revealed a statistically significant difference in response regarding having face-to-face teaching between students at different course levels ($p = 0.03$). Pairwise comparison, explored using a Wilcoxon Rank Sum Test, revealed that first-year students found access to face-to-face teaching significantly more helpful to their studies compared with second-year students ($p<0.05$). Motivation/drive was selected as supporting study by just over half the students, with Master's students more motivated compared with first-year ($p<0.001$), second-year ($p<0.0001$) and third-year students ($p<0.0001$). University processes and services were not generally identified as helpful beyond receiving extended time to complete assessments.

Students were also asked to respond to specific prompts about what had made studying easier during the pandemic. Only four of the 12 factors were rated by the students with any great frequency. Reduction in travel time to University was selected by 54% of students, and changes in the time available to do things by 36%. These factors made studying significantly easier for Master's students compared with first-year students ($p=0.006$). The opportunity for informal contact with lecturers (selected by 28% of students) and online learning (selected by 24%) also facilitated studying. From open-ended survey comments, access to campus facilities was also noted, including 24/7 library services, student information points, studios and laboratories.

When asked what had worked well with online learning, having the right device was a requirement to study effectively (selected by 67% of students). Having online lectures (65%) also worked well for students in supporting their learning, significantly more so for Master's students compared with first-year ($p=0.004$) and third-year ($p=0.004$) students.

Risk factors

From the survey prompts, almost all students selected not being able to see/talk to people as factors that had made their day-to-day living more difficult during the pandemic (Figure 4). Anxiety about the immediate future, lack of normality, and the need to constantly change plans were factors selected by over three-quarters of respondents. Inferential statistical analysis returned significant differences in the responses to these factors according to gender, with females expressing greater anxiety about them than males (p=0.02 or higher). Approximately half of the students expressed concern about finances impacting day-to-day life, often linked to getting or keeping a job, and there was also mention of personal health, looking after relatives, not being able to get hold of people and fear of the coronavirus.

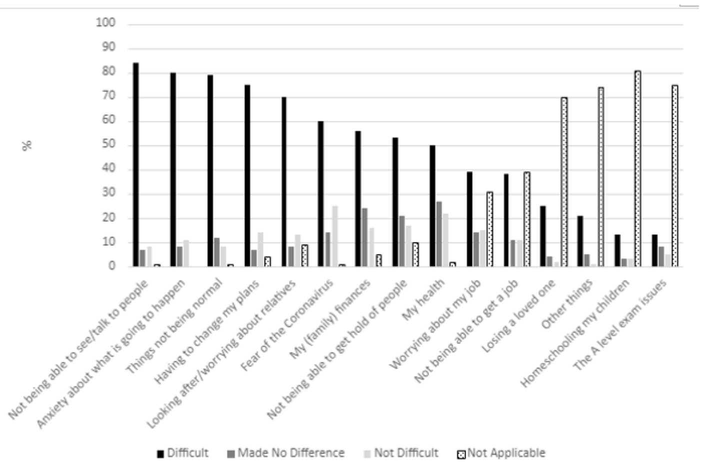


Figure 4. Responses to the question, 'What things have made your day-to-day living more difficult since the pandemic started?'

When students were asked to identify the most significant impact on their day-to-day life arising from the pandemic, the dominant theme that emerged once again was a lack of ability to socialise with friends and relatives and to give and receive embodied physical contact. Numerous students noted a breakdown in their routine and felt a lack of control, which led them to experience a loss of motivation, anxiety and poor mental health. Some students commented that the pandemic had amplified pre-existing anxieties. Such students linked these negative mental states with difficulty coping.

Government handling of the pandemic was chosen by almost three-quarters of students as negatively impacting their studies at university. Some referred to a lack of clarity in communication by the Government, whilst others reported a lack of trust in the ability of the Government to lead the country through the pandemic. News of global crises and events was selected by 68% of the students, with media reporting increasing anxiety and feelings of helplessness. The only other notable factor (selected by 37% of students) impacting negatively on studies was wider community issues, including family responsibilities such as caring and having to shield to protect the health, along with the difficulty of securing food from supermarkets in the early days of the pandemic.

When the students responded to prompts about what had made their studying harder as a result of the pandemic, the most common response was a lack of socialising opportunities (selected by 76% of students). The second most frequently selected factor was falling short in expectations of the overall university experience (selected by 69% of students). Juggling online study and home life was also chosen by 67% of students, but significantly less so for first-year students compared with second-years (p=0.04) and third-years (p=0.03).

Online learning was selected by 67% of students as a factor making their studies harder. Pairwise comparison revealed that first-year (p=0.01), second-year (p=0.03) and third-year students (p=0.00) felt more negatively affected by online learning in comparison to Master's students. Students commented about poor connectivity, slow internet speeds and lack of devices. They found working online difficult due to screen fatigue and feeling disengaged/demotivated as they worked from home or student accommodation, sometimes with distractions, unable to see the faces of their course mates who tended to turn off their cameras during sessions. A final factor, which half the students selected as making their studies harder during the pandemic, was the lack of opportunity for informal contact with lecturers.

Vulnerabilities/invulnerabilities

When things did not go well with their studies, students talked with their peers/housemates more than any other response (Figure 5). Contacting their lecturer or Personal Tutor came above speaking with a family member. Over half the students said they got anxious, and many said they felt low when things did not go well for them. Whilst a minority of students noted they had accessed University welfare services, cross-tabulation revealed these were also the students who had reached out to academic staff, peers and family members.

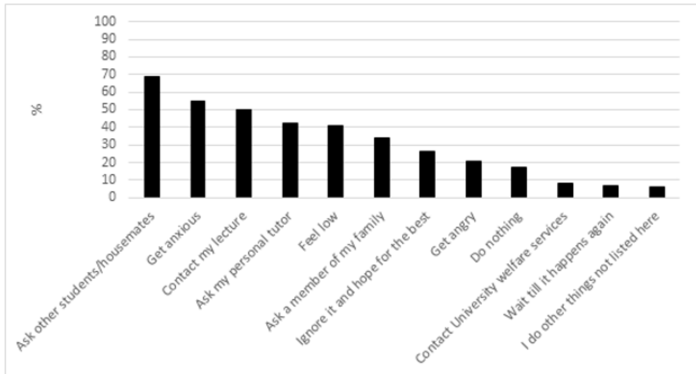


Figure 5. Responses to the question, 'When things did not go very well, what did you do?'

Students were positive about the future in the Autumn of 2020. Almost three-quarters of them believed things would be better for them in a year's time, with a further 21% saying things would be about the same and only 6% saying things would be a little or much worse. The responses from the Master's students were more positive, whilst those from the second years were the most negative of the undergraduate years. The optimism of the students was based on the hope of vaccination rollout, allowing restrictions to be relaxed and opportunities for socialising and attending lectures on campus to return. For final year students there was hope of graduation and the opportunity to gain employment.

Interview and qualitative survey data

Inductive analysis of the interview data and the qualitative comments from the survey yielded 15 main themes. Table 3 lists these themes, with exemplar quotes drawn from the data. DIMoR's contextual systems (based on the Bronfenbrenner ecosystem) were used to group the themes in terms of how proximal and distal they were to the individual.

Table 3. Exemplified themes drawn from the qualitative data.

Main themes	Quotes from data	DIMoR system
Relationships	'I missed seeing the people that I care about and interacting with them in a human way. The feelings of loneliness I felt through the pandemic were really difficult to deal with' 'You can still feel Covid kind of just tearing it apart slowly because you can't meet face-to-face'	(Micro system)
Reaction to change and uncertainty	'The most difficult thing is that nothing is normal. I can't go and do the things I would normally do or see the people I want to see. I feel incredibly anxious, like I've absorbed everybody's fears' 'I feel slightly more vulnerable. There's a level of me that's a creature of habit. And having things thrown up in the air like that, half way through' 'I have had to sort of cling on to the things that are the same, or as close as they can be, um, as markers sort of in my week, as it were'	
Health (mental and physical)	'Because of the stress of the corona virus my depression deepened and I developed suicidal thoughts'	
Hope and optimism	'I am excited for … my studies over the next year. I'm looking forward to getting back in face to face.' 'You have to be optimistic that things will be better. Life will improve and I cannot wait to see what the world has in store next for me'	
Exercise and hobbies	'When gyms and swimming pools and things like that were open, I was using those a lot because it was helping my mental health'	
Motivation	'Online teaching doesn't compare. I lose motivation easily when sat in one room all day. Staring at a screen for six hours a day is difficult mentally and physically' 'And I know that to an extent, as I say, it's self-led and by third year, you should be able to motivate yourself, but the reality of it is there are some people that don't' 'I think that, like personal motivation, has like, it's definitely been a real difficult area for me like. …. I'm not very good at coping without like, variety kind of thing'	
Access to resources/ facilities	'I feel that higher education students were kind of forgotten in terms of having laptops and iPads and things to be able to do their studies. ... It was just assumed that as a student you had access to these things'	(Meso system)
Learning online	Online lectures has meant I can manage my time more effectively and complete tasks at home around lectures' 'Trying to remain focused and have quiet time at home to study in a busy household whilst managing my children is quite hard'	
University policy and communications	'I think the way that got announced [no detriment policy] affected quite a lot … because it came out … like 6:00 o'clock on a Friday or Thursday. So suddenly we had all this information thrown onto us and nobody could reply or answer to us'	
Boundaries and routine	'I spend too much time in the same space where I rest and now do work. It blurs the line between them and makes it harder to relax or be productive'	
University support services	'The welfare services were able to help with having the correct equipment at home … that could help with my disabilities'	
Government policy and communication	'Unclear, unhelpful and non-committal government guidance has made the situation drastically worse than it needed to have been'	
Media	'I'm certainly conscious that a large proportion of society thinks we're [students] to blame to an extent'	(Exo and macro systems)
Work and career opportunities	'I know that not only now is affected, but my future work will be affected'	
Finances	'Losing my job but … I'm still struggling and can't focus on my studies'	

Further analysis using the DIMoR

The DIMoR was used to enable a deeper analysis of the main themes from both the survey and interview data. This revealed that many factors were viewed as protection or risks and vulnerabilities or invulnerabilities depending on the individual context, as illustrated in Table 4. DIMoR helps illustrate the complexity of the themes and how they can shift according to time and circumstance. It should be noted that not all themes had data supporting this dynamic interactive nature.

Table 4. The main themes organised as risk and protective factors.

As a risk	Factor	As protection
Poor, weak, or lack of	⟵ Relationships ⟶	Strong, supportive
Lack of	⟵ Exercise & hobbies ⟶	Participation in
Insecure	⟵ Finances ⟶	Secure
Lack of access to facilities, resources or personnel	⟵ Access to resources ⟶	Good access to facilities, resources and personnel
Technical difficulties, screen time, inappropriate technology	⟵ Learning online ⟶	Appropriate technology, well-staged, effective pedagogy
Poor, lack of, lack of clarity	⟵ Government & University ⟶ policies and communication	Timely, effective
Blurred, difficult to separate, lack of, disrupted	⟵ Boundaries & routines ⟶	Clear, distinct time and space delineation, consistent
Reduced, unavailable, or inefficient	⟵ University support services ⟶	Sufficient, available, effective
Reduced, unavailable, or perceived as so	⟵ Work/career opportunities ⟶	Available, or perceived as so
As a vulnerability	**Vulnerabilities/invulnerabilities**	**As an invulnerability**
Poor adaptability, reluctant to change	⟵ Reaction to change ⟶	Adaptability, acceptance, action orientation
Poor	⟵ Health: mental & physical ⟶	Robust
Lack of, pessimistic, sense of loss, negative mindset	⟵ Hope ⟶	Present, optimistic, forward looking, positive mindset
Weak	⟵ Motivation ⟶	Strong

The DIMoR framework reveals how each theme, depending on its influence and conditions over time, can move across the risk and protective spectrum, thereby illustrating the dynamic and interactive nature of resilience. Taking financial issues as an example, our data showed that this could be both a risk factor if the individual or their close one lost their job or business, but equally, the surveyed students indicated that it could be a protective factor, for example, when the government furlough scheme was introduced designed to support employees in helping them cope during the pandemic:

> I'm on furlough at the moment, which is really useful – to not have to work but getting paid kind of thing, as then I can concentrate more on my actual uni work.

Similarly, relationships with those in the microsystem (family, partner, children) were presented as a source of support, but, at the same time, the intensity of interactions under lockdown conditions meant that it could become a risk factor. This could then be connected to the vulnerabilities of the individuals, for example, if they reported difficulty in managing change or a disruption to routine. The quote below illustrates how the multiple responsibilities of family and work, which may normally have been managed, were disrupted by the pandemic and impacted academic study:

> Juggling University assignments and caring for my mum and my brother, and trying to maintain what social life you can have during Covid, it's a lot. And it's a lot to the point that I have had to get extensions on every one of my assignments this year.

This reflects how many students found the opportunity for self-certified extensions especially helpful.

The DIMoR also demonstrated how factors that may traditionally have been positioned in individuals' exosystems or macrosystems generated influence not just distally but proximally. For example, one of the main themes in Table 3 refers to the broader issues created by the government (often situated in the exo/macrosystem in Western democratic contexts), which have imposed restrictions on peoples' social lives, employment and their day-to-day living. This demonstrates the dynamic nature of the web-like (proximal and distal) systems, where their boundaries become more porous and their influences bleed across levels during periods of disruption. The survey data demonstrated how students generally responded negatively to this increased proximity of government policy, indicating that government decisions and the news impacted on their daily life:

> Whenever we hear that Boris Johnson is making some sort of announcement, we all just collectively groan because we're like, 'God, what is it now'. It's like living in a constant state of dread.

Furthermore, the sudden transition to online learning presented a clear risk to many students but for some students and at certain times, online learning was considered positive. Online learning not only posed challenges and opened up a range of perceived risks for students, especially when all delivery was online but it was also considered practical and flexible given, for example, childcare or other commitments.

This further layer of analysis allows a more granular understanding of the impact of Covid-19 disruption on students and what supported or hindered them. Such analysis can help ascertain additional support that could be offered to develop the resilience of students during disruption and to help develop resilient conditions for a more optimal learning environment.

In summary, the qualitative and quantitative data combined illustrate how participants' assessment of risk or protective factors was complex. Using DIMoR as an overarching lens ensured that the analysis of resilience recognised the influence of these external factors in an inter- and intra-connected way.

Discussion

The qualitative data support the findings of the quantitative data and offer a richness of experience through the interview excerpts. All data sources confirmed much of the extant literature (largely from the US and quantitative in nature) (see Table 5 for those themes where this is the case). The strongest protective factors identified by students were proximal, falling within the students' microsystem, reflecting the importance of close connections in supporting positive mental health and facilitating the continuance of study. This echoes the findings of Sun et al. (2020) around negating symptoms of depression. The students attested to gaining less support from more remote connections within their exosystem.

Table 5. A summary of key themes which confirm extant literature.

Theme	Key Literature
Relationships	Camfield et al. (2021); Quintiliani et al. (2021); Zhang et al. (2021); Hagedorn et al. (2021); Ye et al. (2020); Holdsworth et al. (2019); Sun et al. (2020); Conrad et al. (2021); Rodgers et al. (2020)
Pragmatics of studying	Gonzalez-Ramirez et al. (2021); Wallace et al. (2021); Chu et al. (2021)
Health (Mental and Physical)	Camfield et al. (2021); Bourion-Bedes et al. (2021); Copeland et al. (2021); Gonzalez-Ramirez et al. (2021); Quintiliani et al. (2021); Wallace et al. (2021); Conrad et al. (2021); Rodgers et al. (2020); Panayiotou et al. (2021)
Online learning	Camfield et al. (2021); Conrad et al. (2021); Gonzalez-Ramirez et al. (2021); Quintiliani et al. (2021); Wallace et al. (2021); Chu et al. (2021)
Change and uncertainty	Copeland et al. (2021); Quintiliani et al. (2021);
Policy and communications: government and university	Rodgers et al. (2020); Holdsworth et al. (2018);
Hope and optimism	Browning et al. (2021); Copeland et al. (2021); Wallace et al. (2021); Yang et al. (2020); Ye et al. (2020)
Boundaries and routines	Rodgers et al. (2020); Wallace et al. (2021); Schlesselman et al. (2020); Camfield et al. (2021); Hagedorn et al. (2021)
Exercise and hobbies	Bourion-Bedes et al. (2021); Browning et al. (2021)
Finances	Hagedorn (2021)
Motivation	Camfield et al. (2021); Gonzalez-Ramirez et al. (2021); Wallace et al. (2021)
University support structures	Camfield et al. (2021); Hagedorn et al. (2021); Copeland et al. (2021); Panayiotou et al. (2021); Yang et al. (2020)

Having more time and putting this time to use through hobbies, relaxation and prevailing (online or immediate physical) social networks were positive for some students, reinforcing the protective factors identified by Bourion-Bédès et al. (2021). For other students, however, free time was detrimental as they felt more isolated and became anxious. These tended to be the same students who attested to getting anxious or feeling low when things did not go well for them. This cyclical relationship between mental ill-health and isolation, alongside existing mental health difficulties, was also identified by Browning et al. (2021) and adds additional credence to the importance of understanding existing risk factors in order to put protective factors in place.

In relation to studying, our findings build on the recommendations from Camfield et al. (2021) around the need for 'empathetic responsiveness' from academics in order to provide a flexible approach to meeting students' learning needs as a protective factor of provision. This is particularly so for first-year undergraduates who preferred face-to-face learning on campus with access to facilities and the ability to discuss issues with peers. A particular and new finding was the extent to which Master's students maintained motivation and appreciated online delivery more so than undergraduate students. This may reflect the notion that age is a factor in resilience (Wen et al., 2021), given that Master's students' ages are proportionately higher than those of undergraduate students (see Universities UK, 2019). This may also demonstrate how the flexibility of online learning was conducive to the additional responsibilities that older students tended to have. Findings from this study and that of Bourion-Bédès et al. (2021) highlight the particular impact on females.

Distal factors in the students' macrosystem, such as government and media handling of the pandemic, became more proximal, distracting students from their studies and increased their mental health issues. Juggling online study and home life, and learning via a screen for hours each day without social learning opportunities, demotivated students (particularly undergraduates) and generated a sense of loss as their experiences of university fell short of their expectations. Overall, in common with conclusions by Gonzalez-Ramirez et al. (2021), this loss generated disengagement and a negative impact on motivation. However, the data also show that these risk factors could also be protective, depending on the individual and their own conditions and circumstances.

Our findings support the argument of Gonzalez-Ramirez et al. (2021) that learning is disrupted by local, national or global crises, which can cause personal and academic impact at a range of levels. However, our findings extend the work of, for example, Rodgers et al. (2020), highlighting that the extent of the impact of these factors depends on a range of things, including how proximal or distal they are and how the factors interact with each other and the individual. Of particular importance is the support provided by family and from others with whom students have a close relationship (agreeing with Conrad et al. 2021). In addition, our findings demonstrate how relationships and their intensity during periods of lockdown can exacerbate the inability to cope, and this is particularly so for those with additional caring responsibilities, such as for some Master's students. At the same time, Master's students overall coped better with online provision whilst students at other levels did less so. Though other researchers have highlighted the vulnerability of first year students (Copeland et al., 2021; Gonzalez-Ramirez et al., 2021; Quintiliani et al., 2021), no other studies have found specific differences between levels (e.g. Camfield et al., 2021).

Also novel, and not reflected in the literature, is the extent to which students expressed concern about work and career prospects in light of the risks around the uncertainty caused by the pandemic. Previous literature, such as the more general findings from Holdsworth et al. (2018), indicating the protection of a positive link between the resilience of university students and successful transition from university to workplace, points to this in a much more general way. However, this research has a more nuanced perspective, illustrating the potentially far-reaching consequences that students perceived the pandemic to have.

The added layer of DIMoR as an analytical framework enabled us to see how resilience, in a situation such as that created by the pandemic, is shaped by a wide range of fluctuating and dynamic factors. These interact with the various systems that individuals are situated within, thereby building on the perspective of Schlesselman et al. (2020) that unique contexts impact how individuals respond. This further builds on the point made by Gonzalez-Ramirez et al. (2021) that the extent of the impact of online learning (whether positive or negative) was very much dependent on both individual and broader systems, such as whether students had caring responsibilities, challenging living arrangements, mental health difficulties and so on. This means that we are able to take a more nuanced approach to understanding the impact of the swift move to online learning.

In summary, our results and analysis using DIMoR present a complex picture of student perceptions of their resilience in the context of Covid-19. This enables a deeper understanding of what a higher education sector could do to best support its students during times of disruption or adversity. The findings build upon research on the resilience of HE students and reinforce the need for universities to take a more proactive role in student support during times of major disruption. The data show that students in our study were not very likely to seek support from the university, and where they did, it was where they already had close, supportive relationships with course teams. This reinforces and extends the point made by Hagedorn et al. (2021), who emphasised the vulnerability of the already vulnerable student. Furthermore, students tended to use the university's more 'automated' services, such as uncertified extensions for assignments, which could be organised online without the need for staff contact. Our research suggests that the proximal support provided by personal tutors or lecturers also acts as a lever to access the wider university support services.

Implications for practice in HE

The DIMoR has practical utility in helping stakeholders to understand how institutions such as universities can play a role in fostering conditions to support the resilience of students and how this needs to be deliberate, targeted and granular enough to respond to the diversity of the student body and the broader systems in which the institutions are situated. In doing so, there is a great opportunity for universities (as a proximal microsystem for students) to do more than be the provider of learning opportunities and to become a key place which helps develop resilient individuals and hence resilient communities and societies.

Importantly, this UK study demonstrates that the relationship between risk and protective factors is not binary, and it is important to consider both the nature of the learner and the system in which they are situated in order to understand and then develop the appropriate conditions within which resilience can emerge. Significantly, this research demonstrates the importance of proximal relationships that create a sense of belonging and provide the gateway to accessing wider (university) support systems to better cope during times of disruption so that students can not only survive but thrive in periods of disruption. Focusing on fostering these 'gateways' to wider university services is perhaps a key recommendation for universities.

Placing the results that have emerged from this research onto DIMoR (Figure 6) serves as a reminder of the complexities and fragility of resilience and how it depends on multiple interacting factors.

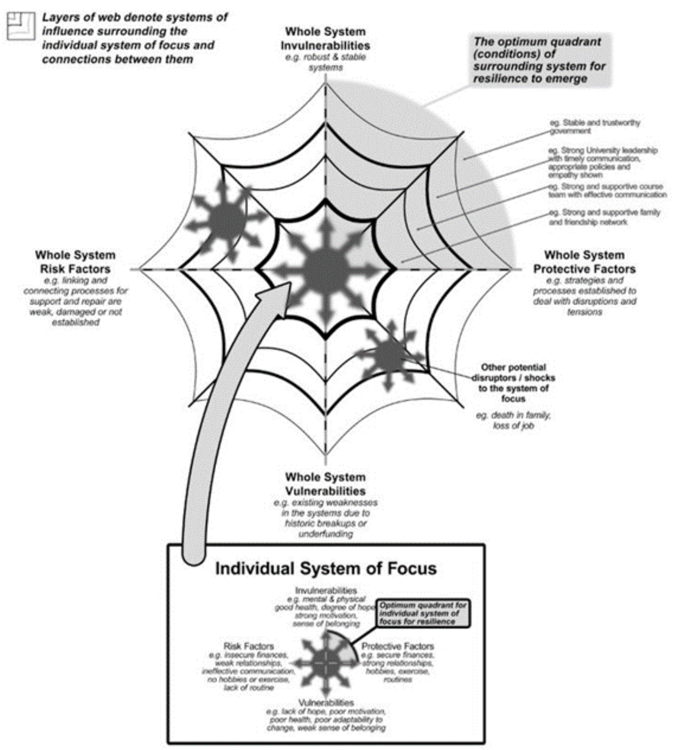


Figure 6. Using the DIMoR as a lens to help analyse optimum conditions for the emergence of resilience in students at times of disruption shows the interplay between the system of focus and surrounding systems.

It illustrates the need to look not only at the students themselves and their individual vulnerabilities and protections, but also at the system in which they sit and its own vulnerabilities and protections. In addition, it reinforces the role that various surrounding influences may play.

If universities want to be environments in which the resilience of students will be supported, then Figure 6 helps to define the conditions that they need to create. We have demonstrated that many system factors can be experienced as both risk and protective by students. As such, institutional managers need to critically examine practices that present risk and identify how they can be adapted to be more protective. This means making interventions that promote student capabilities and impact positively on the wider institutional environment. A culture that nurtures the invulnerabilities of students and provides structures that, for example, support the development of robust mental and physical health and help to develop student independence and action orientation. Universities must also nurture a sense of hope for the future, alongside providing protection against risk by, for example, ensuring that financial support and advice is readily available, and communication is timely and effective. The DIMoR highlights the protective nature of relationships, revealing the importance for universities and course teams to facilitate the development of strong bonds between students, but also between tutors and support staff and students so they can effectively guide students to the support service they need. What is clear is that the simple availability of support services is not going to lead to students accessing them automatically, but they are more likely to do so through their proximal support network. In the case of universities, that is likely to be the lecturers or personal tutors.

Conclusion

Whilst our research was conducted as a case study in a single university, there are nevertheless three main findings that are worthy of wider consideration, particularly given that this is one of the few qualitative studies in this area. These are:

(1) Times of disruption, such as that caused by a pandemic, can affect students' resilience which can then have a detrimental impact on their ability to study;

(2) Factors caused by disruption will not be experienced equally by everyone and can fluctuate depending on individual and context, between protective/risk, vulnerability/invulnerability and proximal/distal. However, there are concrete steps that universities can take to help support student resilience;

(3) The DIMoR model is a useful framework for analysis, enabling a holistic view acknowledging the interactive, dynamic and contextual nature of resilience and the role of individual agency.

Given the uncertainty of our future and the possibility of further shocks and disruptions to the Higher Education system, these findings might prove useful when considering future university culture and making budgetary and policy decisions.

References

Ahmed Shafi, A., Templeton, S., Middleton, T., Millican, R., Vare, P., Pritchard, R., & Hatley, J. (2020a). Towards a dynamic interactive model of resilience (DIMoR) for education and learning contexts. *Emotional and Behavioural Difficulties, 25*(2), 183-198.

Ahmed Shafi, A., Middleton, T., Millican, R., & Templeton, S. (Eds., 2020b). *Reconsidering resilience in education: An exploration using the dynamic interactive model of resilience*. Springer.

Alemany-Arrebola, L., Rojas-Ruiz, G., Granda-Vera, J., & Mingorance-Estrada, A.C. (2020). Influence of Covid-19 on the perception of academic self-efficacy, state anxiety, and trait anxiety in college students. *Frontiers in psychology 11*, 570017. https://doi: 10.3389/fpsyg.2020.570017

BERA. (2018). Ethical guidelines for educational research. *British Educational Research Association*. https://www.bera.ac.uk/wp-content/uploads/2018/06/BERA-Ethical-Guidelines-for-Educational-Research_4thEdn_2018.pdf?noredirect=1

Biesta, G. (2017). Mixing methods in educational research. In R. Coe, M. Waring, L. Hedges, & J. Arthur. (Eds.), *Research methods and methodologies in education* (2nd ed., pp. 159-165). Sage.

Bourion-Bédès, S., Tarquinio, C., Batt, M., Tarquinio, P., Lebreuilly, R., Sorsana, C., Legrand, K., Rousseau, H., & Baumann, C. (2021). Stress and associated factors among French university students under the covid-19 lockdown: The results of the pims-cov 19 study. *Journal of Affective Disorders, 283*, 108–114. https://doi: 10.1016/j.jad.2021.01.041

Braun, V., & Clarke. V. (2021). Conceptual and design thinking for thematic analysis. *Qualitative Psychology, 196*. https://doi.org/10.1037/qup0000196

Browning, M. H. E. M., Larson, L. R., Sharaievska, I., Rigolon, A., McAnirlin, O., Mullenbach, L., Cloutier, S., Vu, T. M., Thomsen, J., Reigner, N., Metcalf, E. C., D'Antonio, A., Helbich, M., Bratman, G. N., & Alvarez, H. O. (2021). Psychological impacts from covid-19 among university students: Risk factors across seven states in the United States. *PloS one, 16*(1), 0245327. https://doi: 10.1371/journal.pone.0245327

Burbules, N. C., Bridges, D., Griffiths, M., & Smeyers, P. (2015). Varieties of interpretation in educational research: How we frame the project. In P. Smeyers, D. Bridges, N. C. Burbules, & M. Griffiths. (Eds.), *International handbook of interpretation in educational research* (pp. 3-16). Springer.

Camfield, E. K., Schiller, N. R., & Land, K. M. (2021). Nipped in the bud: Covid-19 reveals the malleability of stem student self-efficacy. *CBE Life Sciences Education, 20*(2), 25. https://doi: 10.1187/cbe.20-09-0206

Chu, A. M. Y., Liu, C. K. W., So, M. K. P., & Lam, B. S. Y. (2021). Factors for sustainable online learning in higher education during the Covid-19 pandemic. *Sustainability, 13*(9), 5038. https://doi.org/ 10.3390/su13095038

Ciuhan, G. C. & Iliescu, D. (2020). Learning from a failed mixed methods child art-therapy research project. *International Journal of Qualitative Methods, 19*, 1-8.

Conrad, R. C., Hahm, H., Koire, A., Pinder-Amaker, S., & Liu, C. (2021). College student mental health risks during the Covid-19 pandemic: Implications of campus relocation. *Journal of Psychiatric Research, 136*, 117–126. https://doi: 10.1016/j.jpsychires.2021.01.054

Copeland, W. E., McGinnis, E., Bai, Y., Adams, Z., Nardone, H., Devadanam, V., Rettew, J., & Hudziak, J. (2021). Impact of Covid-19 pandemic on college student mental health and wellness. *Journal of the American Academy of Child & Adolescent Psychiatry, 60*(1), 134–141. https://doi: 10.1016/j.jaac.2020.08.466

Guzman-Valenzuela, C. (2016). Connecting theory and practice in qualitative research. In J. Huisman, & M. Tight. (Eds.), *Theory and method in higher education research* (Vol. 2, pp. 115-133). Emerald Group Publishing Limited.

Gonzalez-Ramirez, J., Mulqueen, K., Zealand, R., Silverstein, S., Mulqueen, C., & Bushell, S. (2021). Emergency online learning: College students' perceptions during the Covid-19 pandemic. *College Student Journal, 55*(1), 29–46.

Hagedorn, R. L., Wattick, R. A., & Olfert, M.D. (2021). 'My entire world stopped': College students' psychosocial and academic frustrations during the Covid-19 pandemic. *Applied Research in Quality of Life*, 1–22. https://doi: 10.1007/s11482-021-09948-0

Harari, D., Keep, M., & Brien, P. (2021). Coronavirus: Economic impact. *House of Commons Library*. https://commonslibrary.parliament.uk/research-briefings/cbp-8866/

Holdsworth, S., Turner, M., & Scott-Young, C.M. (2018). ... Not drowning, waving. Resilience and university: A student perspective. *Studies in Higher Education, 43*(11), 1837-1853. https://doi: 10.1080/03075079.2017.1284193

Laborde, D., Martin, W., Swinnen, J., & Vos, R. (2020). COVID-19 risks to global food security. *Science, 369*(6503), 500-502. https://doi:10.1126/science.abc4765.

Leech, N. L., & Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Quality & Quantity, 43*(2), 265-275.

Li, C., & Lalani, F. (2020). The COVID-19 pandemic has changed education forever. *World Economic Forum*. https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/

Mertens, D. M. (2015). *Research and evaluation in education and psychology* (4th ed.). Sage.

Millican, R., Templeton, S. & Hill, J. (2023) Exploring the impact of disruption on university staff using the dynamic interactive model of resilience. *Journal of Applied Learning & Teaching, 6*(1), 112-124. https://doi.org/10.37074/jalt.2023.6.1.19

O'Connor, C. & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Method, 19*, 1-13.

Onwuegbuzie, A. J., Burke Johnson, R., & Collins, K. M. T. (2009). Call for mixed analysis: A philosophical framework for combining qualitative and quantitative approaches. *International Journal of Multiple Research Approaches, 3*(2), 114-139.

Onwuegbuzie, A. J., Leech, N. L., & Collins, K. M. T. (2011). Toward a new era for conducting mixed analyses: The role of quantitative dominant and qualitative dominant crossover mixed analyses. In M. Williams, & W. Vogt. (Eds.), *The sage handbook of innovation in social research methods* (pp. 353-384). Sage.

Panayiotou, G., Panteli, M., & Leonidou, C. (2021). Coping with the invisible enemy: The role of emotion regulation and awareness in quality of life during the Covid-19 pandemic. *Journal of Contextual Behavioral Science, 19*, 17–27. https://doi: 10.1016/j.jcbs.2020.11.002

Quintiliani, L., Sisto, A., Vicinanza, F., Curcio, G., & Tambone, V. (2021). Resilience and psychological impact on Italian university students during Covid-19 pandemic. *Distance Learning and Health. Psychology, Health & Medicine, 1*(12), 1–12. https://doi: 10.1080/13548506.2021.1891266

Rodgers, R. F., Lombardo, C., Cerolini, S., Franko, D. L., Omori, M., Fuller-Tyszkiewicz, M., Linardon, J., Courtet, P., & Guillaume, S. (2020). The impact of the Covid-19 pandemic on eating disorder risk and symptoms. *International Journal of Eating Disorders, 53*(7), 1166–1170. https://doi: 10.1002/eat.23318

Schlesselman, L. S., Cain, J., & DiVall, M. (2020). Improving and restoring the well-being and resilience of pharmacy students during a pandemic. *American Journal of Pharmaceutical Education, 84*(6), 677-682. https://doi: 10.5688/ajpe8144

Sun, Y., Lin, S. Y., & Chung, K. K. H. (2020). University students' perceived peer support and experienced depressive symptoms during the Covid-19 pandemic: The mediating role of emotional well-being. *International Journal of Environmental Research and Public Health, 17*(24), 9308. https://doi.org/10.3390/ijerph17249308

Teddlie, C., & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools, 13*(1), 12-28.

Teddlie, C. & Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. In A. Tashakkori, & C. Teddlie. (Eds.), *The sage handbook of mixed methods in social & behavioral research* (2nd ed, pp. 1–41). Sage.

Universities UK. (2019). *Higher education in facts and figures 2019*. https://www.universitiesuk.ac.uk/sites/default/files/field/downloads/2021-08/higher-education-facts-and-figures-2019.pdf

UK Government. (2020). Making a support bubble with another household. *Department of Health and Social Care.* https://www.gov.uk/guidance/making-a-support-bubble-with-another-household

Wallace, S., Schuler, M. S., Kaulback, M., Hunt, K., & Baker, M. (2021). Nursing student experiences of remote learning during the Covid-19 pandemic. *Nursing Forum, 56*(3), 612-618. https://doi: 10.1111/nuf.12568

Wen, F., Zhu, J., Ye, H., Li, L., Ma, Z., Wen, Z., &. Zuo, B. (2021). Associations between insecurity and stress among Chinese university students: The mediating effects of hope and self-efficacy. *Journal of Affective Disorders, 281*, 447–453. https://doi: 10.1016/j.jad.2020.12.047

World Health Organisation. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. World Health Organisation. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

Yang, D., Tu, C. C., & Dai, X. (2020). The effect of the 2019 novel coronavirus pandemic on college students in Wuhan. *Psychological trauma: Theory, Research, Practice and Policy, 12*(S1), S6-S14. https://doi: 10.1037/tra0000930

Ye, Z., Yang, X., Zeng, C., Wang, Y., Shen, Z., Li, X., & Lin, D. (2020). Resilience, social support, and coping as mediators between Covid-19-related stressful experiences and acute stress disorder among college students in China. *Applied Psychology: Health and Well-Being, 12*(4), 1074–1094. https://doi: 10.1111/aphw.12211

Zhang, X., Huang, P., Li, B., Xu, W., Li, W., & Zhou, B. (2021). The influence of interpersonal relationships on school adaptation among Chinese university students during Covid-19 control period: Multiple mediating roles of social support and resilience. *Journal of Affective Disorders, 285,* 97–104. https://doi: 10.1016/j.jad.2021.02.040

Appendices

Appendix A: Search strategy.

Topic: Resilience of HE students in responding to the pandemic and the resulting ongoing changes to practice.

Key concepts*	Coronavirus	Setting	Students	Mental health/ wellbeing	Transition to distance learning	Ongoing changes to practice	Other (life) factors	Differences between lockdowns
Alternative terms / synonyms	Coronavirus Covid-19 Pandemic Corona Covid lockdown	University College HE Higher Education	Student* University Student* Higher Education Student* HE Student*	"mental* illness*" Wellbeing "Mental* health*" "anxiety" "stress" "trauma" "work life balance" "wellbeing"	Equipment/ access to technology "digital literacy"/ competence/ confidence "distance / remote learning" "online learning" Or "online learn*" "online lectures" Or "online lectur*" "online provision" "face-to-face learning"/ "face-to-face" "learn* engage*" Curriculum/ targets "IT support"	"Changes to practice" "blended learning" New normal Recovery New pedagogy/ curriculum Future Implications "best practice" planning "build back better"	Family concerns Family illness worry bereavement Separation/ isolation Parental duties Home schooling Shielding Caring Financial Carer responsibilities	First (UK) lockdown Second lockdown First wave Second wave Third wave Etc Vaccination New strains New variants
Search terms with operators	Coronavirus OR "Covid-19" OR Pandemic OR Corona OR Covid OR lockdown	University OR College OR HE OR "Higher Education"	Student* OR University Student* OR Higher Education Student* OR HE Student*	"mental* illness*" OR Wellbeing OR "Mental* health*" OR "anxiety" OR "stress" OR "trauma" OR "work life balance"	Technolog* OR Equip* OR "access to technology" OR "digital literacy" OR "digital competence" OR "digital confidence" OR "distance learning" OR "remote learning" OR "online learning" OR "online lectures" OR "online provision" OR "face-to-face learning" OR "face-to-face" OR "learn* engage*" OR curriculum OR target*	"Change* to practice" OR "blended learning" OR "new approaches" OR "New normal" OR Recovery OR pedagog* OR curricul* OR Future OR Implication* OR "best practice*" OR planning OR "build* back better"	Famil* OR "Family concerns" OR "Family illness*" OR worr* OR bereave* OR Separat* OR isolate* OR Parent* OR "Home school*" OR child* OR shield* OR caring OR carer OR finance*	"First (UK) lockdown" OR "Second lockdown" OR "First wave" OR "Second wave" OR "Third wave" OR "wave" OR Vaccin* OR strain* OR variant*

Appendix B: Inclusion criteria.

TYPES OF…	INCLUSION CRITERIA	EXCLUSION CRITERIA	RATIONALE
Topic	Resilience, risk and protective factors identified which are linked to COVID-19 pandemic	Impacts on undergraduate students not linked to COVID-19 pandemic	The study is concentrated on the impact of COVID-19 pandemic in relation to emerging resilience
Article	Peer reviewed journal articles: Empirical studies; conceptual papers based on a clear methodology; meta-analyses	Books; conceptual papers without a clear methodology	Academic focus to concepts
Journal	Fields of Social Sciences, Psychology and Education, Higher Education	Other fields	These fields are most pertinent to a study of University students
Context	University settings	Schools or tertiary education	Current study based in a University
Participants	General population of HE students	Specific groups within HE (i.e. those with disabilities)	Focus of study on HE population as a whole
Time Period	Since 2020	Pre-2020	Time period during and post pandemic
Geographical content	Written in English	Not written in English	Ensure that understanding is maintained as English is the authors first language.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Should I stay or should I go? International students' challenges and opportunities to secure employment in their host country after graduation. A scoping review using PRISMA

David Santandreu Calonge[A]

Mariam Aman Shah[B]

Pablo Medina Aguerrebere[C]

Nadia Abdulla[D]

Melissa Connor[E]

Mira Badr[F]

Emma Blakemore[G]

A *Department of Academic Development, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates*

B *Department of Educational Research, Lancaster University, Lancaster, United Kingdom*

C *Department of Communication and Media, Canadian University Dubai, Dubai, United Arab Emirates*

D *Department of Faculty Affairs, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates*

E *The University of South Australia, UniSA Education Futures, Centre for Research in Educational and Social Inclusion, SA, Adelaide, Australia*

F *Department of Academic Development, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates*

G *Department of Career Services and Internships, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates*

Keywords

Graduate employment;
graduate skills;
international education;
employability;
post-graduation work.

Correspondence

david.santandreu@mbzuai.ac.ae [A]

Article Info

Received 12 June 2023
Received in revised form 3 August 2023
Accepted 14 August 2023
Available online 15 August 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.20

Abstract

The opportunity to find employment is often a key push factor for students to study abroad. However, previous research has established that international graduates often face difficulties in securing employment in their host country and have a lower employment rate compared to local graduates. Although some research has been conducted on this topic in the Australian context, to date, the problem has been under-researched elsewhere. The aim of this scoping review of the literature is to address this gap and examine the challenges faced by international students when seeking employment in their host countries after graduation, as well as the potential opportunities offered to them. 18 articles were identified and were included in the review. Content analysis of the data was undertaken using NVivo 12.0.

Introduction

The internationalization of higher education (HE), student mobility and employment in global labor markets are links on the same chain (Guruz, 2011). They have implications upon each other which not only affect international education but consequently can impact a nation's economy through its policies on employment for graduate skilled labor (Blackmore et al., 2014).

Graduate mobility, often referring to international graduate students' ability to remain in their host countries for employment upon graduation (Wut et al., 2022), is a complex and often understudied area comprised of graduate skills or skills mismatch (Pham & Jackson, 2020; Pham & Saito, 2020; Santandreu Calonge et al., 2019; Calonge & Shah, 2016); graduate outcomes; lucrative international student fees; education policies and immigration policies which are often linked to the processes of economic development (Cameron et al., 2019; Mok & Han, 2016). Despite these complexities, the prospects of finding employment after graduation is often a significant push factor for students to study abroad (Cameron et al., 2019). However, international students who decide to study in a foreign country often face significant challenges in securing employment after graduation (Scott et al., 2015; Khanal & Gaulee, 2019; Tran et al., 2020; Sofat, 2021; Tran et al., 2023).

Previous research in several countries has indicated that these challenges are often due to several factors, including local language proficiency (Mathies & Karhunen, 2021; Zainuddin et al., 2019), mismatch between the skills employers expect and the skills graduates have (Brunello & Wruuck, 2019; Tymon, 2013; Di Pietro & Urwin, 2006), cultural differences and cross-cultural competence (Nguyen & Hartz, 2020; Jackson, 2017; Mehdizadeh & Scott, 2005), a lack of local professional/social networks (Tran et al., 2022b; Alho, 2020), and familiarity with the local job market (Huang & Turner, 2018; Blackmore et al., 2017). In addition, international graduates may, in some contexts, face discrimination and bias from potential employers (Tran et al., 2023; Coffey et al., 2021; Desbiens & Vidaillet, 2010), making it even more difficult for them to find suitable job opportunities. Furthermore, visa restrictions and complex immigration regulations can limit their eligibility for certain types of employment (Tran et al., 2020) and make it difficult for them to remain in the country after graduation. Despite these challenges, many international graduates persist in their efforts to search for work opportunities in their host country, driven by the desire to settle down, gain practical experience and establish a career in their chosen field.

Although some research has been carried out on post-study work in Australia, the United States and in the United Kingdom, to date, only a limited number of studies, apart from perhaps Han et al. (2022), examine international students' employability challenges and opportunities post-graduation globally, in the last five years. Additionally, no single study exists which addresses the two research questions set for this exploratory article. For these reasons, a scoping review was conducted to systematically map the research done in this area, as well as to identify any existing gaps in knowledge. This study, therefore, aims to contribute to this growing area of research by providing fresh insights into the global field of graduate employment.

This study proceeds as follows: Section two reviews the literature and presents the theoretical framework, section three lays out the methodology and conceptual framework, sections four and five present, analyze and discuss the results. The final section concludes and discusses implications.

Background

Employability skills and graduate employability

Employability is a multi-dimensional, competence-based construct (Römgens et al., 2020) that has grown in currency in the last twenty years, used in higher education and government policies globally. However, employability remains a "woolly concept to pin down" (Cranmer, 2006, p. 172) due to the different definitions, meanings, and usage of the term. While there is no one fixed definition of employability, common across the literature is that employability focuses on the lifelong attainment of skills and attributes that will prepare people for gaining and keeping employment (Römgens et al., 2020; Osmani et al., 2019). Yorke's (2004) definition of employability refers to a "set of achievements, skills, understandings and personal attributes – which makes graduates more likely to gain employment and be successful in their chosen occupations" (p. 8). Hillage & Pollard (1998) focus their conceptualization of employability on the individual's ability to "realize potential through sustainable employment" (p. 2) over the course of their working life and have the necessary skills to find fulfilling work. Yorke and Knight's USEM (Understanding, Skills, Efficacy, Metacognition) model proposes that employability needs to be embedded in the curriculum as employability is a strength to 'good learning' rather than something that detracts from the academic curriculum. In their influential model for thinking about employability, Yorke and Knight also stated that employability is "not something static but something a person can develop throughout life" (Yorke & Knight, 2006, p. 3).

Common across these widely referred-to models of employability is it being something more than gaining employment but rather a focus on the transferability of skills across different occupation domains, circumstances, and the lifelong development of employability skills. Often these skills are provided as a list of generic skills (Succi & Canovi, 2020) and knowledge such as "problem solving, leadership, critical thinking, interpersonal skills, adaptability, teamwork, and personal qualities" (Krishnan et al., 2021, p. 29).

As employability is difficult to define and measure, higher education institutions and government policies often interchange the term with employment outcomes (Behle, 2020), resulting in crude statistics on employment rather than employability and a focus on job-getting as opposed to the ability to "create and sustain work, over time" (Bennett, 2019, p. 32). Short-term metrics, league tables and funding have been tied to graduate employment outcomes, such as graduate destination surveys, rather than the actual employability of graduates (Jackson & Bridgstock, 2021).

With the 'massification' of higher education, the university degree has become a standard expectation for many jobs, hence the requirement for graduates to develop additional skills on top of degree knowledge (Barrie, 2006). Within countries such as Australia, the United Kingdom, Canada, and New Zealand, there has been a strategic push by higher education institutions to include career-readiness attributes and employability as part of their offering (Jackson & Bridgstock, 2018). There is also the growing expectancy that university graduates will not only have degree knowledge but also be able to immediately apply an array of skills that are essential to the workplace (Griffin & Coelhoso, 2019).

Work-readiness and the skills mismatch

Graduates are increasingly expected to be work-ready and able to apply both their degree knowledge as well as seamlessly transition to the workplace and use their generic skills (Winterton & Turner, 2019). With this, universities are expected to produce work-ready graduates and prepare their students with a diverse set of skills and capabilities (Pouratashi & Zamani, 2019). The preparation of work-ready graduates is a key purpose of universities due to the coupling of education and the labor market (Jackson, 2014; Tomlinson, 2012). The use of graduate attributes by universities as sets of lists of work-ready skills signals to the industry that the university has adequately prepared the graduate with a set of skills and capabilities that will see them transition successfully to the professional environment (Borg & Scott-Young, 2020; Hatzenbuhler, 2019; Daniels & Brooker, 2014). However, there is concern that universities are not adequately preparing graduates for the skills needed in the labor market and employers' expectations resulting in a skills gap (Salas-Velesco, 2021; Calonge & Shah, 2016; Mocanu et al., 2014). A recurrent complaint from employers is that there are no suitable graduates (Small et al., 2022). This indicates that there is a problem with the supply side and the perception that students are not graduating with the requisite skills and knowledge required by employers (De Lange et al., 2022; Osmani et al., 2019).

In the literature, the exploration of employability from the perspectives of the graduate, the employer, and higher education and the differences between these perspectives yields a gap that adds to the challenges of graduates, particularly international graduates, and perpetuates the skills mismatch. These views are subjective and continuously evolving, which means there is a potential risk of a constant gap between the skills the graduates acquire at university, the employers' needs, and market requirements (Mansour & Dean, 2016).

Logistical challenges

Han et al. (2022) argued that "a country's immigration policies can play a critical role in influencing international graduates' settlement decisions and work integration" (p. 183). Employment visa processes are often complex, lengthy, and costly for employers with no guarantee of obtaining a working visa for an international graduate, so organizations tend to recruit from the local talent pool. This complication puts international graduates at a disadvantage in the host country.

Furthermore, other practices in some countries, such as the prioritization of hiring citizens or permanent residents of a country over hiring of immigrant employees or requiring organizations to provide evidence that the position prioritizes local candidates (Han et al., 2022), may again put international graduates at a significant disadvantage.

Theoretical framework

Various theoretical frameworks have been used in the literature that address the employability concept. For this systematic review, the authors of this article opted for the Human Capital Theory (HCT) (Becker, 2009).

HCT posits that investment in education and training positively affects performance, productivity and, ultimately, the general economy by enhancing knowledge and skills and making graduates employable. This, in turn, supports economic productivity and provides, in theory, better compensation for new market entrants (Herrmann et al., 2023). The aim of the education system is to support the development of human capital, which includes developing transferrable skills and competencies that can add value to graduates' employability. This is considered general human capital. Specific human capital is developed through education, training and experience and may potentially be less transferable and may not support graduate mobility. The education system should be designed to contribute to both categories of human capital development (Mocanu et al., 2014). Human Capital considers multiple dimensions: person, organization, and market. The first dimension suggests that individuals can boost their earning potential and overall economic value by investing in their own human capital. The organization level investigates the collective competencies of employees within an organization setting. Finally, the macro-level, or the labor market, is concerned with the overall competencies available in the workforce by specifically considering academic qualifications (Smaldone et al., 2022).

The relevance of this approach to the current research lies in its emphasis on investing in human capital, with higher education being a crucial component that international students seek from international universities. According to Tran et al. (2020), higher education can enhance the likelihood of securing better job prospects and higher income in the host country where international students complete their degrees.

When discussing the challenges and opportunities of graduate employment in their graduate host country, HCT can thus provide significant insights into the value of education and training as a means of enhancing employability. For instance, graduates who possess specialized skills, capabilities or knowledge that are in high demand in their host country may have greater opportunities for employment and career advancement. Similarly, graduates who invest in additional education or training may be more competitive in the job market, increasing their chances of securing employment. On the other hand, Human Capital Theory can also explain some of the challenges that graduates may face in their job search. For example, if a graduate's education or skills are not valued in their host country, they may face difficulty finding suitable employment opportunities or end up underemployed, with lower-paid, lower-skilled roles, often referred to as brain waste (Mattoo et al., 2008). Additionally, graduates who lack the financial resources to invest in additional education or training may face limited opportunities in the highly competitive entry-level end of the job market.

When employing HCT as the theoretical framework, it is critical to explore the theoretical perspectives of the researchers. The premise of the human capital theory is to provide a universal mechanism of exploring and investigating the relationship between education and employment. It describes these relationships as interconnected concepts, which might be more suitable for a more relativist perspective and may not be as effective when using empirical methods (Blair, 2018; Marginson, 2017). Due to the closed-system nature of this theory, it does not account for the potential external factors that might impact this relationship since both education and employment exist in complex interconnected systems (Marginson, 2017). Another study highlights how the theory does not take into consideration how individuals with educational backgrounds end up in different occupations, although it does predict the association between education and income (Kivinen & Ahola, 1999). Kivinen & Ahola (1999) argue even the highest level of credentials and education cannot guarantee job security, thus providing another limitation of the closed system of HCT. Despite these limitations, HCT provided this paper with a theoretical framework for understanding the challenges and opportunities of graduate employment in their graduate host country, highlighting the importance of education and skills as a form of investment in human capital.

Methods

The study was conducted in the form of a scoping review (Arksey & O'Malley, 2005). A scoping review "differs from a systematic literature review in that it requires broader research aims" (Schwendimann et al., 2018, p. 3) and is an "ideal tool to determine the scope or coverage of a body of literature on a given topic" as it gives "clear indication of the volume of literature and studies available as well as an overview (broad or detailed) of its focus" (Munn et al., 2018, p. 2). Searches by three independent researchers from February to April 2023 focused on Google Scholar and Scopus, yielding a total of 97 results. Five phases were then undertaken: (1) identification of research question(s), (2) identification of relevant studies, (3) selection of studies, (4) charting of data according to issues, codes, and key themes, and (5) collating, summarizing, and reporting of results (Arksey & O'Malley, 2005).

Phase 1: Identify the research question(s)

The following research questions were investigated:

(1) What barriers do international students face when seeking employment in their university education host country, post-graduation?

(2) How do host countries and higher education benefit from opportunities of employment for international students, post-graduation?

Phase 2: Identify relevant studies

To focus on the most current research, database searches were limited to the past 5 years (2019–April 2023). Figure 1 shows Boolean search terms and numbers. The abstract and full-text screening was performed by three authors. The inclusion and exclusion criteria were agreed upon by the research team.

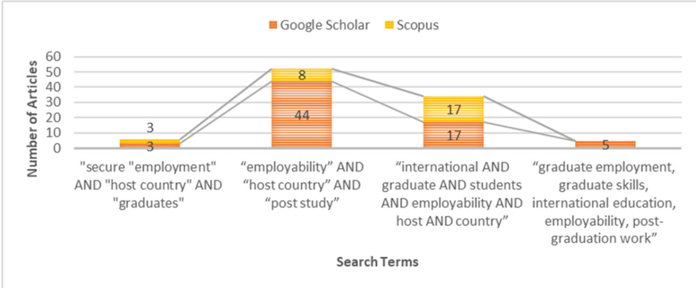


Figure 1. Search Terms, databases, and numbers.

Phase 3: Selection of studies

The review included industry reports, articles, and documents to minimize bias and provide a reliable and reproducible assessment. A protocol was drafted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for scoping reviews (PRISMA-ScR, as shown in figure 2) (Tricco et al., 2018). PRISMA-ScR provides a standard methodology that uses a 20-item guideline checklist. Studies were screened and included in the review if they were: (1) written in English, (2) peer-reviewed (articles/book chapters), (3) reports, (4) Op-eds, (5) conducted in any country, and (6) published between 2019 and April 2023 (Table 1). Studies were excluded if (a) they were published in a language other than English, if (b) they predated 2019, if (c) full text was unavailable, if (d) was not related to employment or employment challenges/barriers/opportunities, if (e) it was not related to employment or employability post-graduation, and if (f) it was an unpublished thesis/dissertation. In total, 18 articles were selected for inclusion. Krippendorff's alpha coefficient (Krippendorff, 2011) was used to determine the degree of inter-rater reliability for abstracts (.85) and full texts (1.00). The three reviewers resolved disagreements on study selection and data extraction by discussion and consensus to reach 100% agreement.

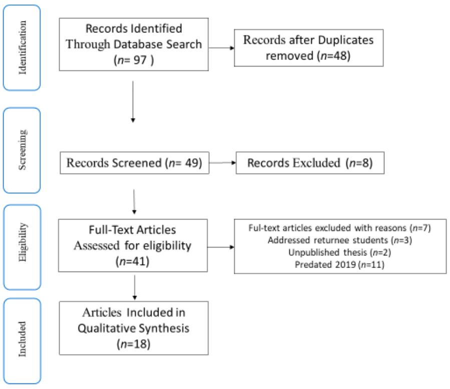


Figure 2. Overview of literature search process using PRISMA-ScR.

Results

Phase 4: Chart data

Data from eligible studies were charted using Excel. Table 1 provides a list of authors, year of publication (reverse chronological order), the title of the article, source, type of article (qualitative/quantitative/mixed methods), context, inclusion criteria and main themes from the article. Figure 3 graphically shows the countries where the included studies were conducted.

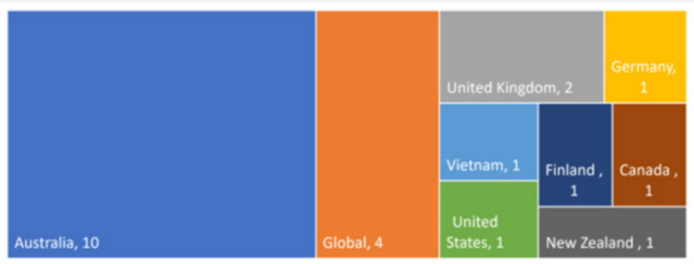


Figure 3. Country and number of articles where the included studies were conducted.

In this scoping review of the literature, 18 studies related to the challenges and opportunities met eligibility for review. Content analysis to identify themes was undertaken using NVivo 12.0.

Table 1. Overview of included studies.

#	Author (s) / Year of Publication	Article Title	Source / Article Type / Research Method	Context	Inclusion Criteria	Themes
1	AGCAS. (2023).	International Graduate Routes Narratives from the UK Job Market	*The Association of Graduate Careers Advisory Services* *Report* Mixed Methods	United Kingdom	1,3,4,5,6	▪ Employment visa ▪ International graduates ▪ Employer's knowledge / experience in hiring international graduates ▪ Graduate internship schemes ▪ Barriers to post-graduate employment ▪ HE careers support
2	Orr, P., Forsyth, L., Caballero, C., Rosenberg, C., & Walker, A. (2023).	A systematic review of Australian higher education students 'and graduates 'work readiness	Higher Education Research & Development, 1-18. Qualitative	Australia	1,2,4,5,6	▪ Gaps between HE and Industry ▪ Graduate work readiness ▪ Difference between work-ready and employability ▪ Industry experience ▪ Industry initiatives with HE – internships, mentorships ▪ Core employability skills ▪ Work Integrated Learning
3	Tran, L. T., Tan, G., Bui, H., & Rahimi, M. (2023).	International graduates on temporary post-graduation visas in Australia: Employment experiences and outcomes.	Population, Space and Place, 29(1), e2602. Qualitative	Australia	1,4,5,6	• Immigration policy • Field of study • Employment trends based on discipline
4	Han, Y., Gulanowski, D., & Sears, G. J. (2022).	International student graduates' workforce integration: A systematic review	*International Journal of Intercultural Relations* Qualitative	Global	1,2,4,5,6	▪ Immigration (permanent residence policies) ▪ Graduates' employment intention (motivation) ▪ Family influence ▪ Country and level of education ▪ Language fluency and skills Cultural knowledge and competence ▪ Local social networks (social capital) ▪ Work experience ▪ Employers' recruitment practices ▪ HE support in Immigration procedures ▪ Immigration policy ▪ Economy ▪ Safety and Stability of host country ▪ Psychological capital (agency) ▪ COVID Pandemic ▪ Pay disparity
5	Tran, L. T., Phan, H. L. T., Tan, G., & Rahimi, M. (2022c).	I changed my strategy and looked for jobs on Gumtree': the ecological circumstances and international graduates' agency and strategies to navigate the Australian labour market	*Compare: A Journal of Comparative and International Education* Mixed Methods	Australia	1,4,5,6	▪ Immigration policy ▪ Employer recruitment practices (preference to local graduates)
6	Tran, L. T., Bui, H., Tan, G., & Rahimi, M. (2022a).	"It is not O.K to think that you are good just because you have graduated from overseas": Agency and contextual factors affecting Vietnamese returning graduates.	International Migration, 60(6), 43-59. Qualitative	Vietnam	1,4,5,6	▪ Graduates' employment intention (motivation) ▪ Immigration policy ▪ Psychological capital (agency)
7	Weilage, C., & Maráz, G. (2022).	Online Study's Influence on International Student Employability Factors in Germany: Germany Vs. Overseas Based Students.	Journal of Teaching in International Business, 33(1), 7-30. Qualitative	Germany	1,4,5,6	▪ Language fluency and skills ▪ Cultural knowledge and adjustment (cultural capital) ▪ Local social networks (social capital)
8	Coffey, J., Farivar, F., & Cameron, R. (2021).	The job seeking experiences of international graduates in the host country: Australia's lost opportunity?	*The International Journal of Human Resource Management* Qualitative	Australia	1,4,5,6	▪ Employer environment and practices (Discrimination, lack of understanding of the immigration policy and procedures)
9	Jackson, D., & Pham, T. (2021).	International students and work-integrated learning Overcoming challenges and looking to the future	*Advances in research, theory and practice in work-integrated learning. Routledge.* Qualitative	Australia	1,4,5,6	▪ Employability and WIL ▪ Local social networks (social capital) ▪ Language fluency and skills ▪ Cultural knowledge and adjustment (cultural capital) ▪ Immigration policy ▪ HE support in WIL ▪ Employer and supervisor support in WIL ▪ Psychological capital (agency)
10	Tran, L. T., Bui, H., Tan, G., & Rahimi, M. (2022b).	Post-Graduation Work Visas and Loopholes: Insights into Support Provision for International Graduates from the Perspectives of Migration Agents, Universities, and International Graduates.	Evaluation Review, 46(4), 438-464. Qualitative	Australia, Canada, New Zealand, the UK, and the U.S.	1,4,5,6	▪ Immigration (permanent residence policies) ▪ Language fluency and skills ▪ Psychological capital (agency)
11	Alho, R. (2020).	You need to know someone who knows someone': international students' job-search experiences.	Nordic journal of working life studies. Qualitative	Finland	1,4,5,6	▪ Immigration ▪ Local social networks (social capital) ▪ Employer recruitment practices ▪ Language fluency and skills ▪ Psychological capital (agency)

12	Baron, G., & Hartwig, K. (2020).	Workplace Experience of International Students in Australia	Journal of International Students, 2020 Vol. 10 No. 2, 218. Mixed-Methods	Australia	1,4,5,6	▪ HE support in WIL (culture, support services)
13	Pham, T., & Jackson, D. (2020).	Employability and determinants of employment outcomes	*Book Chapter,* *Developing and utilizing employability capitals (pp. 237-255). Routledge.* Qualitative	Global	1,4,5,6	▪ Influence of governmental policies ▪ Immigration policy ▪ Graduates' employment intention (family) ▪ Employers' perceptions of graduate skills ▪ Cultural knowledge and adjustment (cultural capital) ▪ Psychological capital (agency)
14	Singh, J. K. N. (2020).	Why do Chinese international students studying in Australia repatriate? Australian Chinese graduates tell it all.	*Journal of Further and Higher Education* Qualitative	Australia	1,4,5,6	▪ Immigration policy ▪ Employer recruitment practices (bias and discrimination) ▪ Family influence
15	Tran, L. T., Rahimi, M., Tan, G., Dang, X. T., & Le, N. (2020).	Post-study work for international graduates in Australia: opportunity to enhance employability, get a return on investment or secure migration?	*Globalization, Societies and Education* Qualitative	Australia	1,4,5,6	▪ Immigration policy ▪ Local social network (social capital) ▪ Language fluency and skills
16	Cameron, R., Farivar, F., & Coffey, J. (2019).	International graduates host country employment intentions and outcomes: Evidence from two Australian universities.	*Journal of Higher Education Policy and Management* Quantitative	Australia	1,4,5,6	▪ Graduates' employment intention (motivation) ▪ Graduate Employability ▪ Language fluency and skills (English) ▪ Employer recruitment practices (bias and discrimination) ▪ HE support in Immigration procedures
17	Khanal, J., & Gaulee, U. (2019).	Challenges of international students from pre-departure to post-study: A literature review.	Journal of International Students, 9(2), 560-581. Qualitative	Global	1,2,4,5,6	▪ Language fluency and skills ▪ Financial issues ▪ Cultural knowledge and adjustment (cultural capital) ▪ Discrimination ▪ Personal issues (homesickness, isolation, culture shock, dietary issues - Psychological capital) ▪ Local social network (social capital) ▪ Immigration policy
18	Pham, T., Tomlinson, M., & Thompson, C. (2019).	Forms of capital and agency as mediations in negotiating employability of international graduate migrants.	Globalization, Societies and Education, 17(3), 394-405. Qualitative	Global	1,4,5,6	▪ Local social network (social capital) ▪ Employability skills ▪ Cultural knowledge and adjustment (cultural capital) ▪ Psychological capital (agency)

Figure 4 shows the number and percentage of word occurrences. Words related to external factors, such as employment practices, immigration-related issues, and local social networks, were among the most cited by the authors.

Assessment of quality, reliability and confidence

Pollock et al. (2022) indicated that critical appraisal and assessment of the quality of articles to be included in a scoping review were "not mandatory" (p. 1099).

Phase 5: Collate, summarize and report the results

The collation of data leading to results was gathered through a process of thematic analysis (Clarke et al., 2015). Through this process, patterns, phrases, and concepts were identified in the literature resulting in four subthemes and unveiling several leading concepts, as shown in Table 2. To present the results identified through the subthemes and leading concepts, the data was further summarized into 3 core overarching themes:

1. Impact of host countries' immigration policies on international graduate students

2. Higher education practices for graduate employment readiness

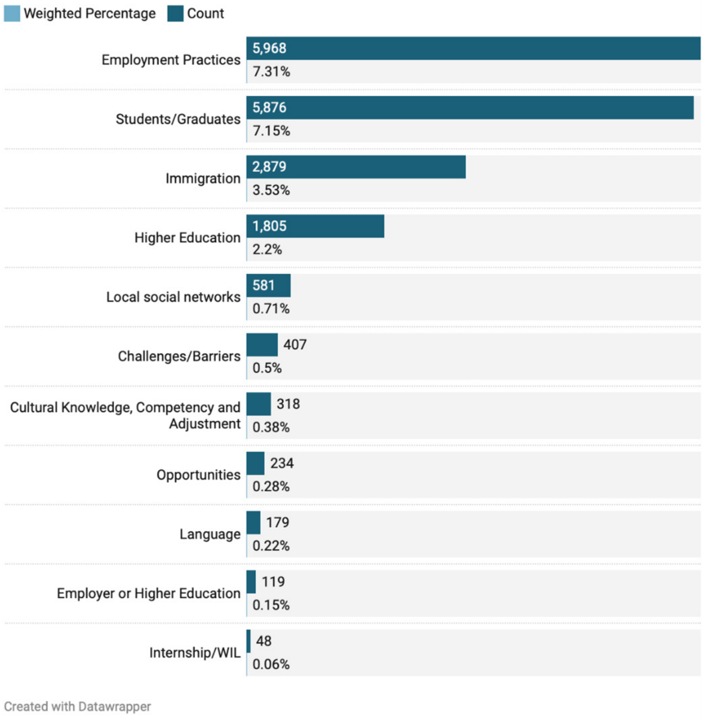


Figure 4. Themes mapped to word count frequency.

Table 2: Overarching themes unpacked.

Overarching Themes	Subthemes	Leading Concepts
Impact of Host Countries' Immigration Policies on International Graduate Students	1. Influence of government policies and employment visa	▪ Host countries economic stability and COVID pandemic ▪ Employment recruitment practices ▪ Higher Education support in immigration procedures
Higher Education Practices for Graduate Employment Readiness	2. Work Integrated Learning (WIL)	▪ Multifaceted graduate skills ▪ Bridging gaps between employers' knowledge, perceptions of graduates' skills and experiences in hiring international graduates ▪ HE career support
International Graduates' Economic and Social Integration Capabilities into Host Countries	3. Psychological capital and integration skills 4. Graduates' employment intentions	▪ Local social networks ▪ Language skills and fluency ▪ Emotional support for personal challenges when considering long-term living in a host country, such as homesickness, isolation and loneliness. ▪ HE support of WIL through culture support services ▪ Graduates' employment intentions based on financial concerns

3. International graduates' economic and social integration capabilities into host countries

Theme 1: Impact of host countries' immigration policies on international graduate students

A host country's international talent pool teeters between the conditions of its immigration policies, growing economic status and the drive for and recognition of the impact which international talent can play on productivity, labor shortages and innovation (Han et al., 2022; Li, 2020). Numerous studies including Singh (2020), Tran et al. (2023), Coffey et al. (2021) and Jackson and Pham (2021), all point to this. However, the data found from these and other studies examined in this paper also indicate the discrepancies which rest between host countries' immigration policies and the challenges they cause for international graduate students' employment. In this regard, the impact on international graduate students is said to be that of a "competitive disadvantage" when it comes to securing employment in their host countries' labor market (Tran et al., 2023, pp. 2-3).

Tran et al. (2023), Coffey et al. (2021) and AGCAS (2023) have highlighted that a major barrier towards employment opportunities for international graduates is due to their visa status. Immigration policies in many host countries, such as the UK and Australia, cater to various forms of temporary graduate visas, which do not have any security of employment, unlike employment security gained through an employer-sponsored visa (Tran et al., (2023). When reflecting on such policies in the UK, a report by the Association of Graduate Careers Advisory Series (AGCAS) (2023), emphasized the "lack of government support" in this process of visas, creating challenges for international graduates' long-term prospects in their host country. This is seen in the literature through Tran et al. (2022c), Tran et al. (2020) and AGCAS (2023) as having reciprocal effects on the potential of a host country's economic development through a loss in the international graduate labor market.

The results found in the data go further to link such challenges in government policies on international graduate student visas to "broader socio-economic contexts" (Tran et al., 2022a, p. 43). This is evident as shifts in economic growth and migration procedures, largely due to the COVID pandemic, led to no ease in visa policies but rather additional constraints for graduate employment opportunities (Han et al., 2022).

In addition to results indicating international graduates being an integral component in fostering economic growth in their host country, as suggested by Han et al. (2022), it was also found that complications in long-term visas for international graduates create barriers to the recruitment practices of potential employers. Pham and Jackson (2020) underscore this by bringing to light that long-term consistencies in the inability to employ international graduate students can develop into a lack of awareness of the talent which such graduates can bring to the growth of locally based industries, thus creating hesitations in the process for their recruitment. Alho (2020) further emphasizes this by stating that under such circumstances, recruitment patterns become "context-bound" (p. 3). This can leave international graduates vulnerable to loopholes and unethical practices in the pursuit of long-term employment in their host countries (Tran et al., 2022b).

Theme 2: Higher education practices for graduates' employment readiness

International graduate mobility, career intentions, employment outcomes and economic growth connect to practices carried out through higher education institutions for their graduates (Cameron et al., 2019). The results found in the literature indicated that to leverage the abilities of international graduates for employment in their host countries, higher education institutions need to be resilient in the development of their programs (Pham & Jackson, 2021; Baron & Hartwig, 2020; Tran et al., 2023). This encompasses aspects of WIL, along with providing career support and advice; cultivating multifaceted skills; and acting as agents which bridge gaps between employers' knowledge of international graduates' skills and their hiring processes (Baron & Hartwig, 2020). As Cameron et al. (2019) expressed, not only will graduates benefit from this, but such practices are strongly in favor of the universities themselves as they become "an important attraction" in terms of (significant) revenue from international students (p. 550), £42 billion to the UK economy in 2022-23, AUS$25.5 billion to the Australian economy in 2022. $33.8 billion to the U.S. in 2022, and $5 billion to France in 2022.

Data also indicated this nevertheless tends to be restricted to "degrees such as business, education, engineering and health sciences" (Baron & Hartwig, 2020, p. viii). Baron and Hartwig (2020) highlighted that such degrees often set requirements for successful graduation, which include WIL through mechanisms such as internships. This, as Orr et al. (2023) and Pham et al. (2019) suggest, do not necessarily cater towards multifaceted skills and resources for international graduate employability. In this regard, employers may be unable to recognize other essential skills which are not discipline specific. The results therefore pointed to gaps which exist in the practices of higher education, the work readiness of their graduates, and the perceptions of skills attained by graduates for potential employers in host countries (Han et al., 2022).

Theme 3: Graduates' economic and social integration capabilities into host countries

The findings in the literature indicate that although immigration policies and WIL are strong components which impact international graduate students' abilities to secure employment in host countries, graduates' psychological capital, social integration capabilities and economic concerns also play a large role (Khanal & Gaulee, 2019; Han et al., 2022; Jackson & Pham, 2021). Pham et al. (2019) pointed this out to be what they called the development of "key forms of capital" (p. 394). Similar sentiments were echoed by Tran et al. (2022a) and Alho (2020) when highlighting international graduates' integration into host countries' labor markets, with Alho (2020) stating that the process of integration is "embedded in national, cultural and institutional contexts" (p. 3). In other words, long-term stay in the context of host countries requires social integration capabilities, such as local language skills, which additionally aid in building psychological capital by lessening isolation through language barriers and strengthening a sense of belonging to the country (Weilage & Maraz, 2022; Khanal & Gaulee, 2019).

In addition to overcoming language barriers, the links between international graduates' social integration, psychological capital and economic concerns when considering employment in host countries requires multiple facets of support (Cameron et al., 2019; Singh, 2020). An example of this is integration through the development of local social networks. This generates an understanding of cultural diversity both by graduates and potential employers

(Jackson & Pham, 2021). This may also contribute towards emotional and psychological support by easing personal challenges for international graduates (Pham et al., 2019; Jackson & Pham, 2021).

The results suggest that support may also be gained through cultural integration services provided by higher education institutions (Weilage & Maráz, 2022, Baron & Hartwig, 2020; Jackson & Pham, 2021). WIL can be utilized to provide not only an understanding of graduate employment related skills, but also as a means to open another channel for social integration, psychological support, and also easing concerns of economic stability for international graduates (Weilage & Maráz, 2022; Han et al., 2022). Higher education services which increase cultural integration, also opens the window to what Tran et al. (2022a) identified as "interrelated contextual factors", enabling international graduates to compare host and home contexts and economic strains or leverages which they may incur within both (p. 43). This awareness of economic stability upon graduating in host countries, in turn shapes the intentions of international graduates as to their decision to stay in their host countries or leave.

Discussion

Higher education institutions play a key role in improving students' skills, enhancing companies' performance, and transforming local societies (Chai et al., 2020). However, these institutions must still improve in several areas, especially in how they fully integrate international students: programs, workshops, internships, and professional opportunities (Dos Santos, 2021). Our results indicate that these students face three main barriers: a) personal issues, b) inefficiencies of higher education institutions, and c) obstacles from public and private organizations.

Personal issues

International students' motivations to study abroad are multiple: to explore a new country, seek out adventure, experience a new environment, learn about other cultures, learn a new language, socialize and to develop networks (Casas Trujillo et al., 2020). Their main motivation, though, is mostly to boost their employability (Cho et al., 2021). However, international students face several issues when they move abroad (Tran et al., 2022): homesickness, isolation, cultural shock, or dietary issues (Khanal & Gaulee, 2019), as well as other barriers related to their psychological capital (Tran et al., 2022). Our analysis shows that support mechanisms (e.g., pastoral care structures) are vital and need to be put in place (Calonge et al., 2022) to ease an often-stressful transition. This critical support plays a key role in helping students overcome their personal issues when studying overseas (Chai et al., 2020).

Helping students to adapt to a new country is a key element because the challenges of acculturation to the stressors of academic study and everyday life in a foreign environment make these students a vulnerable population: they are more likely to suffer from stress, boredom, depression, and mental health issues (Minutillo et al., 2020). Universities should provide international students with extensive pre-arrival information and organize orientation sessions on arrival with local students and international students with similar background to prepare them for their new environment (Jamilah et al., 2020). International students' acculturation modes (assimilation, integration, separation) highly determine their professional career decision-making processes (Li & Lindo, 2022). However, they also face another challenge: the relatively short amount of time they have available to adapt to the new host environment. For example, in Australia, the duration of study is shorter than in other countries, which forces students to keep a high level of academic performance and quickly overcome cultural barriers such as misunderstandings, stereotypes, racial discrimination or conflicts related to lifestyle (Pekerti et al., 2020). In Australia, as well as in other countries, international students face a major issue: their English professional proficiency. Being proficient in English is among the top-sought skills for jobs (Abbas et al., 2021) and highly determines international students' professional careers (Wang, 2020).

Inefficiencies of higher education institutions

According to our results, we can state that international students face several issues related to higher education's low performance in different areas: gaps between tertiary institutions' priorities and industry needs (Orr et al., 2023), lack of support from higher education institutions to help students attend cultural programs about the host country (Baron & Hartwig, 2020), and absence of initiatives to help students adjust to their new environments such as courses, orientation, and procedures (Jackson & Pham, 2021). Whilst international board exam equivalents are, for instance, not often accepted by host country institutions, forcing students to retake courses that they already took in their home country, leading to frustration, anxiety, and disengagement, Lee et al. (2019) spoke about academics' "perceived burden in supervising international students during placement" (p. 1). Additionally, the disconnect between universities' research priorities and new academic program development and companies' needs (staffing, skills) makes it difficult for students to find jobs related to their major (Shams & Thrassou, 2019; Fakunle & Pirrie, 2020).

Obstacles from public and private organizations

With respect to the first research question, we identified some of the most important external barriers affecting international students: host country's immigration policies (Tran et al., 2023), visa programs (AGCAS, 2023), lack of support from higher education institutions when applying to these visa programs (Han et al., 2022; Cameron et al., 2019), local employers' recruitment practices (Tran et al., 2022a), employers' racial discrimination when recruiting international students (Coffey et al., 2021), and the difficulty to develop local networks (Weilage & Maráz, 2022). One of the main challenges is the high level of domestic and international competition for jobs in the local job market. Graduates may also face challenges related to their work experience (or lack thereof), as many employers require prior industry experience, besides internships and/or (unpaid) Work-Integrated Learning (WIL), before hiring. In fact, a report by Chew (2019) in the Australian context indicated that “employment outcomes” for graduates seemed to “improve for those who have accumulated more professional and life experience” (p. 9). Another challenge highlighted by Berquist et al. (2019) is that employers were often “unclear” (p. 21) or had limited understanding or awareness of international graduates’ work rights and entitlements.

Language proficiency and cultural barriers, or discrimination, can also be a significant obstacle, particularly in countries where the official language or cultural norms differ from the graduate's native language and culture, as international graduates, even after having spent three or four years studying, may not have the “same level of local knowledge, understanding of local workplaces, and sustaining connectedness with Australia as those who hold PR or are local citizens” (Berquist et al., 2019, p. 21). In the United States, nativism, for instance, affects international students by restricting them to low-paid jobs (Allen & Bista, 2021). In the United Arab Emirates, there are Emiratization targets, which relate to the number of UAE Nationals employed at that company. Pertaining to the private sector, there is a penalty for companies not meeting targets. In Finland, Anttila (2022) argued that many local employers refused to recruit foreign students because of their level of Finnish or/and because of stereotypes. Additionally, as post-study work visas are somewhat limited (number and time, 2-4 years), graduates often decide to either return home post-graduation (Song & Kim, 2022), or extend their stay by opting to study for another degree, if financial resources allow. For some students whose related family support is back in their home countries, upon graduation, this can lead to a lack of financial and housing support, if this has previously been awarded by the educational institution. In the context of the United Arab Emirates (UAE), aside from the recently launched 10-year Golden Visa, there can be visa challenges for certain nationalities, which may have a knock-on effect on certain candidates successfully being hired. Upon graduation, international students’ visas expire after a 60-day grace period unless the family is resident in the UAE. If this is the case, male students over the age of 25 can stay on their parents’ sponsorship, although they must register for another educational course of at least one year’s duration. For females, this can continue until married, but it must also be for study purposes. The German Academic Exchange Service (DAAD, 2023) estimates that only one-third of international students (around 25,000) remain in Germany each year post-graduation. Australia has recently announced (July 2023) the extension of post-study work rights available to international students who graduate with selected degrees in health and medical fields, teaching, engineering, computer science, and agricultural fields (according to the 2022 Skills Priority List), linked to labor skills shortages in the Australian economy (Parkinson et al., 2023).

On the other hand, and with respect to the second research question, graduates may also encounter opportunities such as networking with professionals in their field, gaining exposure to new industries and work cultures, and accessing specialized training or education programs not available in their home country. International graduates may have a unique advantage of providing new perspectives, diversity in backgrounds and cultures, as well as a global perspective to employers in the host country, as “attracting higher skilled migrants can improve productivity by bringing skills that take years to develop and are in relatively short supply (Parkinson et al., 2023).

Higher education institutions interested in keeping international students should therefore work with public authorities and industry to develop policies aiming to help these students enter the local labour market to gain local experience in industries with high talent shortages: health services, accommodation and food services, schools, family services, employment training, networking services, etc. (Mathies & Karhunen, 2021). These policies should include initiatives against discrimination in the hiring process, as there is tremendous potential in having international students stay post-study for a country’s development (Zhao et al., 2022). In contrast, Chew (2019) highlighted a lost opportunity when he argued that Australia did “not benefit from the full productivity and participation benefits of this young, well-educated, globally competent and highly motivated cohort of graduates” (p. 11).

Limitations

This scoping review has several limitations. First, more than half of the studies included (55.56%) focus on the Australian context, which may give, to a certain extent, a skewed representation of the phenomenon. Employability is a global issue in higher education, with common approaches to enhancing graduate employability used. However, future research looking at non-western approaches, particularly visa limitations and work restrictions, could be useful. The barriers to employment identified in the scoping review are shown to be consistent across disciplines at undergraduate and graduate levels. To understand any nuances, a further research project could include identifying variations using methods such as longitudinal employment outcome data.
Another limitation refers to the theoretical framework adopted for this study. HCT’s limitation is that it oversimplifies the connection between education, skills, and compensation. Education can provide the market with a signal that the graduate is employable and potentially productive. However, it does not always consider the market requirements, technological advancements, and the broader social and economic aspects of the macro-environment (Herrmann et al., 2023).

Conclusion and implications

In considering country-specific human capital, which considers the context of applications and provides a slight adaptation of the original theory to a specific country, international students who are graduating from the host country may face challenges in securing employment and face disadvantages due to barriers in language, differences in culture and limited local networks. Furthermore, various factors, such as immigration policies and regulations, pose a challenge for international graduates. The lack of support from higher education institutions in navigating these regulations and procedures adds to this difficulty. This may be due to the institution's limited resources or unfamiliarity with the requirements. Providing adequate immigration support may encourage international graduates to remain in their host country.

In addition, employers' perceptions of graduate employability and skills are influenced by their internal requirements as well as the market demands that continuously change over time. For instance, depending on the major the student has graduated from, previous experience may be needed for certain roles in the UAE job market, which is reactive to global technological changes and the UAE's ambitious aim to be a global leader.

Therefore, higher education institutions have an opportunity to enhance graduates' skill sets so they can better meet current market needs. It is evident that there exists a complex interplay between educational decisions and migration choices when considering all these factors holistically (Hurley, 2023).

The findings of this study have several important implications for future practice. There is a need to:

a. Strengthen the cooperation between universities, government, and industry: such as the Victorian Government 'Study Melbourne' program that offers international students free career workshops and work experience opportunities with Australian companies. In South Australia, a partnership between government agencies, Regional Development Australia, and Study Adelaide provides the opportunity for international students to tour regional areas with the aim of promoting regional towns as a place to work and live and addressing skill and workforce gaps.

b. Provide more industry-relevant internships and placements: Whilst Nachatar Singh (2023) argued that "South Asian graduates" in Australia were often employed below their skill level, had "experienced unequal opportunities in accessing employability-related programmes" (p. 7), "skewed towards domestic students" (p. 6) and not relevant to their degree, the Review of the Migration System report (Parkinson et al., 2023) indicated that temporary graduate visas "inhibit students' opportunity and ability to show they can succeed in the Australian labour market" (p. 32).

c. Provide more relevant opportunities to develop entrepreneurial skills and social enterprises are required: Many universities have a focus on including entrepreneurial skills in the curriculum and have on-campus business 'startup' hubs and entrepreneurial challenges which are generic and open to the student cohort, which could result in international students experiencing barriers to participation. A suggestion is to increase the relevance of opportunities to develop entrepreneurial skills and social enterprises by learning from the careers office's approaches to tailoring, mentoring, networking, and projects to account for "diverse prior learning" (Ray & Woodier-Harris, 2012, p. 640) and to overcome the "lack of recognition of different experiences, perspectives and background knowledge" (p. 642). Additionally, Rae and Woodier-Harris advise that there should also be support for academics "in designing and running programs for international students" (p. 653).

d. Develop stronger connections with career office and alumni office: such as specialized units in university career offices with staff who are knowledgeable on issues impacting international student employability. Examples of this include the University of Adelaide 'China Career Ready+ Program' which connects Chinese students with both Chinese and Australian employers and offer opportunities for students to be mentored by experienced people from industry who have cross-cultural experience and knowledge. The University of South Australia partners with the Australian business Bupa to offer work experience to international students with the purpose of building social networks and an understanding of Australian business practices and etiquette. In the UAE, Career offices organize, in collaboration with industry and government, interdisciplinary hackathons to tackle global challenges impacting on society.

   An example of providing information on evolving labor market needs is the 'Employable You' interactive web guide designed by the International Education Association of Australian (IEAA) and Australian government department, Austrade. Programs such as the University of Sydney Business School's 'Job-Smart' program help to articulate the specific skills the labor market is seeking. However, there would be benefits in universities, governments and industry collaboratively creating 'one stop shop' websites and resources to make it easy for international students to access information to assist with understanding, evaluating, and articulating their transferable skills as related to the labor market.

e. Tap into "Cultural capital and 'soft-power' for host countries – Australia", for instance, "gains great international relations value from international students having an unambiguously positive experience while in the country" (Parkinson et al., p.105), thus willing to recommend their host university to future international students based on their academic experience.

Authors' contribution: Conceptualization: DSC; Data Curation: NA and MB; Methodology: DSC, Validation: DSC, NA, MB; Writing-original draft: DSC, MAS, MC, PM, EB, and NA; Writing-review, and editing: DSC, MAS, MC. All authors have read and agreed to the published version of the manuscript.

Data availability statement: The datasets used/analyzed during the current study are available from the corresponding author upon reasonable request.

References

Abbas, F., Rana, A., Bashir, I., & Bhatti, A. (2021). The English language proficiency as a global employment skill: The viewpoint of Pakistani Academia. *Humanities & Social Sciences Reviews, 9*(3), 1071-1077. https://doi.org/10.18510/hssr.2021.93106

AGCAS. (2023). *International graduate routes: Narratives from the UK job market*. https://www.agcas.org.uk/write/MediaUploads/Resources/Internationalisation/International_Graduate_Routes_-_Narratives_from_the_UK_job_market.pdf.

Alho, R. (2020). 'You need to know someone who knows someone': International students' job-search experiences. *Nordic Journal of Working Life Studies, 10*(2), 3-22. https://doi.org/10.18291/njwls.v10i2.120817

Allen, R., & Bista, K. (2021). Talented, yet seen with suspicion: Surveillance of international students and scholars in the United States. *Journal of International Students, 12*(1), 175–194. https://doi.org/10.32674/jis.v12i1.3410

Anttila, V. (2022). *The barriers in the employment of international students in Vaasa*. Vaasa: University of Vaasa Press.

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Baron, G., & Hartwig, K. (2020). Workplace experience of international students in Australia. *Journal of International Students, 10*(2), 8-9. https://doi.org/10.32674/jis.v10i2.1946

Barrie, S. C. (2006). Understanding what we mean by the generic attributes of graduates. *Higher Education, 51*, 215-241. https://doi.org/10.1007/s10734-004-6384-7

Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*. Chicago: University of Chicago Press.

Behle, H. (2020). Students' and graduates' employability. A framework to classify and measure employability gain. *Policy Reviews in Higher Education, 4*(1), 105-130. https://doi.org/10.1080/23322969.2020.1712662

Bennett, D. (2019). Graduate employability and higher education: Past, present and future. *HERDSA Review of Higher Education, 5*, 31-61.

Berquist, B., Hall, R., Morris-Lange, S., Shields, H., Stern, V., & Tran, L. (2019). Global perspectives on international student employability. *International Education Association of Australia, 7*, 1-44.

Blackmore, J., Gribble, C., Farrell, L., Rahimi, M., Arber, R., & Devlin, M. (2014). *Australian international graduates and the transition to employment*. Final Report. Melbourne: Deakin University Australia Press.

Blackmore, J., Gribble, C., & Rahimi, M. (2017). International education, the formation of capital and graduate employment: Chinese accounting graduates' experiences of the Australian labor market. *Critical Studies in Education, 58*(1), 69-88. https://doi.org/10.1080/17508487.2015.1117505

Borg, J., & Scott-Young, C. M. (2020). Employers' perspectives on work readiness in construction: Are project management graduates hitting the ground running?. *International Journal of Managing Projects in Business, 13*(6), 1363-1379. https://doi.org/10.1108/IJMPB-10-2019-0238

Brunello, G., & Wruuck, P. (2019). Skill shortages and skill mismatch in Europe: A review of the literature. *Journal of Economic Surveys, 35*(4), 1145-1167 http://dx.doi.org/10.2139/ssrn.3390340

Calonge, D. S., Connor, M., Hultberg, P., Shah, M. A., & Aguerrebere, P. M. (2022). Contactless higher education: A SWOT analysis of emergency remote teaching and learning during COVID-19. *Journal of Educational Studies and Multidisciplinary Approaches, 2*(1). https://doi.org/10.51383/jesma.2022.22

Calonge, D. S., & Shah, M. A. (2016). MOOCs, graduate skills gaps, and employability: A qualitative systematic review of the literature. *The International Review of Research in Open and Distributed Learning, 17*(5), 1-24. https://doi.org/10.19173/irrodl.v17i5.2675

Cameron, R., Farivar, F., & Coffey, J. (2019). International graduates host country employment intentions and outcomes: Evidence from two Australian universities. *Journal of Higher Education Policy and Management, 41*(5), 550-568. https://doi.org/10.1080/1360080X.2019.1646383

Casas Trujillo, J., Mohammed, P. & Saleh, S. (2020). Students' motivations to study abroad: The case of international students at the University of Debrecen. *Central European Journal of Educational Research, 2*, 76-81. https://doi.org/10.37441/CEJER/2020/2/1/5760

Chai, D., Van, H., Wang, C., Lee, J., & Wang, J. (2020). What do international students need? The role of family and community supports for adjustment, engagement, and organizational citizenship behaviour. *Journal of International Students, 10*(3), 571-589. https://doi.org/10.32674/jis.v10i3.1235

Chew, J. (2019). *Economic opportunities and outcomes of* *post-study work rights in Australia*. Melbourne: International Education Association of Australia.

Cho, H., Levesque-Bristol, C., & Yough, M. (2021). International students' self-determined motivation, beliefs about classroom assessment, learning strategies, and academic adjustment in higher education. *Higher Education, 81*, 1-21. https://doi.org/10.1007/s10734-020-00608-0

Clarke, V., Braun, V., & Hayfield, N. (2015). Thematic analysis. *Qualitative psychology: A practical guide to research methods, 3*, 222-248.

Coffey, J., Farivar, F., & Cameron, R. (2021). The job seeking experiences of international graduates in the host country: Australia's lost opportunity?. *The International Journal of Human Resource Management, 32*(4), 846-870. https://doi.org/10.1080/09585192.2018.1504106

Cranmer, S. (2006). Enhancing graduate employability: Best intentions and mixed outcomes. *Studies in Higher Education, 31*(2), 169-184. https://doi.org/10.1080/03075070600572041

DAAD. (2023). *Internationale Studierende als Fachkräfte von morgen. Positions papier des DAAD.* Bonn: German Academic Exchange Service.

Daniels, J., & Brooker, J. (2014). Student identity development in higher education: Implications for graduate attributes and work-readiness. *Educational Research, 56*(1), 65-76. https://doi.org/10.1080/00131881.2013.874157

De Lange, P., O'Connell, B. T., Tharapos, M., Beatson, N., & Oosthuizen, H. (2022). Accounting graduate employability: Employer perspectives on skills and attributes of international graduates. *Accounting Education,* 1-29. https://doi.org/10.1080/09639284.2022.2059383

Desbiens, N. A., & Vidaillet, H. J. (2010). Discrimination against international medical graduates in the United States residency program selection process. *BMC Medical Education, 10*, 1-5. https://doi.org/10.1186/1472-6920-10-5

Di Pietro, G., & Urwin, P. (2006). Education and skills mismatch in the Italian graduate labour market. *Applied Economics, 38*(1), 79-93. https://doi.org/10.1080/00036840500215303

Dos Santos, L. M. (2021). Motivations and career decisions in occupational therapy course: A qualitative inquiry of Asia-Pacific international students in Australia. *Advances in Medical Education and Practice, 13*, 825-834. https://doi.org/10.2147/AMEP.S288885

Fakunle, O., & Pirrie, A. (2020). International students' reflection on employability development opportunities during a one-year master-level program in the UK. *Journal of International Students, 10*(2), 86-100. https://doi.org/10.32674/jis.v10iS2.2719

Fix, B. (2018). *The trouble with human capital theory* (No. 2018/07). Working Papers on Capital as Power.

Griffin, M., & Coelhoso, P. (2019). Business students' perspectives on employability skills post internship experience: Lessons from the UAE. *Higher Education, Skills and Work-Based Learning, 9*(1), 60-75. https://doi.org/10.1108/HESWBL-12-2017-0102

Guruz, K. (2011). *Higher education and international student mobility in the global knowledge economy: Revised and updated second edition.* Suny Press.

Han, Y., Gulanowski, D., & Sears, G. J. (2022). International student graduates' workforce integration: A systematic review. *International Journal of Intercultural Relations, 86,* 163-189. https://doi.org/10.1016/j.ijintrel.2021.11.003

Hatzenbuhler, N. J., & Klein, J. E. (2019). Educational preparation for clinical practice: Reflections of newly graduated RNs. *Nurse Educator, 44*(2), 93-97. https://doi.org/10.1097/NNE.0000000000000550

Herrmann, A. M., Zaal, P. M., Chappin, M. M. H., Schemmann, B., & Lühmann, A. (2023). "We don't need no (higher) education" - How the gig economy challenges the education-income paradigm. *Technological Forecasting and Social Change, 186*, 122136. https://doi.org/10.1016/j.techfore.2022.122136

Hillage, J., & Pollard, E. (1998). Employability: Developing a framework for policy analysis. *Labor Market Trends, 107,* 83–84.

Huang, R., & Turner, R. (2018). International experience, universities support and graduate employability–perceptions of Chinese international students studying in UK universities. *Journal of Education and Work, 31*(2), 175-189. https://doi.org/10.1080/13639080.2018.1436751

Hurley, P. (2023, May 23). Why unis and vocational colleges are key to Australia's temporary migration challenge. *The Conversation.* https://theconversation.com/why-unis-and-vocational-colleges-are-key-to-australias-temporary-migration-challenge-205011?

Jackson, D. (2017). Exploring the challenges experienced by international students during work-integrated learning in Australia. *Asia Pacific Journal of Education, 37*(3), 344-359. https://doi.org/10.1080/02188791.2017.1298515

Jackson, D. (2014). Testing a model of undergraduate competence in employability skills and its implications for stakeholders. *Journal of Education and Work, 27*(2), 220-242. https://doi.org/10.1080/13639080.2012.718750

Jackson, D., & Bridgstock, R. (2021). What actually works to enhance graduate employability? The relative value of curricular, co-curricular, and extra-curricular learning and paid work. *Higher Education, 81*(4), 723-739. https://doi.org/10.1007/s10734-020-00570-x

Jackson, D., & Bridgstock, R. (2018). Evidencing student success in the contemporary world-of-work: Renewing our thinking. *Higher Education Research & Development, 37*(5), 984-998. https://doi.org/10.1080/07294360.2018.1469603

Jackson, D., & Pham, T. (2021). International students and work-integrated learning: Overcoming challenges and looking to the future. In *Advances in research, theory and practice in work-integrated learning* (pp. 179-190). New York: Routledge.

Jamilah, A., Haque, I., Muhammad, F., Harun, G., Chowdhury, A., Akramuzzaman, M., Arafat, Y., & Kabir, R. (2020). Depression and associated factors among international students in a private university of Bangladesh. *Global Psychiatry, 3*(2) 55-61. https://doi.org/10.52095/gp.2020.1406

Khanal, J., & Gaulee, U. (2019). Challenges of international students from pre-departure to post-study: A literature review. *Journal of International Students, 9*(2), 560-581. https://doi.org/10.32674/jis.v9i2.673

Kivinen, O., & Ahola, S. (1999). Higher education as human risk capital. *Higher Education, 38*(2), 191-208. https://doi-org.ezproxy.lancs.ac.uk/10.1023/A:1003788929925

Krippendorff, K. (2011). *Computing Krippendorff's alpha-reliability*. https://repository.upenn.edu/asc_papers/43.

Krishnan, I. A., Ramalingam, S., Kaliappen, N., Uthamaputhran, S., Suppiah, P. C., Mello, G. D., & Paramasivam, S. (2021). Graduate employability skills: Words and phrases used in job interviews. *Australian Journal of Career Development, 30*(1), 24–32. https://doi.org/10.1177/1038416220980425

Lee, D. C. A., Haines, T. P., Maneephong, S., & Zeng, Q. (2019). Barriers to fieldwork placements for international higher degree students: A systematic literature review. *Australian Journal of Career Development, 28*(3), 197-211. https://doi.org/10.1177/10384162198496

Li, D., & Lindo, N. (2022). Acculturation moderating between international students' career decision-making difficulties and career decision self-efficacy. *The Career Development Quarterly, 70*, 38-51. https://doi.org/10.1002/cdq.12284

Li, H. (2020). How to Retain Global Talent? Economic and Social Integration of Chinese Students in Finland. *Sustainability, 12*(10), 1-19. https://doi.org/10.3390/su1210461

Mansour, B. & Dean, J. C. (2016). Employability skills as perceived by employers and university faculty in the fields of human resource development (HRD) for entry level graduate jobs. *Journal of Human Resource and Sustainability Studies, 4*(1), 39-49. https://doi.org/10.4236/jhrss.2016.41005

Marginson, S. (2019). Limitations of human capital theory. *Studies in Higher Education, 44*(2), 287–301. https://doi.org/10.1080/03075079.2017.1359823

Mathies, C., & Karhunen, H. (2021). Do they stay or go? Analysis of international students in Finland. *Globalisation, Societies and Education, 19*(3), 298-310. https://doi.org/10.1080/14767724.2020.1816926

Mattoo, A., Neagu, I. C., & Özden, Ç. (2008). Brain waste? Educated immigrants in the US labor market. *Journal of Development Economics, 87*(2), 255-269. https://doi.org/10.1016/j.jdeveco.2007.05.001

Mehdizadeh, N., & Scott, G. (2005). Adjustment problems of Iranian international students in Scotland. *International Education Journal, 6*(4), 484-493.

Minutillo, S., Cleary, M., Hills, A., & Visentin, D. (2020). Mental health considerations for international students. *Issues in Mental Health Nursing, 41*(6), 494-499. https://doi.org/10.1080/01612840.2020.1716123

Mocanu, C., Zamfir, A. M., & Pirciog, S. (2014). Matching curricula with labor market needs for higher education: state of art, obstacles and facilitating factors. *Procedia - Social and Behavioral Sciences, 149*, 602–606. https://doi.org/10.1016/j.sbspro.2014.08.234

Mok, K. H., & Han, X. (2016). From 'brain drain'to 'brain bridging': Transnational higher education development and graduate employment in China. *Journal of Higher Education Policy and Management, 38*(3), 369-389. https://doi.org/10.1080/1360080X.2016.1174409

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology, 18,* 1-7. https://doi.org/10.1186/s12874-018-0611-x

Nachatar Singh, J. K. (2023). Unequal opportunities in accessing employability-related programmes: South Asian graduates' experiences in Australia. *Higher Education Quarterly*. https://doi.org/10.1111/hequ.12416

Nguyen, T., & Hartz, D. (2020). International students in Australia, employability and cultural competence. In J. Frawley, G. Russell, & J. Sherwood, J (Eds.), *Cultural competence and the higher education sector: Australian perspectives, policies and practice* (pp. 331-348). Singapore: Springer.

Orr, P., Forsyth, L., Caballero, C., Rosenberg, C., & Walker, A. (2023). A systematic review of Australian higher education students' and graduates' work readiness. *Higher Education Research & Development,* 1-18. https://doi.org/10.1080/07294360.2023.2192465

Osmani, M., Weerakkody, V., Hindi, N., & Eldabi, T. (2019). Graduates employability skills: A review of literature against market demand. *Journal of Education for Business, 94*(7), 423-432. https://doi.org/10.1080/08832323.2018.1545629

Parkinson, M., Howe, J., & Azarias, J. (2023). *Review of the migration system, final report*. Melbourne: Australian Government, Department of Home Affairs.

Pekerti, A., Van de Vijver, F., Moeller, M. & Okimoto, T. (2020). Intercultural contacts and acculturation resources among International students in Australia: A mixed-methods study. *International Journal of Intercultural Relations, 75*, 56-81. https://doi.org/10.1016/j.ijintrel.2019.12.004

Pham, T., & Jackson, D. (2020). Employability and determinants of employment outcomes. In T. Pham, T., & D. Jackson (Eds.), *Developing and utilizing employability capitals* (pp. 237-255). Routledge.

Pham, T., & Saito, E. (2020). Career development of returnees: Experienced constraints and navigating strategies of returnees in Vietnam. *Journal of Further and Higher Education, 44*(8), 1052-1064. https://doi.org/10.1080/0309877X.2019.1647333

Pham, T., Tomlinson, M., & Thompson, C. (2019). Forms of capital and agency as mediations in negotiating employability of international graduate migrants. *Globalisation, Societies and Education, 17*(3), 394-405. https://doi.org/10.1080/14767724.2019.1583091

Pollock, D., Tricco, A. C., Peters, M. D., McInerney, P. A., Khalil, H., Godfrey, C. M., Alexander, L., & Munn, Z. (2022). Methodological quality, guidance, and tools in scoping reviews: A scoping review protocol. *JBI Evidence Synthesis, 20*(4), 1098-1105. https://doi.org/10.11124/JBIES-20-00570

Pouratashi, M., & Zamani, A. (2019). University and graduates employability: Academics' views regarding university activities (the case of Iran). *Higher Education, Skills and Work-Based Learning, 9*(3), 290-304. https://doi.org/10.1108/HESWBL-12-2017-0103

Rae, D., & Woodier-Harris, N. (2012). International entrepreneurship education: Postgraduate business student experiences of entrepreneurship education. *Education+ Training*. https://doi.org/10.1108/00400911211274800

Römgens, I., Scoupe, R., & Beausaert, S. (2020). Unraveling the concept of employability, bringing together research on employability in higher education and the workplace. *Studies in Higher Education, 45*(12), 2588-2603. https://doi.org/10.1080/03075079.2019.1623770

Salas-Velasco, M. (2021). Mapping the (mis)match of university degrees in the graduate labor market. *Journal for Labour Market Research, 55*, 14. https://doi.org/10.1186/s12651-021-00297-x

Santandreu Calonge, D., Aman Shah, M., Riggs, K., & Connor, M. (2019). MOOCs and upskilling in Australia: A qualitative literature study. *Cogent Education, 6*(1), 1687392. https://doi.org/10.1080/2331186X.2019.1687392

Schwendimann, R., Blatter, C., Dhaini, S., Simon, M., & Ausserhofer, D. (2018). The occurrence, types, consequences and preventability of in-hospital adverse events–a scoping review. *BMC Health Services Research, 18*(1), 1-13. https://doi.org/10.1186/s12913-018-3335-z

Scott, C., Safdar, S., Trilokekar, R. D., & El Masri, A. (2015). International students as 'ideal immigrants' in Canada: A disconnect between policy makers' assumptions and the lived experiences of international students. *Comparative and International Education, 43*(3), 1-18. https://doi.org/10.5206/cie-eci.v43i3.9261

Shams, R., & Thrassou, A. (2019). Theorization and industry-based research project development: Bridging the industry–academia research gap. *Industry and Higher Education, 33*(2), 79–82. https://doi.org/10.1177/0950422218797919

Singh, J. K. N. (2020). Why do Chinese international students studying in Australia repatriate? Australian chinese graduates tell it all. *Journal of Further and Higher Education, 44*(9), 1207-1219. https://doi.org/10.1080/0309877X.2019.1669771

Smaldone, F., Ippolito, A., Lagger, J., & Pellicano, M. (2022). Employability skills: Profiling data scientists in the digital labour market. *European Management Journal, 40*(5), 671–684. https://doi.org/10.1016/j.emj.2022.05.005

Small, L., McPhail, R., & Shaw, A. (2022). Graduate employability: The higher education landscape in Australia. *Higher Education Research & Development, 41*(3), 919-933. https://doi.org/10.1080/07294360.2021.1877623

Sofat, S. (2021). *Barriers of international graduates seeking employment after graduation: A systematic review* (Doctoral dissertation, Auckland University of Technology). https://openrepository.aut.ac.nz/items/9b3a3dff-0fdf-4220-88f8-29cab959d594.

Song, K., & Kim, H. (2022). Measuring perceived influences of immigration rules on career exploration among international students. *Journal of College Student Development, 63*(5), 572-587. https://doi.org/10.1353/csd.2022.0048

Succi, C., & Canovi, M. (2020). Soft skills to enhance graduate employability: comparing students and employers' perceptions. *Studies in Higher Education, 45*(9), 1834-1847. https://doi.org/10.1080/03075079.2019.1585420

Tomlinson, M. (2012). Graduate employability: A review of conceptual and empirical themes. *Higher Education Policy, 25*, 407-431. https://doi.org/10.1057/hep.2011.26

Tran, L. T., Tan, G., Bui, H., & Rahimi, M. (2023). International graduates on temporary post-graduation visas in Australia: Employment experiences and outcomes. *Population, Space and Place, 29*(1), e2602. https://doi.org/10.1002/psp.2602

Tran, L. T., Bui, H., Tan, G., & Rahimi, M. (2022a). "It is not ok to think that you are good just because you have graduated from overseas": Agency and contextual factors affecting Vietnamese returning graduates. *International Migration, 60*(6), 43-59. https://doi.org/10.1111/imig.12954

Tran, L. T., Bui, H., Tan, G., & Rahimi, M. (2022b). Post-graduation work visas and loopholes: Insights into support provision for international graduates from the perspectives of migration agents, universities, and international graduates. *Evaluation Review, 45*(4), 438-464. https://doi.org/10.1177/0193841X221098211

Tran, L. T., Phan, H. L. T., Tan, G., & Rahimi, M. (2022c). 'I changed my strategy and looked for jobs on Gumtree': The ecological circumstances and international graduates' agency and strategies to navigate the Australian labour market. *Compare: A Journal of Comparative and International*

Education, 52(5), 822-840. https://doi.org/10.1080/03057925.2020.1837613

Tran, L. T., Rahimi, M., Tan, G., Dang, X. T., & Le, N. (2020). Post-study work for international graduates in Australia: Opportunity to enhance employability, get a return on investment or secure migration?. *Globalisation, Societies and Education, 18*(5), 495-510. https://doi.org/10.1080/14767724.2020.1789449

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467-473. https://doi.org/10.7326/M18-0850

Tymon, A. (2013). The student perspective on employability. *Studies in Higher Education, 38*(6), 841-856. https://doi.org/10.1080/03075079.2011.604408

Wang, Z. (2020). Bonding and bridging social capital: The determinants of English language fluency and its effects on the labour market outcome of international students in Ireland. *The Economic Social Review, 51*(1), 35-61.

Weilage, C., & Maráz, G. (2022). Online study's influence on international student employability factors in Germany: Germany vs. overseas based students. *Journal of Teaching in International Business, 33*(1), 7-30. https://doi.org/10.1080/08975930.2022.2089441

Winterton, J., & Turner, J. J. (2019). Preparing graduates for work readiness: An overview and agenda. *Education + Training, 61*(5), 536-551. https://doi.org/10.1108/ET-03-2019-0044

Wut, T. M., Xu, J., & Sum, C. K. (2022). Staying local or going back: A study on international university graduates' mobility. *PLOS ONE, 17*(8), e0268821. https://doi.org/doi.org/10.1371/journal.pone.0268821

Yorke, M. (2004). Employability in the undergraduate curriculum: Some student perspectives. *European Journal of Education, 39*(4), 409-427. https://doi.org/10.1111/j.1465-3435.2004.00194.x

Yorke, M., & Knight, P. T. (2006). *Embedding employability into the curriculum*. https://www.heacademy.ac.uk/resource/embedding-employability-curriculum

Zainuddin, S. Z. B., Pillai, S., Dumanig, F. P., & Phillip, A. (2019). English language and graduate employability. *Education+ Training, 61*(1), 79-93. https://doi.org/10.1108/ET-06-2017-0089

Zhao, J., Chapman, E., Houghton, S., & Lawrence, D. (2022). Perceived discrimination as a threat to the mental health of Chinese international students in Australia. *Frontiers in Education, 7*, 726614. https://doi.org/10.3389/feduc.2022.726614

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : **2591-801X**

Content Available at : **http://journals.sfu.ca/jalt/index.php/jalt/index**

An analysis of the learning styles in online environments of graduate students studying distance education

Gökhan Alptekin[A]

Yusuf Yıldırım[B]

Hakan Altınpulluk[C]

Onur Yumurtaci[D]

[A] *PhD candidate, Distance Education Department, Anadolu University*

[B] *PhD candidate, Distance Education Department, Anadolu University*

[C] *Associate Professor, Distance Education Department, Anadolu University*

[D] *Assistant Professor, Distance Education Department, Anadolu University*

Keywords

Distance education;
e-learning styles;
Master's program;
online learning;
quantitative cross-sectional screening;
Turkey.

Correspondence

hakanaltinpulluk@anadolu.edu.tr [C]

Article Info

Received 4 May 2023
Received in revised form 9 July 2023
Accepted 15 July 2023
Available online 20 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.11

Abstract

This research was conducted with the purpose of analyzing the learning styles in online environments of students in the Anadolu University Institute of Social Sciences distance education non-thesis Master's program. To this end, a quantitative cross-sectional screening model was applied to a total of 271 students in the distance education non-thesis Master's degree program. The data for the study was gathered online using the "E-Learning Styles Scale for Electronic Environments". Some of the findings of the study are as follows: (1) The learning styles in online environments of students do not show statistically significant differences based on sex, income, and average daily use of technological devices. (2) Age appears to have a high level of influence on the visual and aural learning levels of students in online environments and a medium level of influence on their active learning levels. (3) Students who are retired have lower levels of audiovisual learning and active learning compared to students in other vocational groups. (4) As the technology use efficacies of students increase, their logical learning levels in online environments increase. (5) Students who use technological devices for an average of seven or more hours per day have higher independent learning levels in online environments compared to those who use them between 0-3 hours.

Introduction

One of the most important factors that influence the learning of students in online learning environments, along with the effectiveness and efficiency of an online class, is learning styles (Birdal, 2022; Fatahi et al., 2016; Kurnaz & Ergün, 2019; Mutluay, 2018; Uçar, 2022). Kolb (1984) defined learning style as an individual and independent way of learning based on the requirements of the environment in which learning will take place, along with learning experiences acquired through previous learning processes. Learning styles are important for students in order for them to adapt their own cognitive, affective, psychomotor skills and learning experiences to the activities they are expected to execute throughout the online course process (Gülbahar & Alper, 2014). Another definition of learning styles that emphasizes this importance is that it is an indicator of how a student perceives, processes, understands, interprets and memorizes information and is influenced by intellectual, physical, emotional, social, mental, environmental, and cultural factors (Kadam et al., 2021). Learning style is an individual difference that influences the learning requirements and preferences of students throughout the process of acquiring, processing and interpreting information which differentiates them from other students (Şimşek, 2004).

One of the most significant individual differences that influence learning processes while supporting the academic achievements and learning permanence of students is learning styles (Arslan & Uslu, 2014; Fatahi et al., 2016; Kadam et al., 2021; Şimşek, 2004). Yurdal et al. (2021) state that online learning environments are better than face-to-face learning environments for students with different learning styles. Learning in online learning environments, within the capabilities of distance education, takes place in different learning styles and, more significantly, at the pace of the learners themselves (Moore & Kearsley, 2012). From this perspective, it may be stated that learning styles are highly important for learning itself (Özonur et al., 2020). When online learning environments are designed in accordance with the learning styles of students, the motivation, joy, and participation of learners increases, their learning develops (Latham et al., 2012), their academic achievement increases (Kurnaz & Ergün, 2019), and a more effective learning experience is provided (Özonur et al., 2020).

Based on the definitions and findings of the literature on online learning styles provided above, it is understood that as an individual and independent way of learning, learning styles are an individual difference that influences the quality, effectiveness and efficiency of learning experiences and activities to be conducted by students in learning environments while it increases their learning motivations, participation in learning activities and academic achievement.

In situations where learning takes place in online learning environments rather than face- to-face learning environments, the learning experiences to be executed by students will change depending on the opportunities and facilities presented to the student by the online learning environments (Oktay, 2022). The changing roles of students in face-to-face and online learning environments may cause differences in students' learning styles (Badge et al., 2012). Therefore, students will feel the need to develop different learning styes in online learning environments compared to face-to-face learning environments (Özonur et al., 2020). Based on this requirement expressed in the literature, a new definition has been adopted: e-learning styles are learning characteristics that aid students in effectively using the information they require with their own unique learning method in online learning environments (Gülbahar & Alper, 2014).

Gülbahar and Alper (2014) stated that e-learning styles of students may be listed as follows: audiovisual learning, where students learn best through seeing and hearing; logical learning, where students learn through problem solving resulting in detailed and deep thought; independent learning, where students learn individually at their own pace; intuitional learning, where students learn by association with feelings and emotions; verbal learning in which learning takes place through reading; social learning in which interaction is established with other students and learning takes place collaboratively within group work; and active learning in which students learn by doing, living, and experiencing.

One of the significant ways of increasing effectiveness, efficiency and quality in online learning is to design the online learning environments in accordance with the e-learning styles of the students (Birdal, 2022). In online learning environments where learners are responsible for their own learning, determining the learning styles of students assist in discovering the strengths and weaknesses of their learning experiences and makes them prone to learn easily and permanently (Dağ & Geçer, 2009). Thus, designing online learning environments taking students' e-learning styles would increase the effectiveness of personalized educational programs (Yurdal et al., 2021).

Students' learning styles should be determined and analyzed, and the learning processes and environments should be planned and designed based on their learning styles (Evin-Gencel, 2007). Therefore, it is important that when adaptive online learning environments uniquely differentiated by students' learning styles are being designed, the e-learning styles of students are known, and the online learning environments are differentiated in accordance with these e-learning styles (Oktay, 2022).

This study focuses on the e-learning styles of non-thesis Master's degree students studying through distance learning at the Anadolu University Institute of Social Sciences. Students take online courses in virtual classrooms on the Canvas Learning Management System (LMS) in the evening hours under the guidance of an instructor. The classes start between six and ten p.m. and are conducted by turning on the cameras and microphones of the instructor and students. Since these students are mostly employed, these classes are held in the evening hours. Online courses are usually conducted with a lecture by the instructor and a question-and-answer session at the end of the lecture.

This review of the literature revealed no study on determining the e-learning styles of graduate students in online environments. Within the scope of this study, data was gathered from graduate students studying in the distance education non-thesis program distance learning setting, and the gap in the literature may be filled to a degree.

Research purpose

The purpose of this research is to analyze the learning styles of non-thesis Master's students studying at a distance in online learning environments regarding different variables. To achieve this goal, answers to the following research questions were sought:

1. Do the e-learning styles of students in the Distance Education Non-Thesis Master's Degree programs vary based on gender?
2. Do the e-learning styles of students in the Distance Education Non-Thesis Master's Degree programs vary based on age?
3. Do the e-learning styles of students in the Distance Education Non-Thesis Master's Degree programs vary based on occupation?
4. Do the e-learning styles of students in the Distance Education Non-Thesis Master's Degree programs vary based on monthly income?
5. Do the e-learning styles of students in the Distance Education Non-Thesis Master's Degree programs vary based on technological competence?
6. Do the e-learning styles of students in the Distance Education Non-Thesis Master's Degree programs vary based on the average daily use duration of technology?

Method

The study was conducted using the quantitative method of descriptive research. Descriptive research is a method used when a subject is to be studied as is in order to determine the current apparent status (Karakaya, 2014). In this method, a current situation is explained as carefully as possible, and relationships between events are determined (Büyüköztürk et al., 2014). This method attempts to define the subject of the research by evaluating individuals, events or objects within their own current circumstances (Karasar, 2012).

Research design

This study was conducted in order to analyze different variables of the online environment learning styles of students studying at the Anadolu University Institute of Social Sciences Distance Education Non-Thesis Master's Degree program. One of the general screening models, a cross-sectional screening model, was used in the study. Screening models, which are an integral part of the descriptive method, are ways of organizing a population or sample to gain a general impression regarding a population when the population consists of many elements (Karasar, 2012). Cross-sectional screening models, however, deal with large sample sizes containing individuals with different qualities. In this model, the variables within the study that are to be described are measured all at once (Büyüköztürk et al., 2014; Fraenkel & Wallen, 2006). In this study, in order to determine the online learning styles of non-thesis Master's students studying via distance education based on different variables in a single pass, a cross-sectional screening model was used.

Study group

The population of the study consisted of students studying in the Anadolu University Institute of Social Sciences Distance Education Non-Thesis Master's program during the 2022-2023 educational year. The sample consisted of 271 students who responded to the data gathering tool distributed to all of the students of the program. The non-random method of convenience sampling was used when establishing the sample group of the study. Based on the principles of availability and accessibility, this sampling method saves time and cost to the researcher allowing for rapid data gathering (Büyüköztürk et al., 2014). Researchers who use this method work with voluntary participants (Erkuş, 2005). The demographic characteristics of the students who constitute the workgroup of the study are presented in Table 1.

Table 1. Demographic information of students.

Feature	Variable	N	%
Sex	Female	146	53.9
	Male	125	46.1
Age	18-34	142	52.4
	35-54	121	44.6
	55-64	8	3.0
Occupational Status	Unemployed	34	12.5
	Public Employee	121	44.6
	Private Sector Employee	93	34.3
	Retired	9	3.3
	Self-Employed	14	5.2
Monthly Income	5500 TRY and below	36	13.3
	5501-9999 TRY	82	30.2
	10000 TRY and above	153	56.5
Technology Use Competence	Basic Level	27	10.0
	Medium Level	136	50.1
	Advanced Level	108	39.9
Average Daily Usage Time of Technological Devices	0-3 hours	42	15.5
	3-5 hours	70	25.8
	5-7 hours	73	26.9
	7 and above hours	86	31.8

Data collection tools

Data were gathered for this study using a personal information form and the E-Learning Styles for Electronic Environments Scale' developed by Gülbahar and Alper (2014). The data gathering was conducted electronically. The electronic questionnaire created using Google Forms was distributed to the students in the Anadolu University Institute of Social Sciences Distance Education Non-Thesis Master's programs between November 14, 2022, and January 5, 2023. The electronically created questionnaire was sent to the students' e-mail addresses using shortened links. The voluntary participants were limited to a single response to the questionnaire, and the necessary information was presented to the participants in advance. Care was taken to avoid a biased sample group of participants. Data from participants who did not express their consent of free and voluntary participation were considered false and omitted from the study. A pilot study with 68 participating students was conducted to test the validity and reliability of the data-gathering tools. The pilot study was also conducted in a similar fashion using Google Forms, while the main study utilized data from a total of 271 participants.

Personal information form

The personal information form was created to determine the demographic characteristics of participants, such as sex, age, occupation, income, technology competence, and daily technology use duration.

E-learning styles for electronic environments scale

This measure used in the study was developed by Gülbahar and Alper (2014) and consists of seven sub-factors: "Audiovisual Learning", "Verbal Learning", "Active Learning", "Social Learning", "Independent Learning", "Logical Learning", and "Intuitional Learning". The scale, as a whole, measures the learning styles of individuals in online environments. The scale consists of 38 items and seven sub-factors and is of the 5-point Likert type. Items 1 through 8 measure the audiovisual learning levels of students while items 9-15 measure their verbal learning, 16-21 measure their active learning, 22-27 measure their social learning, 28-31 measure their social learning, 32-34 measure their logical learning, and 35-38 measure their intuitional learning levels. Exploratory factor analysis (EFA) was conducted to test the reliability and construct validity of the scale. Prior to the EFA, the fit of the data was tested, and a Kaiser-Meyer-Olkin (KMO) value of 0.960 was calculated, while the Bartlett test of sphericity was statistically significant ($p<0.01$). The EFA did not reveal any unloaded factor, and 18 of the factors with loading under 0.30 were omitted from the analysis. Next, confirmatory factor analysis (CFA) was conducted, and the analysis revealed that the scale in question could be successfully applied to the students. Additionally, the reliability of the scale was tested using Cronbach's Alpha internal consistency test, and a value of $\alpha= 0.94$ was observed for the whole scale. Furthermore, the Cronbach's Alpha values for the seven sub-factors that the scale consists of were determined to be between 0.72 and 0.87.

Since this study was conducted on a different population, the construct validity and fit values were determined again using CFA. The scale was confirmed on a separate group of students with similar characteristics prior to being applied to the main sample. Using AMOS 21.0 (Analysis of Moment Structures) software, the CFA revealed a good fit statistical value of corrected chi-square $\chi2/sd = 1.471$. Kline (2011) states that a value between $0 \leq \chi2/sd \leq 2$ indicates perfect fit. Therefore, the value obtained for this study presents a good fit value. Additionally, RMSEA was calculated as a separate measure of fit. The analysis provided an RMSEA value of 0.042, while the literature indicates a value between .00 and .05 would provide a perfect fit interval (Browne & Cudech, 1993). Studying other goodness of fit indexes resulted in a Comparative Fit Index value calculation of 0.907. Baumgartner and Homberg (1996), and Bentler and Bonett (1980) stated that a value of $.90 \leq CFI \leq .95$ is an acceptable fit measurement. As such, the value calculated for this study was also deemed acceptable. Another goodness of fit index that was calculated was the Tucker-Lewis Index. This value was calculated to be TLI = 0.904. Byrne (1994) stated that this value must be at least 0.90, indicating that the TLI value obtained from the CFA is acceptable. The incremental fit index was determined to be IFI = 0.909. Bollen (1989) stated that a value above 0.90 for this index indicates a good fit. Within this study, the adjusted goodness of fit index was also calculated, resulting in a value of AGFI = 0.850. Shermelleh-Engel and Moosbrugger (2003) stated that the acceptable range of values for this index is $.85 \leq AGFI \leq .90$, resulting in an acceptable value for this study. Lastly, the Standardized Root Mean Squared Residual Value was calculated to be SRMR = 0.068, with Hu and Bentler (1999) stating that a value below .080 is the requirement for a good fit. To determine the reliability of the scale used in the study, the Cronbach's Alpha (α) value of the internal consistency test was conducted, resulting in an internal consistency coefficient of α=0,807 and reliability coefficients for the sub-factors of the scale were all greater than 0.70.

Ethical statement

The E-Learning Styles in Electronic Environments Scale used in the study was developed by Gülbahar and Alper (2014). The required permission for the use of this scale in this study was obtained from the researchers via e-mail, and the study was conducted with the approval of the Anadolu University Humanities and Social Sciences Research and Publication Ethics Committee number E-54380210-050.99-432702 dated 27 October 2022.

Data analysis

The data gathered electronically for the study was first input into the Microsoft Excel spreadsheet program to organize it in order for the data to be successfully analyzed by IBM's SPSS 26.0 software. The organized data were coded in accordance with their responses to the personal information form and the e-learning styles for electronic environments scale and input into SPSS. A total of 296 participants in the voluntary questionnaire were identified. However, 25 of these participants did not provide their explicit consent to the questionnaire and were therefore considered invalid and removed from the study. Thus, 271 questionnaires were included in the analysis after verifying that all the data was correctly entered and normality analysis was conducted. The kurtosis, skewness, z scores and histograms of the data were analyzed to determine whether or not normal distribution was achieved. With a sample size between 50 and 300, z scores should not exceed 3.29 (Kim, 2013). The z scores of the data set were found to be below 3.29, the kurtosis and skewness values were within the -1/+1 interval, and the histogram indicated normal distribution (Huck, 2012). In order to determine the correlation levels between the percentages and scale variables of the data set, various measurement techniques were implemented sequentially, such as frequency analysis, independent samples t-test, and one-way analysis of variance (ANOVA). Parametric and non-parametric tests were conducted after the data set was confirmed to have a normal distribution. The kurtosis and skewness value calculations, determination of z scores, frequency analysis, independent samples t-test, one-way ANOVA and other parametric and non-parametric tests were conducted using SPSS, whereas due to the different populations being analyzed, the CFA was conducted using AMOS 21.0.

During data analysis, a high number of groups causes an increase in the margin of error. Therefore, in order to regulate the alpha value, Bonferroni correction was conducted prior to the multiple comparison tests. Bonferroni correction is a statistical correction conducted with a binary combination formula being applied to the significance coefficient/ group number (Vialatte & Chchocki, 2008). Therefore, the corrected alpha coefficients in multiple comparison tests are calculated to be 0.05/3=0.016 for groups of 3, 0.05/6=0.008 for groups of 4, and 0.05/10=0.005 for groups of 5. These new significance coefficients were utilized as measurements in the multiple comparison tests conducted in the study. To determine the effect sizes of the significant differences, Cohen's d values and eta-squared (η2) values were calculated (Cohen, 1988a; 1988b). In the analysis tables of the sub-factors of the scale used in the study, the sub-factors were summarized in the table as audiovisual, verbal, active, social, independent, logical, and intuitional. These refer to the following sub-factors of students' learning levels respectively: audiovisual learning levels, verbal learning levels, active learning levels, social learning levels, independent learning levels, logical learning levels, and intuitional learning levels. The total sum of the sub-factors that consist the scale measures students learning styles in online learning environments.

Results

This section of the study presents the statistical analyses conducted in order to determine the online learning styles of students in the distance education non-thesis Master's program. The findings are presented as tables and interpreted further. Independent samples t-test was conducted to measure any significant difference between the sex of students and their learning styles. The results of that analysis are presented in Table 2.

Table 2. Differentiation of learning styles of students in online environments based on sex.

Sub-factor/ Scale	Sex	N	X̄	sd	T	df	p
Audiovisual	Female Male	146 125	4.121 4.098	.5198 .5461	.364	269	.539
Verbal	Female Male	146 125	3.607 3.610	.6099 .6572	-.034	269	.473
Active	Female Male	146 125	3.455 3.092	.8054 .8863	3.53	269	.163
Social	Female Male	146 125	3.679 3.686	.8343 .8703	-.072	269	.794
Independent	Female Male	146 125	3.808 3.884	.7422 .6813	-.087	269	.185
Logical	Female Male	146 125	3.121 3.346	.9238 .9767	-1.95	269	.717
Intuitional	Female Male	146 125	3.429 3.304	.7991 .8023	1.28	269	.846
Total	Female Male	146 125	3.667 3.618	.4682 .4950	.822	269	.823

Based on the information presented in Table 2, the learning styles of students in online learning environments did not statistically significantly differ based on sex: audiovisual learning ($t(269)= 0.364$, $p>0.05$), verbal learning ($t(269)= -0.034$, $p>0.05$), active learning ($t(269)= 3.530$, $p>0.05$), social learning ($t(269)= -0.072$, $p>0.05$), independent learning ($t(269)= -0,087$, $p>0.05$), logical learning ($t(269)= -1.950$, $p>0.05$) and intuitional learning ($t(269)= 1.280$, $p>0.05$). Similarly, the total scores of the students in online environments obtained from the learning styles scale did not result in statistically significant differentiation based on sex as a variable ($t(269)= 0.822$, $p>0.05$). This situation indicates that sex is not an influential variable in the learning styles of students in online learning environments. The result of the analysis did not reveal any significant difference, and therefore Cohen's d value was not recorded.

One-way ANOVA was conducted to determine if statistically significant differences existed between students' online environment learning styles and age. The findings of this test are presented in Table 3.

Table 3. Differentiation of learning styles of students in online environments based on age.

Sub-factor/ Scale	Variables (ages)	N	X̄	Sd	df	F	p	Difference
Audiovisual	18-34 35-54 55-64	142 121 8	4.194 4.096 2.843	.4628 .5099 .3644	270	29.80	**.000***	18-34 > 55-64 35-54 > 55-64
Verbal	18-34 35-54 55-64	142 121 8	3.586 3.643 3.482	.6315 .6289 .7037	270	.430	.651	-
Active	18-34 35-54 55-64	142 121 8	3.478 3.147 2.020	.7849 .8730 .4833	270	15.19	**.000***	18-34 > 35-54 18-34 > 55-64 35-54 > 55-64
Social	18-34 35-54 55-64	142 121 8	3.642 3.736 3.583	.9102 .7682 .9677	270	.462	.630	-
Independent	18-34 35-54 55-64	142 121 8	3.950 3.721 3.781	.6679 .7536 .6870	270	3.467	.033	-
Logical	18-34 35-54 55-64	142 121 8	3.293 3.148 3.166	.8863 1.032 .8728	270	.766	.466	-
Intuitional	18-34 35-54 55-64	142 121 8	3.376 3.390 3.000	.8114 .8009 .5976	270	.896	.410	-
Total	18-34 35-54 55-64	142 121 8	3.699 3.617 3.088	.4555 .4910 .4071	270	6.741	**.001***	18-34 > 55-64 35-54 > 55-64

$*p<0.016$

Studying the findings in Table 3 shows a statistically significant difference in the learning styles of students in online learning environments based on age ($F(2.270)=[6.741]$, $p<0.016$). Thus, it may be stated that young and middle-aged students have higher levels of learning in online learning environments compared to students of older ages. An analysis of the sub-factors of the scale indicated significant differences in audiovisual learning levels and active learning levels depending on their ages ($F(2.270)=[29.80]$, $p<0.016$; $F(2.270)=[15.19]$, $p<0.016$). In order to determine the source of this difference, first, a Levene test was conducted. The results of the Levene test showed that the requirement of homogenous variances was satisfied. In order to determine which groups were the source of the statistically significant differences, Tukey's range test (Tukey's Honestly Significant Difference – HSD) was applied. The test results indicated that students aged 18-34 ($\bar{X}$= 4.194, sd= .4628) had higher levels of audiovisual learning than those aged 55-64 ($\bar{X}$= 2.843, sd= .3644). Similarly, students aged 35-54 ($\bar{X}$= 4.096, sd= .5099) also had higher audiovisual learning levels compared to those aged 55-64 ($\bar{X}$= 2.843, sd= .3644). No difference was found between students in the young and middle age groups regarding audiovisual learning. Additionally, students aged 18-34 ($\bar{X}$= 3.478, sd= .7849) were found to have higher levels of active learning compared to students aged 35-54 ($\bar{X}$= 3.147, sd= .8730) and 55-64 ($\bar{X}$= 2.020, sd= .4833). Similarly, students aged 35-54 ($\bar{X}$= 3.147, sd= .8730) had higher active learning levels than students aged 55-64 ($\bar{X}$= 2.020, sd= .4833). This finding supports the notion that as age reduces, active learning levels of students in online learning environments increases. In order to determine the effect sizes of the differences obtained in the test, an analysis of the eta-squared ($\eta 2$) values was chosen. The literature in the field indicates values between 0 and 0.01 as very small effects, 0.01 and 0.06 as small effects, 0.06 and 0.14 as medium effects, and values above 0.14 as large effects regarding effect size ranges (Cohen, 1988b). In this regard, the effect size of age on the audiovisual learning levels of students in online environments was found to be large ($\eta 2$= 0.181), and medium on active learning levels ($\eta 2$= 0.101).

One-Way ANOVA test was conducted to determine whether or not students' learning styles in online learning environments differentiated based on occupation, and the results of the test are presented in Table 4.

The ANOVA test results presented in Table 4 were studied, revealing a statistically significant difference in the learning styles of students in online learning environments based on occupation ($F(2.270) = [4.885]$, $p<0.005$). Similarly, significant differences were recorded in the audiovisual learning and active learning sub-factors. In order to determine the source of the significant differences in both the whole of the online learning differences scale and the sub-factors, Tukey's HSD multiple comparison test was conducted. The analysis revealed that retired students ($\bar{X}$= 3.064, sd= .4019) differed in their online learning styles compared to other students. An analysis of the sub-factors revealed that retired students ($\bar{X}$= 3.097, sd= .6428) had lower levels of audiovisual learning compared to other occupational groups, and a similar situation was observed for active learning and retired students ($\bar{X}$= 2.222, sd= .8036). No statistically significant difference was observed with the remaining sub-factors.

Table 4. Differentiation of learning styles of students in an online environment based on occupation.

Sub-Factor/ Scale	Variables (Occupation)	N	X	sd	df	F	p	Difference
Audiovisual	Unemployed	34	4.220	.4502	270	10.76	.000*	Unemployed> Retired Public S.> Retired Private S.> Retired Freelance> Retired
	Public Sector	121	4.116	.5203				
	Private Sector	93	4.186	.4661				
	Retired	9	3.097	.6428				
	Freelance	14	3.937	.4896				
Verbal	Unemployed	34	3.836	.6041	270	2.401	.050	-
	Public Sector	121	3.628	.6298				
	Private Sector	93	3.569	.6348				
	Retired	9	3.349	.5895				
	Freelance	14	3.316	.5719				
Active	Unemployed	34	3.480	.8058	270	4.223	.002*	Unemployed> Retired Public S.> Retired Public S.> Retired Freelance> Retired
	Public Sector	121	3.305	.8714				
	Private Sector	93	3.320	.8213				
	Retired	9	2.222	.8036				
	Freelance	14	3.131	.8143				
Social	Unemployed	34	3.759	.8868	270	.902	.463	-
	Public Sector	121	3.672	.8868				
	Private Sector	93	3.740	.7568				
	Retired	9	3.296	.8849				
	Freelance	14	3.452	.9987				
Independent	Unemployed	34	4.022	.5271	270	1.541	.191	-
	Public Sector	121	3.754	.8036				
	Private Sector	93	3.924	.6550				
	Retired	9	3.638	.6627				
	Freelance	14	3.767	.6237				
Logical	Unemployed	34	3.245	.8696	270	.755	.556	-
	Public Sector	121	3.305	.9809				
	Private Sector	93	3.157	.9588				
	Retired	9	2.814	.8992				
	Freelance	14	3.190	.9310				
Intuitional	Unemployed	34	3.375	.7468	270	1.509	.200	-
	Public Sector	121	3.431	.8391				
	Private Sector	93	3.387	.7995				
	Retired	9	3.027	.5651				
	Freelance	14	2.964	.6419				
Total	Unemployed	34	3.773	.4207	270	4.885	.001*	Unemployed> Retired Public S.> Retired Private S.> Retired Freelance> Retired
	Public Sector	121	3.654	.5168				
	Private Sector	93	3.672	.4147				
	Retired	9	3.064	.4019				
	Freelance	14	3.439	.4845				

$p<0.005$

In order to determine the effect size of these significant differences, eta-squared ($\eta 2$) values were noted. Analysis of these values indicated that occupation had a medium effect size ($\eta 2$= 0.139) on the audiovisual learning levels of students in online environments, while the effect size on active learning levels was small ($\eta 2$= 0.059). To determine if the learning styles of students in online environments differed based on income, an ANOVA test was conducted, and the results of the test are presented in Table 5.

Table 5. Differentiation of online learning styles based on monthly income.

Sub-Factor/ Scale	Variables	N	X	sd	df	F	p	Difference
Audiovisual	5500 TRY and below	36	4.197	.4690	270	.829	.438	-
	5500-9999 TRY	82	4.062	.6064				
	10000 TRY and above	153	4.116	.5016				
Verbal	5500 TRY and below	36	3.857	.6296	270	4.170	.016*	5500 TRY and below > 10000 TRY and above
	5500-9999 TRY	82	3.644	.6370				
	10000 TRY and above	153	3.531	.6152				
Active	5500 TRY and below	36	3.518	.7916	270	2.570	.078	-
	5500-9999 TRY	82	3.363	.9619				
	10000 TRY and above	153	3,192	.8100				
Social	5500 TRY and below	36	3.768	.9234	270	.319	.727	-
	5500-9999 TRY	82	3.634	.9043				
	10000 TRY and above	153	3.688	.8042				
Independent	5500 TRY and below	36	3.680	.6697	270	1.760	.174	-
	5500-9999 TRY	82	3.942	.8082				
	10000 TRY and above	153	3.828	.6660				
Logical	5500 TRY and below	36	3.055	.8600	270	.905	.406	-
	5500-9999 TRY	82	3.191	.9399				
	10000 TRY and above	153	3.283	.9813				
Intuitional	5500 TRY and below	36	3.388	.8544	270	.041	.960	-
	5500-9999 TRY	82	3.387	.8714				
	10000 TRY and above	153	3.359	.7537				
Total	5500 TRY and below	36	3.730	.4651	270	.801	.450	-
	5500-9999 TRY	82	3.655	.5652				
	10000 TRY and above	153	3.619	.4331				

$p<0.016$

Table 5 presents data indicating there was no statistically significant differentiation in the learning styles of students in online environments and income level (p>0.016). A study of the sub-factors revealed a significant difference in students' verbal learning levels in online environments and their income levels (F(2.270)= [4.170], p<0.016). Tukey's HSD was conducted to determine the source of this difference, resulting in students with a monthly income of 5500 TRY and below ($\bar{X}$= 3.857, ss= .6296) having higher verbal learning levels in online environments compared to students with 10000 TRY and above of monthly income ($\bar{X}$= 3.531, sd= .6152). When the eta-squared (η2) values of the observed significant difference are calculated to determine the effect size, the value was found to be η2=0.030 indicating a small effect size. On the other hand, no significant difference was found regarding income influencing the other sub-factors of the scale.

Another ANOVA test was conducted to determine whether a statistically significant difference existed between the learning styles of students in online environments and their technology use competencies. The results of the test are presented in Table 6.

Table 6. Learning style differentiation based on technology use competencies.

Sub-Factor/ Scale	Variables (Competence Levels)	N	X	sd	df	F	P	Difference
Audiovisual	Basic	27	3.717	.7194	270	8.969	.000*	Intermediate > Basic Advanced > Basic
	Intermediate	136	4.132	.4945				
	Advanced	108	4.181	.4828				
Verbal	Basic	27	3.576	.7202	270	.054	.948	-
	Intermediate	136	3.618	.6264				
	Advanced	108	3.604	.6190				
Active	Basic	27	2.864	1.003	270	3.701	.026	-
	Intermediate	136	3.338	.8500				
	Advanced	108	3.330	.8161				
Social	Basic	27	3.530	.8709	270	.922	.399	-
	Intermediate	136	3.745	.7721				
	Advanced	108	3.642	.9339				
Independent	Basic	27	3.481	.7169	270	7.813	.001*	Advanced > Basic
	Intermediate	136	3.773	.7810				
	Advanced	108	4.020	.5679				
Logical	Basic	27	2.370	.6292	270	18.02	.000*	Intermediate > Basic Advanced > Basic Advanced >Intermediate
	Intermediate	136	3.166	.9694				
	Advanced	108	3.512	.8603				
Intuitional	Basic	27	3.037	.6992	270	4.816	.009*	Intermediate > Basic Advanced > Basic
	Intermediate	136	3.501	.8263				
	Advanced	108	3.291	.7646				
Total	Basic	27	3.324	.5452	270	7.023	.001*	Intermediate >Basic Advanced > Basic
	Intermediate	136	3.670	.4870				
	Advanced	108	3.692	.4266				

*p<0.016

Table 6 portrays whether or not the learning styles of students in online environments differ based on ICT us competence, where ICT competence was found to have a statistically significant influence on online learning styles (F(.270)=[7.023], p<0.016). A significant difference was also found when the sub-factors were analyzed. Tukey's HSD test was conducted in order to determine the source of these significant differences, revealing that students with intermediate ($\bar{X}$= 4.132, sd= .4945) and advanced ($\bar{X}$= 4.181, sd= .4828) technology use competencies had higher audiovisual learning levels compared to those with basic competencies ($\bar{X}$= 3.717, sd= .7194). This significant difference had a medium effect size (η2= 0.062). Students with advanced competencies in technology use ($\bar{X}$= 4.020, sd= .5679) had higher levels of independent learning compared to those with basic competencies ($\bar{X}$= 3.481, sd= .7169). This difference, however, had a smaller effect size (η2=0.055). A further significant difference was found in the logical learning sub-factor, where logical learning levels increased as technology competence increased, with a medium effect size (η2=0.118) being calculated for this correlation. Similarly, a significant difference was found where medium ($\bar{X}$= 3.501, sd= .8263) and advanced ($\bar{X}$= 3.291, sd= .7646) ICT competencies led to higher intuitional learning levels compared to students with basic ($\bar{X}$= 3037, sd= .6992) competencies, with a small (η2=0.034) effect size. To determine whether average daily technology use created a significant difference in the online learning styles of students, an ANOVA test was conducted and the findings are presented in Table 7.

Table 7. Differentiation of learning styles of students in online learning environments based on daily average use duration of technological devices.

Sub-Factor/ Scale	Variables (Use Time)	N	$\bar{X}$	sd	df	F	p	Difference
Audiovisual	0-3 hours	42	4.098	.5532	270	1.250	.292	-
	3-5 hours	70	4.075	.5948				
	5-7 hours	73	4.047	.5058				
	7 hours or more	86	4.199	.4821				
Verbal	0-3 hours	42	3.710	.6800	270	1.131	.337	-
	3-5 hours	70	3.655	.6353				
	5-7 hours	73	3.616	.5841				
	7 hours or more	86	3.515	.6397				
Active	0-3 hours	42	3.226	.9793	270	.439	.726	-
	3-5 hours	70	3.228	.8503				
	5-7 hours	73	3.376	.8174				
	7 hours or more	86	3.290	.8545				
Social	0-3 hours	42	3.781	.8035	270	.886	.449	-
	3-5 hours	70	3.545	.8796				
	5-7 hours	73	3.716	.7488				
	7 hours or more	86	3.717	.9249				
Independent	0-3 hours	42	3.452	.9127	270	8.643	.000*	7 hours or more > 0-3 hours
	3-5 hours	70	3.707	.6634				
	5-7 hours	73	3.962	.6047				
	7 hours or more	86	4.043	.6396				
Logical	0-3 hours	42	3.134	.9851	270	.422	.738	-
	3-5 hours	70	3.166	.8695				
	5-7 hours	73	3.237	.9852				
	7 hours or more	86	3.306	.9851				
Intuitional	0-3 hours	42	3.452	.9535	270	.209	.890	-
	3-5 hours	70	3.328	.8215				
	5-7 hours	73	3.373	.6973				
	7 hours or more	86	3.366	.7986				
Total	0-3 hours	42	3.627	.5799	270	.498	.684	-
	3-5 hours	70	3.591	.4713				
	5-7 hours	73	3.666	.4651				
	7 hours or more	86	3.679	.4509				

*p<0.008

A quick glance at the findings of the ANOVA test presented in Table 7 clearly shows no significant difference was found between students' online learning styles and average daily use time of technology (F(2.270)=[0.498], p<0.008). Regarding the sub-factors of the scale, only independent learning revealed a significant difference (F(2.270)= [8.643], p<0.008). To determine the source of this difference, firstly, a Levene test was conducted, resulting in the finding that the variance did not portray homogenous distribution, leading to the necessity for a Games-Howell test. As a result of the non-parametric post hoc multiple comparison, students who used technological devices for seven or more hours during the day were found ($\bar{X}$= 4.043, sd= .6396) to have higher levels of independent learning compared to those who only used them between 0-3 hours per day on average ($\bar{X}$= 3.452, sd= .9127). The effect size of this significant difference was found to be medium (η2= 0.088), while none of the other sub-factors of the scale presented any significant difference.

Discussion

This study analyzed the learning styles in online learning environments of students in a non-thesis Master's program studying through distance education regarding different variables. The analysis resulted in many findings regarding online learning styles.

Firstly, the online learning styles of the participants were analyzed to determine whether they significantly differed based on sex. The results indicate that online learning styles do not differ by sex, and similar findings emerged regarding the other sub-factors of which the scale of measurement consists. This shows that the distance education non-thesis Master's students may have a common learning style regardless of sex. The literature in the field reveals other scientific research that supports these findings (Arslan & Babadoğan, 2005; Birdal, 2022; Demir, 2015; Dikmen, 2020; Mutluay, 2018; Özgür, 2013; Yeşilyurt, 2014). Conversely, there are also studies which have found statistically significant differences in online learning styles of students and their sex (Dikbaş, 2006; Kuru, 2018; Özüdoğru, 2022; Şentürk, 2016; Şentürk & Ciğerci, 2018; Uçar, 2022; Yetiş, 2018).

When a significant difference in the learning styles of students in online environments based on age was sought, a meaningful effect was found between the ages of students and their learning styles (Arslan & Babadoğan, 2005). The findings were that young (18-34) and middle-aged (35-54) students had higher levels of audiovisual learning compared to older (55-65) students. The effect size for this significant difference was also found to be quite large. One consideration may be that changes in the sensory perception and cognition of students as they age may be the cause for this situation, causing differences between students. Additionally, the active learning levels of students aged 18-34 were higher than those of students aged 35-54 and 55-64, and the levels of middle-aged (35-54) students were also higher than the levels of students aged 55-64 indicating an increase in active learning with a decrease of age. This may be caused by the higher capacity to process information of younger individuals, an ability which may decline with age. The effect size of this significant difference was found to be medium, however, studies in the field also indicate no significant difference between age and online learning styles (Özgür, 2013).

A review of the findings regarding the occupation of the participating students shows that occupation causes significant differences in their online learning styles. Retired students were found to have lower levels of active learning compared to students in other occupational groups. The eta-squared (η2) values of these significant differences were studied in order to determine the effect size of this variable. The analysis revealed a medium effect size of occupation on audiovisual learning in online learning environments and a small effect size on active learning. This situation may once again be related to the fact that retired students tend to be older, which would draw parallels to the findings of the age variable.

No statistically significant correlation was found between the monthly incomes of the participating students and their online learning styles. The analysis of the sub-factors of the online learning styles scale revealed a statistically significant difference between monthly income and verbal learning levels. The multiple comparison test conducted to find the cause of this difference revealed that students making 5500 TRY or less had higher verbal learning levels than those earning 10000 TRY or more each month. This may be caused by the need for low-income students to use their verbal skills more in order to communicate and express their thoughts in their daily lives, further developing this learning ability. The eta-squared (η2) value of this difference was analyzed, and a small effect size between the two variables was found.

Another research question this study attempted to answer was whether the learning styles of students in online environments differed based on their competencies in using information and communication technologies. The results show a statistically significant difference in the learning styles of students in online environments and their ICT competencies. This difference was observed to take place in certain sub-factors of the online learning styles scale. The findings were that students with intermediate and advanced technology use competencies had a medium size effect on their online audiovisual learning styles compared to students with only basic competencies. Students with advanced ICT competencies also had higher levels of independent learning compared to those with basic competencies. However, the effect size of this correlation was small. A similar difference was found regarding logical learning levels in that an increase in ICT competence also led to an increase in this style of learning in online environments, with the effect size determined to be medium. Students with intermediate and advanced technology competencies also had higher intuitional learning levels than those with basic competencies. However, the effect size of this significant difference was small. Achieving a certain level of competence when using technology requires not only higher levels of learning skills but also constant active use of technology which is why it is believed that individuals who achieve this level of competence also increase their independent, logical, and intuitional learning levels over time.

Lastly, homogenous differentiation between the average daily technology use of students and their online learning styles was studied. The analysis concluded that the duration for which students used technological devices did not cause any changes in students' online learning styles (Kuru, 2018; Yetiş, 2018). Conversely, studies also exist indicating statistically significant differences in online learning styles based on how long students use technology throughout the day (Mutluay, 2018). Further analysis of the sub-factors only resulted in a statistically significant difference in independent learning. The finding was that students who used technology for seven or more hours a day had higher independent learning levels compared to those who only used them between 0-3 hours on average. No statistically significant difference was found in this regard between students who used technology 3-5 hours a day and 5-7 hours a day. This indicates that heavy (seven hours or more per day on average) ICT users have significant differences regarding their independent learning styles. This significant difference was found to have a medium effect size.

Limitations of the study

This study has various limitations. Being a study on the learning styles in online learning environments of non-thesis Master's degree students enrolled in distance education, one limitation may be the selection of students enrolled at the Anadolu University Institute of Social Sciences. This study is also limited to the e-learning styles in online environments scale. Additionally, it is limited by the variables being measured, namely sex, age, occupation, income, ICT competence, and daily average technology use duration. The self-reporting nature of the responses to the scale items during the data gathering process may also be considered a limitation. Lastly, the requirement for gathering data through an online environment such as Google Forms due to the global COVID-19 pandemic health crisis may be considered to be a limitation.

Recommendations

Various recommendations may be made based on the outcomes of the study. Studies with larger sample sizes of non-thesis Master's students' learning styles in distance education and online learning environments may be conducted. While this study was conducted on non-thesis distance educated Master's students, various other studies on thesis-required face-to-face Master's students and doctorate students may be recommended. Age appears to be an important factor in online learning styles, with young and middle-aged students having higher levels of audiovisual learning compared to older students. Similarly, retired students were found to have lower levels of audiovisual and active learning levels compared to other occupational groups. Training for older students regarding audiovisual practice and techniques may prove to be beneficial. Lastly, considering technology competence and use time appears to have a positive effect on various sub-factors of online learning styles, a moderated increase in the use of electronic devices such as computers, telephones, and tablets for the purpose of learning may be recommended.

Conclusions

The findings of this study analyzing the learning styles of non-thesis Master's degree students in online learning environments are presented sequentially below.

The learning styles of non-thesis Master's degree students studying through distance education did not differ based on gender. This finding led to the conclusion that non-thesis Master's degree students in distance education had a common learning style regardless of gender, and therefore gender-specific arrangements are not required in the instructional design of learning activities during the development of distance education programs.

The learning styles of non-thesis Master's degree students studying through distance education differed based on age. The findings revealed a decrease in audio-visual and active learning levels as age progressed. This led to the conclusion that age-based learning activities may be effective when designing distance education programs, which would consider the reduction in sensory perception and information processing capacity with age. This conclusion may be supplemented with learning activities that reduce cognitive load and appeal to the available visual, aural and affective perception levels of students in accordance with their ages.

The learning styles of non-thesis Master's degree students studying through distance education differed based on occupation. Similar (and obviously related) to the age variable, retired students were older than students of other occupations resulting in lower levels of audio-visual and active learning. As such, it was concluded that learning activities that reduce the cognitive load and appeal to the visual, aural and affective perception levels of retired students would be beneficial during the instructional design of distance education programs.

A statistically significant difference was found between the verbal learning levels and monthly incomes regarding the learning styles of distance education students in non-thesis Master's degree programs. This difference may be due to lower-income students needing to use their verbal skills to communicate and express their thoughts more frequently in their daily lives. Thus, scholarships and other financial aid may be offered to lower-income students to support their financial status, or they may be provided access to the technology they need. Other preventative measures may be taken, considering the learning styles of low-income students may be negatively impacted by their lack of access to technology.

Technology competence was a statistically significant variable that influenced the learning styles of distance education students in non-thesis Master's programs. Increased competence regarding technology resulted in a medium sized increase in audio-visual learning levels. Achieving a certain level of competence when using technology requires not only higher-level learning skills but also constant and active use of technology. Therefore, it may be stated that individuals who achieve this level of competence eventually also achieve higher levels of independent, logical, and intuitional learning. As such, during the instructional design of distance education programs, more technologically focused environments may be created for students with higher technology competencies. Learning activities may be arranged such that students may indulge in higher levels of interaction with other students, instructors, and content.

The average daily duration of technology use did not result in a significant difference in the learning styles of distance education students in non-thesis Master's degree programs. Despite this, the study revealed that students who used technology for seven hours or more each day had higher independent learning levels compared to those who used technology for only 0-3 hours a day. Extra attention may be paid to certain aspects to accommodate students with independent learning styles and high durations of technology use. Learning activities that take advantage of self-directed learning skills may be created while preparing learning activities.

References

Arslan, B., & Babadoğan, C. (2005). The relationship between 7th and 8th grade students' learning styles and academic achievement level, gender and age. *Eurasian Journal of Educational Research, 21*(14), 35-48. https://search.trdizin.gov.tr/tr/yayin/detay/50217/

Arslan, H., & Uslu, B. (2014). The relationship between learning styles and leadership orientations of pre-service teachers. *Education and Science, 39*(173), 340–355. http://egitimvebilim.ted.org.tr/index.php/EB/article/view/2305

Badge, J. L., Saunders, N. F., & Cann, A. J. (2012). Beyond marks: New tools to visualise student engagement via social networks. *Research in Learning Technology, 20*(1), 1-14. https://doi.org/10.3402/rlt.v20i0.16283

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing, 13*(2), 139-161. https://doi.org/10.1016/0167-8116(95)00038-0

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588. https://doi.org/10.1037/0033-2909.88.3.588

Birdal, E. (2022). *Investigation of the relationship between university students' e-learning styles and technology addiction*. Master Thesis. Yeditepe University, Graduate School of Educational Sciences, Istanbul, Turkiye.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*(3), 303-316. https://doi.org/10.1177/0049124189017003004

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In: Bollen, K.A., & Long, J.S. (Eds.), *Testing structural equation models* (pp. 136-162). Sage. https://doi.org/10.1177/0049124192021002005

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. & Demirel, F. (2014). *Bilimsel araştırma yöntemleri*. Ankara: Pegem.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. SAGE.

Cohen, J. (1988a). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers. https://doi.org/10.4324/9780203771587

Cohen, J. (1988b). *Statistical power analysis for the social sciences*. Erlbaum. https://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf

Dağ, F., & Geçer, A. (2009). Relations between online learning and learning styles. *Procedia-Social and Behavioral Sciences, 1*(1), 862-871. https://doi.org/10.1016/j.sbspro.2009.01.155

Demir, Ö. (2015). *Examining the readiness levels of students and instructors for e-learning: Hacettepe university education faculty example*. Master Thesis. Hacettepe University, Graduate School of Educational Sciences. Ankara, Turkiye.

Dikbaş, E. (2006). *Investigation of pre-service teachers' attitudes towards e-learning*. Master Thesis. Dokuz Eylül University, Graduate School of Educational Sciences. İzmir, Turkiye.

Dikmen, M. (2020). Investigation of university students' e-learning styles in terms of various variables. *Paper presented at the 2nd International symposium on social sciences and educational sciences,* Asos Publishing, Elazığ. https://kongre.akademikiletisim.com/files/usves2/usves2_tam_metin_kitapcigi.pdf

Erkuş, A. (2005). *Bilimsel araştırma sarmalı*. Ankara: Seçkin Publishing.

Evin-Gencel, İ. (2007). Adaptation of learning styles inventory-3 based on Kolb's experiential learning theory into Turkish. *Dokuz Eylul University The Journal of Graduate School of Social Sciences, 9*(2), 120-139. http://hdl.handle.net/20.500.12397/5475

Fatahi, S., Moradi, H., & Kashani-Vahid, L. (2016). A survey of personality and learning styles models applied in virtual environments with emphasis on e-learning environments. *Artificial Intelligence Review, 46*(3), 413-429. https://doi.org/10.1007/s10462-016-9469-7

Field, A. (2005). *Discovering statistics using spss: And sex and drugs and rock 'n' roll* (2th Edition). London: Sage Publications.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th Ed.). Mac Graw Hill, Inc.

Gülbahar, Y., & Alper, A. (2014). Development of e-learning styles scale for electronic environments. *Education and Science, 39*(171), 421-435. http://egitimvebilim.ted.org.tr/index.php/EB/article/view/2078

Huck, S. W. (2012). *Reading statistics and research* (6th ed.). Pearson Education. http://students.aiu.edu/submissions/profiles/resources/onlineBook/J5E3k4_Reading_Statistics_and_Research-_6th.pdf

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Kadam, P. R., Gaikwad, U. K., & Bhamre, P. D. (2021). Student centred learning using student's learning style. *Journal of Engineering Education Transformations, 34,* 391-395. https://doi.org/10.16920/jeet/2021/v34i0/157185

Karakaya, İ. (2014). Bilimsel araştırma yöntemleri. A. Tanrıöğren (Ed.), *Bilimsel Araştırma Yöntemleri içinde* (s. 57-83). Ankara: Anı Publishing.

Karasar, N. (2012). *Bilimsel araştırma yöntemi.* Nobel Publishing.

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics, 38*(1), 52-54. https://doi.org/10.5395/rde.2013.38.1.52

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). The Guilford Press. http://ndl.ethernet.edu.et/bitstream/123456789/74702/1/35.pdf

Kolb, D. (1984). *Experiential learning: Experience as The source of learning and development.* Prentice Hall. https://www.researchgate.net/publication/235701029_Experiential_Learning_Experience_As_The_Source_Of_Learning_And_Development

Kurnaz, F. B., & Ergün, E. (2019). Examining the relationship between learning styles and academic achievement in e-learning environments. *Journal of Theoretical Educational Science, 12*(2), 532-549. https://doi.org/10.30831/akukeg.407029

Kuru, O. (2018). Investigation of classroom teaching students' e-learning styles in terms of various variables. *Kahramanmaraş Sütçü İmam University Journal of Social Sciences, 15*(1), 105-118. https://dergipark.org.tr/tr/pub/ksusbd/issue/37007/372419

Latham, A., Crockett, K., McLean, D., & Edmonds, B. (2012). A conversational intelligent tutoring system to automatically predict learning styles. *Computers & Education, 59*(1), 95-109. https://doi.org/10.1016/j.compedu.2011.11.001

Moore, M., & Kearsley, G. (2012). *Distance education: A systems view of online learning.* Wadsworth.

Mutluay, Y. (2018). *Examination of the relationship between vocational college students' media and technology use and their e-learning styles.* Master Thesis, University of Kahramanmaraş Sütçü İmam, Graduate School of Natural and Applied Sciences, Kahramanmaraş, Turkiye.

Oktay, Ö. (2022). Uyarlanabilir Öğrenme Ortamları. T. V. Yüzer ve M. Kesim (Ed.). *Açık ve Uzaktan Öğrenmenin Dijital Dönüşüm Boyutu içinde* (s.352-377). Ankara: Pegem Akademi.

Özgür, H. (2013). Learning styles of distance education students: Trakya university sample. *Trakya University Journal of Education, 3*(2), 85-91. https://dergipark.org.tr/en/pub/trkefd/issue/21474/230169

Özonur, M., Kamışlı, H. & Solmaz, M. İ. (2020). Identifying distance learners' learning styles. *Elementary Education Online, 19*(3), 1858-1863. https://doi:10.17051/ilkonline.2020.735341

Özüdoğru, G. (2022). Preservice teacher's e-learning styles and attitudes toward e-learning. *İ.e.: Inquary in Education. 14*(1), 1-14. https://eric.ed.gov/?id=EJ1345423

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74. https://www.stats.ox.ac.uk/~snijders/mpr_Schermelleh.pdf

Şentürk, C. (2016). Investigation of the attitudes of teachers towards e-learning based on several variables. *Journal of International Social Research, 9*(43), 1501-1511. https://www.sosyalarastirmalar.com/abstract/investigation-of-the-attitudes-of-teachers-towards-elearning-based-on-severalrnvariables-74857.html

Şentürk, C., & Ciğerci, F. M. (2018). Investigation of classroom teachers' e-learning styles. *Kahramanmaraş Sütçü İmam University Journal of Social Sciences, 15*(1), 69-88. https://dergipark.org.tr/en/pub/ksusbd/issue/37007/336918

Şimşek, A. (2004). Öğrenme biçimi. Y. Kuzgun ve D. Deryakulu (Ed.), *Eğitimde Bireysel Farklılıklar içinde* (s. 95-136). Ankara: Nobel Yayıncılık.

Uçar, D. (2022). *Examination of science teacher candidates' attitudes towards social media and e-learning styles in terms of various variables.* [Unpublished Master's Thesis]. University of Hacettepe Graduate School of Educational Sciences.

Vialatte, F., & Cichocki, A. (2008). Split test Bonferonni correction for QEEG statistical maps. *Biological Cybernetics, 98*(4), 295-303. https://doi.org/10.1007/s00422-008-0210-8

Yeşilyurt, E. (2014). Determination of teachers' learning styles and evaluation of the relationship among learning styles. *Journal of Theory and Practice in Education, 10*(4), 999-1021. https://dergipark.org.tr/tr/pub/eku/issue/5462/74155

Yetiş, Ü. (2018). *The relationship between university students 'social media use habits and e-learning style.* Master Thesis, University of Kahramanmaraş Sütçü İmam, Graduate School of Natural and Applied Sciences, Kahramanmaraş, Turkiye.

Yurdal, M. O., Sahin, E. M., Aytuğ-Kosan, A. M., & Toraman, Ç. (2021). Development of medical school students' attıtudes towards online learning scale and its relationship with e-learning styles. *Turkish Online Journal of Distance Education, 22*(3), 310-325. https://doi.org/10.17718/tojde.961855

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

A systematic review of graduate training on cultural competence

David Freire[A] — [A] University of Virginia, USA

Aaliyah Churchill[B] — [B] University of Virginia, USA

Noelle Hurd[C] — [C] University of Virginia, USA

Keywords

Cultural competence;
discrimination;
graduate training;
mental health;
structural inequality.

Correspondence

df9gz@virginia.edu [A]

Article Info

Received 14 July 2023
Received in revised form 1 October 2023
Accepted 1 October 2023
Available online 5 October 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.31

Abstract

A systematic review was conducted to investigate scholarship from the last ten years regarding graduate training for the provision of culturally competent mental health care to individuals who hold marginalized identities (e.g., those marginalized based on their race, ethnicity, gender, sexual orientation). This review furthered a conceptualization of cultural competence that views clients as individuals embedded within their own cultures and communities while also recognizing the interplay of systems of power and oppression within an individual's life that create unique lived experiences. This was accomplished by conducting a systematic literature review following PRISMA guidelines. Seven databases (i.e., PsycINFO [EBSCO], PubMed, Psychology and Behavioral Sciences Collection [EBSCO], Academic Search Complete [EBSCO], SocIndex [EBSCO], Science Direct, ProQuest) were searched using a priori-defined search strings that encompass graduate training, cultural competence, and the various mental health care fields. Recommendations for improving cultural competence conceptualizations, engendering innovative training interventions, and increasing rigorous evaluation tools are provided.

Introduction

Graduate programs, governing boards, higher education faculty, and clinical directors play a crucial role in setting the stage for a foundation of training that emphasizes diversity and multicultural awareness within the training and skills instilled in their students in mental health training programs. Effective anti-racist and social justice-oriented training prepares clinicians to work effectively with marginalized communities. A social justice orientation, for the purposes of this review, is an ideology that reflects the pursuit of social, economic, and political equality with a basis on acceptance and celebration of difference and diversity (Craig, 2002). This involves culturally competent care, which considers the systems of oppression individuals must navigate while actively working to dismantle these systems that harm the clients one works with (Chung & Bemak, 2011; Matthew & Adams, 2009).

Cultural competence (with a particular focus on understanding systems of power, privilege, and oppression) among mental health service providers holds promise for improving the quality of mental health care provided to individuals who hold marginalized identities (Ali & Sichel, 2014). Within the United States, oppressive structures of power perpetuate unequal distribution of resources across various societal institutions, marginalizing many groups that fall out of the privileged majority (e.g., White, heterosexual, cisgender, middle class; Moradi, 2017). Despite this harrowing context, many mental health care workers lack the proper training and understanding to acknowledge how systems of power and domination uniquely affect the developmental and mental health outcomes of marginalized individuals.

Since their inception, the mental health care fields (e.g., psychology, psychiatry, social work) have neglected to include a diverse set of voices within training, research, and practice (Hall, 2014), and the underrepresentation of marginalized groups in the field has contributed to treatment that often times fails to account for the unique lived experiences of marginalized community members (Koç & Kafa, 2019). However, over the past 40 years, incorporating culturally competent care into treatment services for marginalized individuals has become a growing area of focus within graduate training programs (Sue et al., 2009). Despite growing attention to this construct, cultural competence within mental health care training and practice lacks a standardized definition (Benuto et al., 2018), and definitions that are frequently endorsed do not always account for the structural forces that define the lived experiences of oppression and marginalization that individuals must contend with on a daily basis (Danso, 2018). The current systematic review is inclusive of studies published within the past ten years that have focused on training graduate students in the mental health field on cultural competence.

Historical context

Historically, the mental health care fields' relationship with marginalized groups (e.g., people of color, people with disabilities, LGBTQ individuals) has been stigmatizing and oppressive. Ideologies that have centered ideas of White supremacy, misogyny, homophobia, and ableism have dominated research and practice. The mental health care field, alongside countless other fields and disciplines, have failed to actively resist these oppressive philosophies and instead have allowed them to permeate and persist (Constantine, 2007; Tasca et al., 2012; Tievsky, 1988). Consequently, across a number of instances, diagnoses and treatment have been disenfranchising through the advancement of discriminatory theories (e.g., Morton's theory of craniometry, idea of "feeble-mindedness"; Allen, 1984; Radford, 1991). Moreover, research in the field of mental health has been grounded in ideas of homophobia, white supremacy, and misogyny (Hall, 2014). For centuries, social welfare programs (the precursors to the field of social work) were complicit in the identification and removal of individuals who exhibited any kind of "abnormal behavior," with a privileged minority arbitrarily deciding which groups to other and oppress (Mackelprang & Salsgiver, 1996). While stigmatization originally concentrated on individuals with cognitive and physical disabilities (e.g., Buck vs. Bell; Lombardo, 1985), the mental health care field's use of diagnoses to marginalize individuals broadened to include gender minorities, racial/ethnic minorities, and sexual minorities. For the purpose of this review, "minority" does not describe a quantitative state of being but rather is used to describe groups that have been excluded from mainstream social, economic, and educational life.

Multicultural education

In order to understand the importance of multicultural education within mental health care training, it is essential to first define "culture." Most definitions of culture consist of describing it as a set of discrete behavioral norms and thought processes shared by individuals within a definable population that are distinct from those shared within other groups (Lehman et al., 2004). More contemporary definitions conceptualize culture as a system, specifically an interconnected relationship between peoples, places, and practices, for the ultimate purpose of enacting, justifying or challenging power within a social context (Causadias, 2020). Causadias (2020) defines people in this cultural model as referring to population dynamics, cultural groups, and social relations; places refer to institutional influences, ecological dynamics, and cultures within various contexts; practices refer to community engagement and culture being enacted. All three of these cultural components exist in mutual relationship with one another in order for certain groups to obtain power, the ability to force others into compliance, force others to behave as desired, and the ability to control access to spaces (Causadias, 2020).

An individual's culture is affected by the social context in which they inhabit, with sociocultural factors arising from the interconnected nature between social issues and cultural phenomena (Yamada & Brekke, 2008). These factors include a variety of social issues associated with minority status (e.g., immigration stress, racial discrimination), which culminate in defining how individuals within marginalized populations experience culture and oppression within the Western context. Conceptualizing culture as a system designed to enact power onto groups of individuals allows us to understand how different groups experience oppression within the U.S. context. Oppression, which operates as a cultural system, has negative effects on physical and psychological health outcomes and thus is pivotal to understanding how mental health is conceptualized for members of marginalized groups (Seaton et al., 2018).

Many concepts have arisen over the years that have attempted to capture the interplay between sociocultural experiences and individual mental health. Cultural sensitivity was posited as the awareness of cultural information and schemas that a client holds and incorporating this cultural information into one's own behavior and thought processes when interacting with clients (Kumpfer et al., 2002). Relatedly, the idea of cultural competence has been advanced to capture the awareness of culture and the application of this knowledge to diverse clients (Betancourt et al., 2003; Huey et al., 2014; Lakes et al., 2006; Whaley & Davis, 2007). Cultural humility, building off of both cultural competence and sensitivity, goes a step further by incorporating the importance of critical self-examination of one's own cultural awareness, openness to new cultural information, and emphasizing a lifelong motivation to learn from others (Hook et al., 2017; Mosher et al., 2017).

While many previous definitions have focused more broadly on an examination of difference and culture without an examination of structural and institutional forces, we aim to further the conceptualization of culture as a system in which individuals are both acted upon as well as agents themselves in relation to power structures. By viewing culture as a mutually interacting system designed to perpetuate various forms of subordination, we are able to integrate both cultural awareness and an examination of hegemonic power within the U.S. context that contributes to different lived experiences for marginalized populations (relative to privileged groups) that lead to negative physical and psychological outcomes (Causadias, 2020). This review, therefore, aims to highlight the importance of integrating a conceptualization of culture that goes beyond individual difference, but focuses on practitioners understanding that marginalized populations experience cultural systems that often can have deleterious effects on their development and everyday lives. Therefore, we will be using the definition espoused by the National Association of Social Workers (NASW) in 2015, which defines cultural competence as including: (1) the awareness of how diverse populations experience their uniqueness in a larger context; (2) an understanding of intersectionality that examines oppression, discrimination, and domination; and (3) a recognition of the individual's position of prerogative and entitlement in relation to the populations they serve and with a recognition of the need to exercise cultural humility (Lusk et al., 2017).

Marginalized individuals within the U.S. from different cultures and identity groups experience distinct social upbringings that collide with various oppressive forces, and these experiences shape behavior, cognition, and reaction patterns (Lehman et al., 2004). This can affect how one conceptualizes the formation and maintenance of mental disorders as well as an individual's willingness to receive psychotherapy, suggesting that failing to take into account systems of power and the resulting marginalized experiences of individuals may lead to poor treatment conceptualization and outcomes (Koç & Kafa, 2019). All individuals, regardless of their background, should be able to have access to responsive and effective treatment while having freedom from harm from incompetent providers.

Current models of multicultural training aim to move past conceptualizations that homogenize marginalized groups, though attempts to standardize treatment recommendations may reify stereotypes of marginalized groups and promote reductive stereotypes when teaching clinicians how to take into account the identities of marginalized individuals. Additionally, evaluation approaches to cultural competence training have been variable and inconsistent (Benuto et al., 2018; Curtis-Boles & Bourg, 2010; Merta et al., 1998; Roysicar et al., 2005; Stanhope et al., 2005). The diverse nature of current evaluation methods lacks standardization in both methodology and scope of what defines cultural competence and how to best measure it, complicating the ability of graduate programs to effectively determine if their training for students is effective in method and outcome.

Current study

Political landscapes defined by racist discourse that promote the establishment of oppressive policies have led to an increase in the visibility of state-sanctioned violence against various marginalized communities (Aymer, 2016; Grills et al., 2016). Oppression has become endemic to the U.S. context, and structures of power have allowed marginalizing policies and rhetoric to permeate the lived experiences of targeted groups (Anderson et al., 2022). This recent increase in salience of marginalization in the U.S. has led to increased stress and anxiety within marginalized groups, pushing the importance of cultural competency and multicultural education (Williams & Etkins, 2021). Relatedly, there has been an uptick in the development of cultural competence trainings that have been disseminated to graduate programs regarding best practices for instruction (Celinska & Swazo, 2021; Dameron et al., 2020; Ratts et al., 2016).

Culture, however, has been defined in a variety of ways, oftentimes failing to account for the sociohistorical context and power structures that marginalized individuals must contend with and often bring into the clinical setting. Educational systems often reproduce colonial power structures and fail to include the voices and experiences of marginalized individuals necessary to provide culturally competent clinical care (McLeod et al., 2020). Some argue that clinical care of marginalized groups should include more than just an understanding that they hold distinct marginalized identities (Hansen et al., 2018). Clinical care to marginalized groups could also center critical reflection on how individual psychopathology is engendered from interactions between biology, environment, and historically (and currently) marginalized realities; integrate this understanding into treatment provision; and consider opportunities to change structural realities to make society more equitable (Kirmayer et al., 2018). In order to map the evidence base and identify knowledge gaps and weaknesses within the current conceptualizations of cultural competence, a systematic literature review was employed, as this methodology has the potential to inform evidence-based policies and practice within clinical care (Mallet, 2012). The current review aims to take inventory of recent training and intervention methods aimed at increasing cultural competence (specific to the treatment of individuals who hold marginalized identities) among those training to become mental health care providers. This review also aims to document advances in the conceptualization of cultural competence specific to service provision to marginalized individuals.

Methods

The systematic review method is a rigorous and powerful tool to summarize the evidence base and identify gaps within the literature. Systematic reviews, however, can often face the challenge of subjective screening and quality appraisal methods used to assess the relevance and inclusion of articles and studies within the scope of a review (Mallet et al., 2012). Therefore, our approach was to adopt strict inclusion and exclusion criteria to minimize the chances of bias influencing the selection of articles in the final analysis. Inclusion criteria included: (a) the article concerned graduate training or topics related to graduate training and cultural competence, such as multicultural competence, learning, and cultural sensitivity (b) the article used original quantitative or qualitative data, or presented theoretical frameworks related to the topic (c) the article, if it was an empirical study, included graduate students, such as masters students, doctoral students, or other post-baccalaureate programs (d) the topic of multicultural competence training concerned individuals who held historically marginalized identities (e.g., racial/ethnic minorities, sexual minorities, gender minorities, low-income individuals) (e) the article focused on fields that concerned mental health care (i.e., psychology, psychiatry, social work) and (f) the article had undergone a peer-review process and was published in an empirical journal. Given that marginalization and oppression manifest differently across diverse global contexts and that training models for graduate education in mental health professions vary meaningfully across countries, we only included studies in our review that were conducted in the United States. Exclusion criteria included: (a) the study was conducted outside of the U.S., (b) the article concerned cultural competence training in a field outside of the mental health care fields, (c) the study sample was not comprised of graduate students, (d) the article did not focus on cultural competence training and (e) the article did not promote an understanding of cultural competence that accounted for systems of power, oppression, and social hierarchy. The review was conducted following the PRISMA guidelines (see Figure 1).

Seven databases (i.e., PsycINFO [EBSCO], PubMed, Psychology and Behavioral Sciences Collection [EBSCO], Academic Search Complete [EBSCO], SocIndex [EBSCO], Science Direct, ProQuest) were searched using an a priori-defined search string: (“cultural competence” OR “cultural competency” OR “cultural awareness” OR “cultural sensitivity” OR “multicultural competence” OR “multicultural competency”) AND (training OR education OR development OR learning) AND (“graduate students” OR “masters students” OR “doctoral students”) AND (“mental health” OR “psychology” OR “social work” OR “psychiatry”). Previous reviews, reference lists, and additional searches were also utilized to identify potential articles. We restricted our search to articles published in or after 2010, as this was the year of the last comprehensive review of multicultural competence training in the field of mental health that we were able to find (i.e., Rogers & O’Bryon, 2014). The search strategy produced 2914 articles. After duplicates were removed using Microsoft Excel, there were 1874 remaining articles to be screened using the article title and abstract. Of the articles screened 202 articles were identified as potentially relevant based on inclusion criteria. Then, the full-text article was retrieved to determine whether all inclusion criteria were met, removing an additional 149 articles. The final sample consisted of 53 articles.

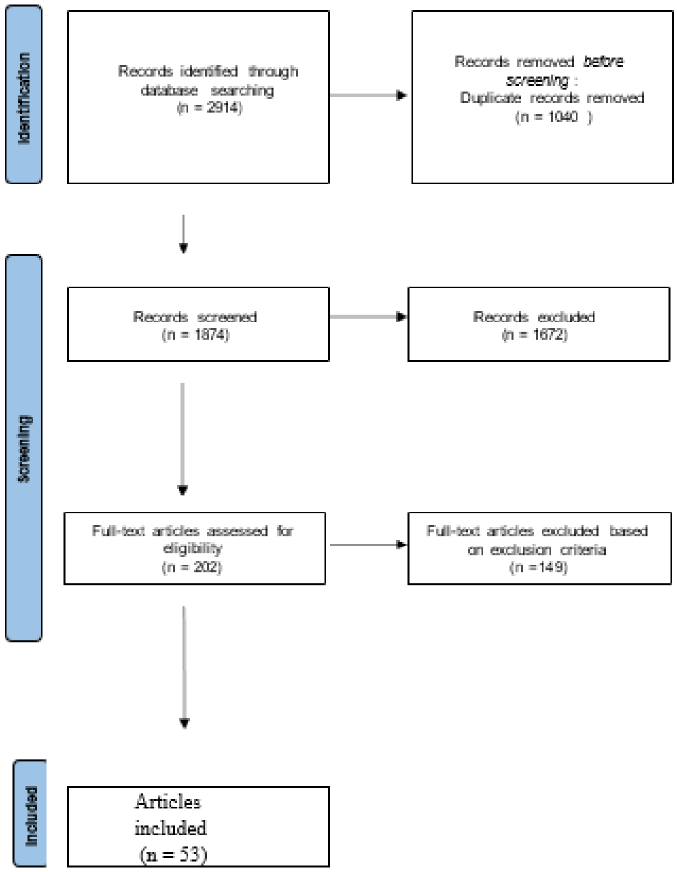


Figure 1. Prisma flow diagram.

Results

53 articles were included in this review. 30 articles (54%) were conceptual in nature and concentrated on presenting recommendations for clinical practice with individuals who hold various marginalized identities; introducing innovative theoretical conceptualizations of multicultural competence training; and understanding how standards for multicultural competencies set by governing bodies can best be incorporated into training. Two articles (5%) were descriptive in nature, focused on disseminating data regarding perceptions of student competence and correlates of multicultural competence. Finally, 21 articles (41%) were

evaluative in nature, focused on investigating the utility and effectiveness of various training interventions used to increase multicultural competency in graduate students. The systematic analysis yielded results that were organized into the following categories: 1) intergroup contact and cultural competence, 2) cultural humility and cultural competence, 3) intersectionality and cultural competence, and 4) antiracism and cultural competence. This review will continue to use the definition of cultural competence espoused by the NASW, which defines cultural competence as including: (1) the awareness of how diverse populations experience their uniqueness in a larger context; (2) an understanding of intersectionality that examines oppression, discrimination, and domination; and (3) a recognition of the individual's position of prerogative and entitlement in relation to the populations they serve and with a recognition of the need to exercise cultural humility (Lusk et al., 2017).

Intergroup contact and cultural competence

Intergroup contact, based on Allport's contact hypothesis (1954), argues that positive contact with an outgroup member can lead to positive attitudes toward the outgroup (Allport, 1954; Imperato et al., 2021). This kind of contact provides the basis for effective communication between groups which may lead to increased cultural knowledge, more accurate beliefs about the other, and an overall gained respect for the outgroup (Kormos et al., 2014). Intergroup contact, within clinical science, may offer a tool for clinicians in training to foster their cultural competence when didactic methods of instruction are insufficient or hold the potential to engender negative experiences for faculty members from marginalized backgrounds who are often uniquely tasked with being on-the-spot experts on multicultural topics (Dorn et al., 2020).

A review of the literature yielded scholars who presented various ways to conceptualize and implement intergroup contact into curriculum aimed at increasing cultural competence within graduate clinicians in training. Thibeault (2019) furthered the idea of intergroup contact for graduate students within the context of service learning with indigenous populations. Service learning, in this case, refers to an approach that implements learning objectives within community service in order to provide a valuable learning experience for students while meeting the needs of the community. Through this culturally immersive experience with another group, students may learn about another culture, increase their skills in cultural competency, and provide the community the assistance it may need in various projects such as building structures and creating gardens (Thibeault, 2019).

Relatedly, Fondacaro & Harder (2014) presented a training model called Connecting Cultures for promoting cross-cultural immersion and contact regarding working with refugee populations. Using a social justice framework, Fondacaro & Harder promoted a culturally sensitive context in which graduate students engage with refugee communities through community service and professional development to prepare them for intercultural contact. This model promoted the idea of cultural competence as a life-long endeavor instead of an acquired skill, tasking graduate students to continually challenge their beliefs and learn from others about their unique lived experiences by continuously seeking out opportunities to engage with different communities (Fondacaro & Harder, 2014). Lorelle et al. (2021) presented the use of cultural immersion as a way to prompt intergroup contact that can increase individual multicultural and social justice competencies. Lorelle and colleagues argued that seeking out opportunities to engage in collaboration and communication through intentionally putting one's self in a culturally different environment can spark discussion of privilege, oppression, and bias that can increase trainee cultural competence.

A review of the literature also yielded various ways in which graduate training can incorporate intergroup contact into clinician instruction, curriculum, and practice. Shannonhouse et al. (2018) investigated the use of cultural immersion and intergroup contact among graduate students in Council for Accreditation of Counseling and Related Educational Programs (CACREP) accredited programs. Establishing baseline data regarding how cultural immersion is conducted in counselor education, results from a survey disseminated to program directors indicated that nearly half of respondents' programs engaged in cultural immersion, though they varied widely in the degree to which they facilitated conditions for successful intergroup contact (e.g., providing sociohistorical context, sustained duration of immersion, diverse opportunities to interact with the community, and reflection), implemented evaluation of program effectiveness, and/or engaged students in reflection. Survey results additionally indicated that graduate programs may have a limited understanding of cultural immersion outcomes among their students (Shannonhouse et al., 2018).

Bolea (2012) presented a curricular approach that utilized service learning with indigenous populations as a cross-cultural immersion tool to build cultural competency and critical thinking. Preliminary qualitative findings through anecdotal reports and a course specific evaluation survey showed students reporting increases in empathy for indigenous communities, improved knowledge about the sociohistorical context of the community, and improved relational skills to be an effective social worker (Bolea, 2012). McDowell et al. (2012) examined the use of intergroup contact by investigating student experiences following engagement with an international study abroad course designed to increase cultural competence within counseling students. Following semi-structured interviews to gauge student experience and feedback, results showed increased social awareness, changes in world views, increased awareness of societal structures, recognition of one's own privilege, and enhanced contextual and systemic thinking.

Platt (2012) engaged marriage and family therapy students in a 5-week Spanish language and cultural immersion program focused on improving clinical service delivery to Latinx communities. The 5-week program consisted of critical dialogues with other students and within the unfamiliar culture, the use of online forums, daily Spanish language courses, and home-stay living arrangements with Mexican host families. Results from student qualitative interviews following the immersion program indicated an increased awareness of one's own culture, increased complexity in how students perceived Latinx culture, and reports of plans to incorporate knowledge into clinical work by tuning into contextual factors that may influence the community and monitoring one's own biases (Platt, 2012).

Dessel & Rodenbord (2017) conducted a study examining outcomes of a master's in social work (MSW) program that implemented an intergroup dialogue (IGD) course in an urban Midwest MSW program, which was based on Sue's multicultural education model. Pretest and posttest data was collected over the span of two semesters using a survey that measured three domains: social identity awareness, knowledge about inequality, and micro/macro social work skills. Results from paired t-test analyses indicated that students reported statistically significant improvement on several measures of cultural competence, including knowledge of racial inequality, understanding the causes for poverty and economic inequality, and motivation to bridge differences; however, students did not increase their comfort in communicating with people of other groups or dealing with conflict (Dessel & Rodenbord, 2017).

Parikh et al. (2020) implemented the use of digital storytelling as a way to immerse students in a culture different form their own. Students were tasked with engaging with a 10-hour mini-immersion experience in which they were tasked with interviewing members of a culture different from their own. Results from qualitative interviews revealed an increased awareness about other cultures, an increase in knowledge of other cultures and their own racial identity, and development of skills in multicultural competence.

Toporek & Worthington (2014) implemented a service-learning program called Project Homeless Connect (PHC), a service that offers employment counseling to individuals from low socioeconomic backgrounds while providing counseling training aimed at increasing cultural and structural competence around impoverished communities in counseling students. Results from qualitative interviews showed students reporting a broadening of their understanding of homelessness, increased self-efficacy from collaboration with community members, and increased confidence in applying counseling skills.

Killian & Floren (2020) explored the relationship between teaching methods and multicultural and social justice competencies within clinicians in training through the implementation of a quasi-experimental study that assigned students to different cultural competence courses. Results suggested that pedagogical approaches that utilized direct exposure, such as community service learning, best facilitated clinician self-reported cultural competence and readiness to be social advocates (measured as 6 levels of empowerment to enact change at the intrapersonal and global level) for marginalized populations, supporting the use of immersive experiences in promoting multicultural competence within clinicians in training.

Lee (2014) implemented a virtual community exercise in order to increase graduate clinician self-awareness of bias and engagement with issues of diversity in relation to marginalized communities through a simulated immersion exercise. A virtual community is a computer-generated display that allows users to feel present and interact with a simulated environment. Qualitative interviews as well as independent t-tests comparing student scores on the multicultural awareness, knowledge, and skills survey (MAKSS) were conducted to determine intervention effectiveness. Results indicated statistically significant increases in learning about diversity concepts, statistically significant increases in self-awareness of perception of issues related to marginalized individuals, and reports of understanding of issues related to one's own identity and community.

Cultural humility and cultural competence

Cultural humility is defined as critical self-reflection of how one's own culture and identities inform how clinicians view clients, emphasizing the role that systems of power and oppression have in shaping individuals' experiences (Abe, 2020). Cultural humility aims to go beyond definitions of cultural competence that only emphasize working across difference by also focusing on an awareness and knowledge of how societal structures and institutions have been organized in a way that engenders inequality (Fisher et al., 2015). Within clinical science, cultural humility may help broaden trainees' understanding of their marginalized clients' lived experiences by tasking clinicians in training to critically examine their own biases potentially increasing opportunities to minimize harm while striving to obtain accurate representations of client experience and psychopathology (Fisher et al., 2015).

Holyoak et al. (2020) attended to critical examination and self-reflection by presenting the importance of the clinician-in-training's "way of being" within family therapy. This "way of being" framework emphasized helping graduate trainees to conceptualize clients as fully human and complex (i.e., seeing clients as individuals with rights, desires, and cultures that have consequences for therapy outcomes), and viewing their clients from a culturally sensitive and humble space while reflecting on the cultural misunderstandings that may arise in treatment.

Stabb & Reimers (2013) spoke to the importance of cultural humility as it relates to working with low-income populations through their presentation of how the APA Training Council Benchmark Competencies can be integrated when conceptualizing graduate training for working with impoverished individuals. Stabb & Reimer argued that a clinician's professional identity should be driven by critical self-assessment and cultural humility. Accordingly, they recommended that clinicians in training be educated on the structural causes of class inequality while engaging in difficult dialogues (focused on issues of class and privilege) within the classroom that hold the potential to build clinician empathy for their marginalized clients.

Trinh et al. (2021) argued for the implementation of cultural humility within psychiatric education, arguing that lifelong engagement with self-evaluation and self-critique within the context of the client-patient relationship is necessary for understanding how systemic and social structures affect the lives of marginalized clients. By reflecting on personal sociocultural identities, biases, and assumptions, clinicians in training may be able to improve their understanding of diverse patient populations and improve the delivery of mental health care to marginalized individuals (Trinh et al., 2021). Venner & Verney (2015) acknowledged that critical self-reflection and engagement with difficult topics of privilege and oppression can oftentimes engender reluctance and hesitance within clinicians in training. Through the building of trust and rapport, instructors can create conversations that utilize open-ended questions to draw out students' perspectives and goals, aiding students in engaging with topics of multiculturalism within a safe and supportive environment (Venner & Verney, 2015).

Chan et al. (2018) addressed the importance of counselor educators assisting their students in understanding their own privilege by attending to systemic issues of power that have real mental health implications for marginalized individuals. Educators, according to Chan, should communicate to students that although they did not actively create the privilege or discrimination that other groups experience, it is their duty to understand how systems of oppression create sociopolitical forces that affect relationships, services, and advocacy that clinicians provide their clients. Relatedly, Estrada et al. (2013) identified student understanding of privilege as an avenue for promoting socially just practice within clinical work. They presented a student orientation training model involving a mandatory orientation for incoming counseling students, which consists of various icebreakers and group activities. They shared meals aimed at teaching individuals how to recognize their own privileges within an environment that normalizes these discussions and provides constructive feedback.

Dessel et al. (2019) advanced the need for critical self-reflection and cultural humility within the context of training conservative Christian counselors to engage with ideas of power and privilege when working with LGBTQ clients. They argued that faculty can prepare themselves to teach LGBTQ subject matters to graduate trainees by engaging in critical self-assessment and developing knowledge about power and privilege prior to classroom instruction. This may allow faculty to create an environment that empowers students to reflect on their own journeys toward LGBTQ cultural competence (Dessel et al., 2019).

Ridley et al. (2021) introduced the process model of multicultural competence, a framework that advances the intentional integration of cultural knowledge acquisition and understanding within the standard therapeutic goals. Ridley and colleagues argued that a process model of cultural competence promotes an ongoing commitment to cultural understanding, deviating from a content-focused model that may lead to stereotyping once it is "achieved." By continuously engaging in self-awareness of one's own limitations in understanding the complex systems of oppression and power that individuals must navigate, one can continuously strive for education and knowledge acquisition that can improve care for marginalized clients (Ridley et al., 2021).

Our review also yielded descriptions of applications of cultural humility into graduate training for increasing cultural competence when working with marginalized individuals. Tormala et al. (2018) documented how they translated the outline for cultural formulation, an outline within the DSM-V describing cultural factors that may influence client functioning and the therapeutic relationship, into an assignment to increase cultural humility in graduate students. Thematic analysis was then used on both cultural formulation assignments, yielding 6 themes that were analyzed for changes over time: 1) cultural self-awareness 2) intersectionality 3) perspective taking 4) unsupported cultural statements 5) scientific mindedness and 6) power/privilege differentials. Results from these analyses showed increased cultural self-awareness and decreased use of biased cultural statements (i.e., stereotypical or overgeneralized statements) at the end of their academic semester relative to their levels at the beginning of the semester.

Jones et al. (2016) implemented a pilot intervention aimed at increasing cultural competence, knowledge, and self-awareness of graduate clinicians in training. Students in a control group were trained in traditional cognitive behavioral therapy (CBT) frameworks, while separate treatment group received additional instruction that included cultural frameworks and treatment perspectives that were guided by culturally responsive care specific to working with marginalized clients. Results indicated little change for the CBT only group, though the CBT + cultural responsiveness training group showed improvements in cultural competence over time.

Jones & Lee (2021) introduced the use of a one-semester course following Sue et al.'s three domains of cultural competency (awareness, knowledge, skills) in order to increase student cultural competency for working with marginalized individuals. The required multicultural course comprised building student self-awareness through engagement with ideas of privilege, oppression, and color-evasive racism. Results from a pretest-posttest design study utilizing the self-assessment of multicultural awareness, knowledge, and skill (SAMAKS) indicated statistically significant increases across all three areas of cultural competency within clinical trainees with large effect sizes at the completion of the course (Jones & Lee, 2021).

Hoke & Robbins (2011) presented an educational approach that was implemented with beginning graduate psychiatric nursing students with the goal of promoting cultural competence and cultural humility. This educational intervention consisted of integrating cultural competence within a graduate community mental health nursing course, with the expectation that at the end of the course, students would demonstrate understanding of one's own cultural beliefs as well as the cultural beliefs of others within the mental health care setting. Results from qualitative student interviews indicated that students felt they had become more aware of how cultural factors (their client's as well as their own) may affect the nursing care they provide (Hoke & Robbins, 2011). Killian & Floren (2020), mentioned above, also reported within their study that students who completed the multicultural counseling inventory and reported engaging in self-reflection practices demonstrated statistically significant increases in clinician cultural competence.

Intersectionality and cultural competence

Intersectionality is a theoretical framework for understanding the diversity of experiences of privilege and oppression within groups, with particular attention placed on how marginalized individuals belonging to more than one marginalized group navigate interlocking systems of power and subordination (Bubar et al., 2016; Crenshaw, 2017). Within clinical science and clinician training, this theoretical conceptualization is meaningful because it provides a lens for how educators can address patterns of institutional oppression their clients' experience. Additionally, an intersectional approach to graduate instruction may allow clinicians in training to engage in dialogue and research that highlights the intersection between privileged and oppressed identities marginalized clients hold (Chan et al., 2018). Understanding client experiences of oppression may aid in clinician understanding of factors that differentially lead to negative developmental and mental health outcomes for singularly and multiply marginalized individuals.

The current review yielded scholars who presented various conceptualizations for integrating an intersectional lens within graduate training in the mental health care professions. Rosenthal (2016) noted the common trend for intersectional scholarly work to focus on identities of marginalized individuals rather than on the social inequities they experience, suggesting that intersectional scholarship should attend to the social justice core of intersectionality and work to reduce structural oppression that affects the well-being of marginalized communities. Rosenthal presented several ways that mental health practitioners can make social justice and equity more central agendas in their work, including addressing and critiquing social structures and engaging and collaborating with local communities. Buchannan & Wiklund (2020) provided further critique, positing that while intersectionality is a core competence for mental health practitioners, most graduate programs fail to integrate feminist, critical race, and social justice theories within their curriculum which are pivotal for teaching intersectional competence. Buchanan and Wiklund (2020) presented various methods for incorporating concepts of intersectionality within education and graduate training, including conducting an audit of syllabi to determine if intersectional scholarship is present; establishing a social justice and diversity practicum/specialization; and diversifying the clinical psychology workforce. Grzanka et al. (2021) go further by presenting a conceptual framework for combatting White supremacy through an interdisciplinary intersectional approach, arguing that mental health workers must adopt a commitment to critiquing and transforming oppressive systems. Grzanka et al. provide various recommendations for combatting White supremacy by leveraging current multicultural approaches to clinical work, including utilizing intersectionality to understand clients' relationships to structures of inequality and incorporating interdisciplinary scholarship within intersectional approaches. Gutierrez (2018) furthered the importance of integrating an intersectional lens within the supervisory relationship between clinician in training and instructor. Gutierrez posited that culturally competent supervision involves a responsibility from the supervisor to help students recognize how multiply marginalized clients experience various cultural interactions and systems of power, and aiding supervisees in understanding how their own privilege, biases, and worldviews can influence how they conceptualize and understand multiply marginalized individuals' mental health outcomes (Gutierrez, 2018).

Greene & Flasch (2019) presented a developmental model for integrating intersectionality into clinical supervision in order to foster multicultural competence within clinicians in training. They posited that intersectionality should be integrated into each level of supervisee development to promote a healthy working alliance with clients. This consisted of instructors initiating the conversations about diversity and intersectionality during early stages of student development and transitioning into challenging students to present their own understanding of how client experiences of intersecting systems of power and oppression create unique lived experiences that must be considered for case conceptualization (Green & Flasch, 2019).

Rio (2017) presented considerations for how to integrate an intersectional lens as instructors aim to implement anti-oppressive and liberatory praxis in the classroom. Rio warned about potential issues that may arise when attempting to integrate an anti-oppressive and intersectional approach, including reification of stereotypes by focusing on the "minority client" during discussions of cultural competence. Additionally, Rios spoke to various elements that should be implemented when teaching with an intersectional lens. This included curating a diversity of voices in course syllabi, including content that is pertinent to current events, presenting content aimed at engaging in critical thought and resistance to privilege, and educator preparedness for creating safe spaces for engaging with difficult dialogue that can yield various emotional reactions from students (Rios, 2017). Chan et al. (2018), mentioned in the previous section, also argued that utilizing an intersectional framework may provide a pathway for educators to enhance their students' critical thinking about issues of cultural competence by prompting an analysis of the linkages among various identities.

Our review also yielded various ways in which graduate training can incorporate an intersectional lens into clinician instruction. Bubar et al. (2016) conducted a study in which they tasked graduate students to write a paper on the concepts of structural power and professional power using an intersectional lens in order to assess student understanding of these theoretical concepts. Results indicated that students were unaware of how the provider-client dynamic enacts power and privilege onto their clients. Additionally, student language choice and visual representation of oppression indicated a limited understanding of how oppression is multilayered and intersectional. Bubar et al. (2016) posited that student understanding must therefore be increased by providing them the tools they need to consider paradigms in addition to race and class, employing an intersectional lens that allows them to understand their own place in systems of oppression that permeate our society. Hage et al. (2020) provided a theoretical rationale for implementing social justice practicums (SJP) within doctoral psychology programs, presenting 3 case examples of implementations while discussing considerations and challenges. Hage et al. provided contextual examples using three different academic institutions (Boston College, Springfield College, University of Tennessee) who implemented SJP requisites within their clinical programs; this included courses geared towards preparing students to intervene in the sociocontextual factors that limit community well-being, exploring questions hindering the mental health fields' focus on cultural diversity and social justice, and preparing students to serve as social justice advocates in their varied roles.

Robinson et al. (2016) presented the implementation of an intersectionality assignment designed to create a paradigm shift for students in order for them to incorporate power and privilege into their conceptualization when working with marginalized clients rather than assigning them to siloed identity groups. Students were instructed to read various manuscripts and then tasked to formulate a definition of intersectionality which they would then present to the class. This assignment was paired with an intersectionality exercise in which students received a random identity (with preassigned character descriptions) and were asked to respond to certain statements by moving forward or backward to aid in understanding how power and privilege plays out differently across various co-occurring identities. Reflections from students and instructors showed that the intersectionality assignment and paradigm allowed individuals to view clients as individuals living within an ever-changing environment (Robinson et al., 2016).

Brinkman & Donohue (2020) conducted a study to determine if a course designed to implement an intersectional lens increased student understanding from the beginning to the end of the term. The course was designed to integrate advocacy to promote systems-level changes and engagement and utilized the awareness, knowledge, skills, and action framework proposed in the Multicultural and Social Justice Counseling Competencies (MSJCC) while introducing the students to intersectionality theory. Quantitative results from this pre-/post- assessment study showed that students showed statistically significant increases in cultural competence and their self-efficacy in examining clients' multiple social identities in their conceptualization of them; additionally, qualitative interview results showed students reported initially not understanding intersectionality.

Nagy et al. (2022) implemented a training curriculum designed to introduce the history of multiculturalism and how the intersection of multiple identities leads to unique lived experiences. Trainees were introduced to the cultural formulation interview of the DSM-V and, through a process-oriented model, were taught how to flexibly respond to salient cultural content and view treatment as a dynamic process that can be influenced by cultural context at various stages. While the study sample size prevented analysis of statistically significant changes over time, results showed increases in mean self-rated cultural competence from the beginning to the end of the course. Additionally, qualitative data at the end of the training showed participants reporting a desire for multicultural training to be infused in other parts of their professional education beyond training.

Antiracism, social justice, and cultural competence

Antiracism can be defined as practices that promote racial equity and actively oppose racism through changes in policies, behaviors, and beliefs that perpetuate racist ideas and actions (Toraif et al., 2021). Antiracist teaching aims to create equitable education for students from all cultural groups, fighting for racial justice and reducing societal inequities that societal institutions perpetuate (Lawrence & Tatum, 1997). Within clinical science, antiracism strives to correct the homogenization of curriculum that prioritizes White voices and excludes marginalized communities by meaningfully incorporating teaching that depicts the lived experiences of marginalized groups (Haskins & Singh, 2015). A related concept, social justice has the aim of achieving full and equitable participation of all groups within a society that is structured to meet everyone's respective needs (Bell, 2016). Clinical work with marginalized individuals often involves helping clients navigate a world that was designed to enact harm and perpetuate inequality, and thus should also involve combatting systems of power that create environments that harm clients (Sheely-Moore & Kooyman, 2011).

The current review yielded scholars who presented various ways to conceptualize integration of antiracist and social justice ideologies within clinical training. Galan et al. (2021) argued for consistent efforts to be taken up by the mental health fields towards dismantling policies, practices, and systems that have contributed to racial inequities in clinical science, training, and practice. Galan et al. provided recommendations for structural reform in regard to clinical training and supervision, curriculum and pedagogical approaches, as well as graduate recruitment and retention. This included requiring trainees to complete training in treating racial trauma; establishing and maintaining inclusive clinic procedures; implementing innovative evaluation strategies for measuring cultural competency; and increasing the use of telehealth services to provide accessible options to marginalized clients. Haskins & Singh (2015) presented the integration of critical race theory (CRT) as a framework for promoting antiracist and equitable training within counseling psychology. They posited that educators should strive to investigate how racism and intersectional systems of power may influence their curriculum and their teaching practices. The integration of CRT within curriculum development and implementation may allow students to engage with curriculum that will uplift the voices of marginalized populations, prepare them to identify and address oppressive environments, and better prepare them to work with marginalized individuals.

Several scholars have furthered the idea of structural rather than cultural competence when understanding the lived experiences of marginalized clients (Ali & Sichel, 2014; Hansen et al., 2018; Metzl & Hansen, 2018; Waite & Hassouneh, 2021). Presented within the context of psychiatry, structural competence goes beyond identifying clinician bias and improving communication, pushing clinical providers to 1) understand how social, economic, and political systems produce mental health inequalities in marginalized groups and 2) work to correct these conditions (Metzl & Hansen, 2018). Kirmayer et al. (2018) presented structural competency as pivotal for addressing not only the biological bases of mental illness, but also the social and cultural factors that contribute to psychopathology within marginalized groups. Kirmayer and colleagues argued that clinicians should engage in advocacy outside of the therapy room, understanding the structural determinants that inform client mental health and then supporting coalitions and collective actions aimed at correcting the inequalities that lead to further harm.

Ali & Sichel (2020) spoke to how structural competence can provide a framework for linking psychological practice and social justice. They argued that often, the goals of psychological practice and social justice are at odds with each other (i.e., small group change focus vs. systemic change). Ali and Sichel posited that clinicians can be prepared to integrate structural competence while maintaining their efforts for person-level transformation. This may entail 1) actively understanding the political and social landscape in this country, 2) teaching trainees how a lack of access to resources may lead to negative outcomes in marginalized communities, and (3) engaging in community activism that can lead to greater awareness of the lived experiences of different cultural groups.

Sudak et al. (2020) identified the ways in which psychiatric education institutionally disenfranchises marginalized groups through biases in diagnoses, structural racism in recruitment, and inadequate faculty efforts to prepare students to work with marginalized groups. Sudak and colleagues presented the inclusion of cultural and structural competency trainings as a method to improve the inequity that exists in mental health treatment. Structural competency, in this case, involves a shift from instruction that emphasizes cross-cultural understanding to pedagogy that aims to instruct on the societal and institutional forces that contribute to the mental health of marginalized groups (Sudak et al., 2020). Relatedly, Sheely-Moore & Kooyman (2011) argued for the use of a developmental framework to instill social justice competencies in clinicians in training. They described how clinicians in training should engage with a social action plan, focused on identifying how personal privilege as well as systems of oppression inform how they conceptualize and understand their marginalized clients' experiences of violence and marginalization.

A review of the literature also yielded various ways in which graduate training can incorporate antiracism and social justice into clinician instruction. Brinkman & Hirsch (2019) furthered the idea clinical training should empower students to engage in social justice through the identification of and resistance against oppressive forces. They implemented a study that utilized an advocacy proposal assignment which tasked students with reflecting on a form of oppression and developing a realistic plan to effect change. Results from paired-sample t-tests revealed that the students who completed this assignment had statistically significant increases in their intentions to engage in advocacy as well as their confidence to engage in activism when compared to the control group. Student education that informs them about the institutions that create inequality may serve as a catalyst for counselors to not only focus on individual intervention but on actively changing institutions.

Lenes et al. (2020) worked to promote antiracism within clinical training by presenting a training model that paired multicultural instruction with mindfulness activities. The Color Conscious Multicultural Mindfulness (CCMM) training involved multimodal multicultural content delivery (e.g., videos, art, etc.) which was paired with mindfulness practices to aid participants in reflecting on their emotional reactions to content before they responded. Results indicated that the training model ultimately led to statistically significant increases in multicultural awareness, decreases in color-blind racial attitudes, and increased mindfulness practices. Presseau et al. (2019) investigated student reporting of how integrating multiculturalism and social justice within clinician training environments may inform their social justice attitudes and behaviors. Results from this correlational study indicated that students who underwent social justice and multicultural training were more likely to score higher on a social issues and advocacy scale (comprised of political and social advocacy, political awareness, social issues awareness, and confronting discrimination), suggesting that social justice advocacy may be a natural extension of multicultural competence. By revealing how multicultural competence may extend into anti-oppressive educational and practical opportunities for clinicians in training, Presseau and colleagues furthered the idea that prioritizing multicultural competence training within a social justice environment may allow students to become more comfortable with integrating social justice advocacy within their professional identity.

Rohrbaugh et al. (2019) introduced the Yale Department of Psychiatry Structural Competency Community Initiative (YSCCI), an educational approach designed to introduce psychiatry students to the daily lives of individuals in the community and the structural challenges to mental health they face. This approach is comprised of four components: 1) residents, peers and community leaders form teams and discuss challenging topics, 2) small group leaders lead a tour to experience barriers to safe housing, healthy food, and jobs, 3) students are tasked one week later to present their neighborhood experience with a structure al focus, 4) separate focus groups provide an opportunity for feedback. Results from qualitative interviews indicate students found the experience meaningful; reported deeper appreciation for social determinants of health; and increased engagement with issues of power, hierarchy, and systemic oppression (Rohrbaugh et al., 2019).

Mathis et al. (2019) introduced a drawing exercise for increasing structural competence within psychiatry residents. This consisted of an introductory exercise that tasked psychiatry residents to draw the neighborhoods where they grew up, followed by drawing their patients' neighborhoods, highlighting the factors that impacted their patients' and their own health. Psychiatric residents reported an increased frequency of discussion of social determinants of health after participating in this activity, indicating a more nuanced understanding of the structures that affect their patient's physical and mental health outcomes (Mathis et al., 2019).

Gomez (2022), delineating the lack and need for diversity within psychology, presented a case example of a graduate course within her department which served as a foundation for understanding and dismantling structural inequality by adopting an interdisciplinary perspective. The semester long course tackled topics of structural inequality, critical race theory, queer theory, and highlighted the structural inequities within the mental health field. The ultimate aim of this intervention was a shift to a non-pathologizing conceptualization of marginalized individuals affected by intersectional systems of oppression.

Discussion

This systematic review aimed to take inventory of recent publications addressing the conceptualization of cultural competence (with attention to power, privilege, and oppression) and recent efforts to implement and evaluate cultural competence trainings for graduate trainees in the mental health care fields. Accordingly, this review attended to conceptualizations of cultural competence as well as documentation of implemented training interventions and attempts at evaluation of these interventions. Below we highlight key takeaways from this review, identify existing gaps, and make recommendations for how to move the field forward.

A common theme across the articles reviewed was the importance of adopting a conceptualization of marginalized clients' experiences that accounts for the systems of power and oppression that impact marginalized clients' lived experiences. For example, the structural competence approach pushes for clinicians to understand how social structures influence client access to resources, economic opportunities, and susceptibility to experiencing violence (Metzl & Hansen, 2018). This direction in theoretical considerations for culturally competent care goes beyond acknowledgement of difference by incorporating an interdisciplinary lens that accounts for social structures, cultural systems, and historical contexts that influence the lived experiences of marginalized populations. Multiple social structures may also have a concurrent influence on the lived experiences of marginalized individuals. Scholars have attended to the role of multiple intersecting structures of power, arguing that understanding how each client uniquely experiences oppression through multiple intersecting power structures may improve their understanding of the factors that contribute to different mental health outcomes for multiply marginalized individuals (Chan et al., 2018). The literature highlighted various techniques to increase an intersectional structural lens within graduate curricula, including diversifying the authors included in course syllabi, relating class content to current sociopolitical events, and the preparation of a safe space for engaging with difficult dialogues (Rios, 2017). Additionally, it is possible that these findings may have relevance for clinical training globally as racism and other forms of oppression exists in many forms across the world.

Within the training context, the integration of a structural approach includes having students engage with literature that bolsters their understanding of the sociopolitical landscape in a national and local sense so that they can acknowledge how disparate access to resources can contribute to negative outcomes in marginalized clients. Adopting a structurally competent approach to clinical care may require students to engage in critical self-reflection and humility about their own gaps in knowledge, biases, and cultural identities (Abe, 2020). Pushing students to think critically about their own identities, learning, and experiences can solidify concepts and ideas of oppression and power that influence their clients' lives (Hicks et al., 2019). Furthermore, education is posited as a lifelong commitment, given that marginalized clients exist within an everchanging environment and there may always be gaps in clinicians' knowledge of how other cultural groups navigate systems of power and oppression (Fondocaro & Harder, 2014).

The challenge then exists for training programs instructing students with a structural competence lens to move beyond appreciation of the importance of these issues and understand when and how to address these issues within clinical work. It is much easier to keep discussions of privilege and oppression in the abstract while empathizing with marginalized clients about their different lived experiences. It is much more difficult to explore how these issues of marginalization and hierarchies of power influence the therapeutic relationship and clients' perception of mental health care. Furthermore, clinicians in training should be prepared to acknowledge and discuss how current events related to different systems of power (e.g., race related hate crimes, debates about civil rights legislation) may influence their clients' lives (Cardemill & Battle, 2003). Graduate programs and clinicians in training must understand that failing to discuss and address societal issues in the context of psychotherapy may unintentionally invalidate clients and harm the therapeutic relationship.

Students may also be encouraged to engage in antiracist and social justice action in their communities in attempts to mitigate the structural oppression their clients face (Ali & Sichel, 2020). Clinicians working with marginalized clients may find themselves at a loss as they attempt to help clients cope with a larger social environment that is organized to prevent their access to resources, opportunities, and basic necessities. Clinicians striving to adopt a structurally competent approach to clinical care may benefit from incorporating a social justice orientation into their core mission. This would involve going beyond acknowledgment of systemic injustices and adopting a professional responsibility to affect change at the societal level (Alegria et al., 2022). Scholars have argued for clinicians to engage in advocacy work outside of the therapy room, educating themselves on the structural forces that influence their marginalized clients and then supporting collective action (e.g., community service, campaigning, protesting) aimed at correcting these injustices (Kirmayer et al., 2018, Sheely & Kooyman, 2011). Clinicians in training may benefit from instruction that bridges psychological practice and social justice advocacy, preparing them to aid marginalized individuals while becoming agents of change that can help to ameliorate these oppressive environments.

Recommendations

The current review indicates that there may be growth in the attention to the role of power structures in conceptualizations of cultural competence. However, in reviewing articles for inclusion, many articles were excluded because they relied on definitions that focused primarily on difference without attention to power, privilege, and oppression. A failure to incorporate a structural approach to cultural competence may lead training programs to prepare trainees to utilize therapeutic tools (e.g., cognitive restructuring, exposure techniques, mindfulness) in a way that is unintentionally harmful. For example, instructing students to prepare clients to adapt to harmful social environments (e.g., restructuring experiences of discrimination) without acknowledging the larger social injustices that are in need of change may lead clinicians to be complicit within these systems of oppression.

Unfortunately, there does not appear to be much standardization in how mental health training programs define or implement cultural competence trainings (Benuto et al., 2018). Cultural competence definitions have evolved over time but not necessarily linearly or uniformly. Cultural sensitivity and cultural humility have emerged as part of these evolving definitions with a move to center clinicians' awareness of their positionality and biases as well as growing attention to differences within groups as well as between them (Hook et al., 2017; Kumpfer et al., 2002). Yet our review of the recent literature suggests that mental health programs are still struggling to incorporate ideas of power, privilege, and oppression within their understanding of culturally competent care (Metzl & Hansen, 2014). Variability within conceptualization may lead to inconsistent design and applications of cultural competence trainings within graduate school curricula.

Regarding training interventions, our literature review suggested that many programs continue to rely on a single course to teach students about diversity science and cultural competence (Bubar et al., 2016; Hoke & Robbins, 2011; Jones et al., 2016; Jones & Lee, 2021; Nagy et al., 2022; Robinson et al., 2016; Tormala et al., 2018). Single-course approaches to cultural competence training have been shown to be insufficient and fail to integrate necessary content throughout students' coursework and training and neglect to reinforce critical skills (e.g., initiating difficult conversations with clients, reflecting on one's own biases, etc.) throughout students' academic and clinical trajectory (Anderson et al., 2022). Moreover, single-course approaches may inadvertently promote the idea that cultural competence is a merit to be achieved rather than a lifelong journey that clinicians must constantly reinforce through education, self-introspection, and practice. More contemporary conceptualizations of training advocate for a developmental sequence of multicultural competence which targets student's awareness earlier on and then progresses towards a focus on knowledge acquisition and skill development throughout their academic program (Bardone-Cone et al., 2016).

Additionally, the literature emphasized interventions geared toward instilling cultural competence within graduate trainees such as cross-cultural immersion, exposure to marginalized communities, self-reflection exercises, social justice advocacy assignments, and community service learning (Brinkman & Hirsch, 2019; Jones & Lee, 2021). Many interventions aimed to increase awareness of the role of systems of power and oppression, as well as one's own biases toward marginalized groups. It is worth noting, however, that well-intentioned interventions, such as immersion activities, may have the potential to be exploitative. Depending on the approach implemented, immersion activities may further marginalize marginalized communities by using them as tools for instruction. Moreover, if conceptualized or implemented improperly, these types of activities may reify stereotypes or neglect within-group variability, which would undermine rather than further clinicians' cultural competence capacities.

Furthermore, this review revealed that approaches to measuring and evaluating cultural competence may vary substantially across programs (Bolea, 2012; Brinkman & Donohue, 2020; Dessel & Rodenbord, 2017; Hoke & Robbins, 2011; Jones et al., 2016; Mathis et al., 2019; McDowell et al., 2012). While some measures focused on students' attitudes, self-perceptions, and self-efficacy, other measures assessed specific cultural competence knowledge, awareness, and skills (Sue et al., 1982) that were evaluated using validated and reliable assessment tools (e.g., MAKSS). Assessment tools that measure student cultural competence skills may be of more value for assessing the effectiveness of cultural competence training interventions; however, students' comfort and confidence in their cultural competence abilities seems to be a secondary, but important additional measure. The literature review also indicated that few programs may be using rigorous evaluation methods that include control groups, pre-/post- research designs, and large enough sample sizes to meaningfully and effectively analyze evaluation data (Brinkman & Donohue, 2020; Brinkman & Hirsch, 2019; Jones et al., 2016; Jones & Lee, 2021; Lenes et al., 2020).

To move the field forward, graduate training programs may benefit from a centralized repository which can improve communication across programs and may provide more consistency in how cultural competence is defined, implemented, and evaluated. For example, the APA has disseminated resources for clinicians in training aimed at improving support for culturally competent care, including a multicultural training database and guides for students from marginalized backgrounds (American Psychological Association, 2013). This can be extended further by incentivizing research globally and allowing open communication access between institutions from various nations in order to bolster the availability of culturally and structurally competent training models and evaluation methods. By providing access to resources and the most current research on cultural competence training, best practices can be promoted. This may include providing explicit instruction on how training programs can infuse cultural competence throughout their curriculum to provide many points of intervention (Boroughs et al., 2015). A centralized site could also facilitate the sharing of research methods and data across programs which would allow for analyses to be conducted with data from multiple programs. Lastly, more funding opportunities are needed to better incentivize rigorous research on cultural competence interventions (Galán et al., 2021).

Conclusion

Clients from marginalized backgrounds navigate structures of power that determine their access to resources, economic opportunities, and basic necessities. Clinical care for marginalized individuals may benefit from a structural approach that accounts for oppressive forces within case conceptualization, treatment implementation, and outcome tracking. Graduate programs contend with the question of how to best prepare students to practice culturally competent care with marginalized individuals. Instilling within graduate trainees the importance of lifelong education, social justice advocacy, and critical self-reflection may prepare trainees to provide effective and structurally competent care. The current review suggests that cultural competence conceptualizations, trainings, measures, and evaluation approaches vary across mental health training programs. We suggest that centralized resources and funding opportunities may be necessary to move the field forward. Finally, we believe that the core themes of the results of this review, such as the importance of taking a structurally competent approach to clinical care, can be applied across contexts and may be of benefit to clinicians across the globe.

References

Abe, J. (2020). Beyond cultural competence, toward social transformation: Liberation psychologies and the practice of cultural humility. *Journal of Social Work Education, 56*(4), 696–707. https://doi.org/10.1080/10437797.2019.1661911

Ali, A., & Sichel, C. E. (2014). Structural competency as a framework for training in counseling psychology. *The Counseling Psychologist, 42*(7), 901–918. https://doi.org/10.1177/0011000014550320

Ali, A., & Sichel, C. E. (2020). Radicalizing advocacy in service settings: Using structural competency to address tensions between social action and psychological practice. *Psychological Services, 17*(S1), 22–29. https://doi.org/10.1037/ser0000382

Allen, G. E. (1984). Essay review: The roots of biological determinism. *Journal of the History of Biology, 17*(1), 141-145. https://www.jstor.org/stable/4330882

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

American Psychological Association. (2013, December 12). *APAGS multicultural training database*. https://www.apa.org/apags/governance/subcommittees/cultural-competency

Anderson, R. E., Heard-Garris, N., & DeLapp, R. C. T. (2022). Future directions for vaccinating children against the American endemic: Treating racism as a virus. *Journal of Clinical Child and Adolescent Psychology, 51*(1), 127–142. https://doi.org/10.1080/15374416.2021.1969940

Aymer, S. R. (2016). "I can't breathe": A case study—Helping Black men cope with race-related trauma stemming from police killing and brutality. *Journal of Human Behavior in the Social Environment, 26*(3-4), 367–376. https://doi.org/10.1080/10911359.2015.1132828

Bardone-Cone, A. M., Calhoun, C. D., Fischer, M. S., Gaskin-Wasson, A. L., Jones, S. C. T., Schwartz, S. L., Wise, E. H., & Prinstein, M. J. (2016). Development and implementation of a diversity training sequence in a clinical psychology doctoral program. *The Behavior Therapist, 39*(3), 65–75.

Bell, L. A. (2016). Theoretical foundations for social justice education. In *Teaching for diversity and social justice* (pp. 3-26). Routledge.

Benuto, L. T., Casas, J., & O'Donohue, W. T. (2018). Training culturally competent psychologists: A systematic review of the training outcome literature. *Training and Education in Professional Psychology, 12*(3), 125–134. https://doi.org/10.1037/tep0000190

Betancourt, J. R., Green, A. R., & Carrillo, J. E. (2002). *Cultural competence in health care: Emerging frameworks and practical approaches (Vol. 576)*. Commonwealth Fund, Quality of Care for Underserved Populations.

Bolea, P. S. (2012). Cross-cultural service learning with Native Americans: Pedagogy for building cultural competence. *Journal of Teaching in Social Work, 32*(3), 284-299. https://doi.org/10.1080/08841233.2012.687684

Boroughs, M. S., Bedoya, C. A., O'Cleirigh, C., & Safren, S. A. (2015). Toward defining, measuring, and evaluating LGBT cultural competence for psychologists. *Clinical Psychology: Science and Practice, 22*(2), 151–171. https://doi.org/10.1111/cpsp.12098

Bowleg, L. (2008). When Black + lesbian + woman ≠ Black lesbian woman: The methodological challenges of qualitative and quantitative intersectionality research. *Sex Roles: A Journal of Research, 59*(5-6), 312–325. https://doi.org/10.1007/s11199-008-9400-z

Brinkman, B. G, & Donohue, P. (2020). Doing intersectionality in social justice oriented clinical training. *Training and Education in Professional Psychology, 14*(2), 109–115. https://doi.org/10.1037/tep0000274

Brinkman, B. G., & Hirsch, K. (2019). From proposal to action: Supporting student advocacy during graduate counseling training. *Journal for Social Action in Counseling & Psychology, 11*(1), 51-66.

Bubar, R., Cespedes, K., & Bundy-Fazioli, K. (2016). Intersectionality and social work: Omissions of race, class, and sexuality in graduate school education. *Journal of Social Work Education, 52*(3), 283–296. https://doi.org/10.1080/10437797.2016.1174636

Buchanan, N. T., & Wiklund, L. O. (2020). Why clinical science must change or die: Integrating intersectionality and social justice. *Women & Therapy, 43*(3-4), 309–329. https://doi.org

/10.1080/02703149.2020.1729470

Buchtel, E. E. (2014). Cultural sensitivity or cultural stereotyping? Positive and negative effects of a cultural psychology class. *International Journal of Intercultural Relations, 39*, 40–52. https://doi.org/10.1016/j.ijintrel.2013.09.003

Burns, L., Einaudi, P., & Green, P. (2009). S&E graduate enrollments accelerate in 2007: Enrollments of foreign students reach new high. *InfoBrief. NSF 09-314.* National Science Foundation.

Cardemil, E. V., & Battle, C. L. (2003). Guess who's coming to therapy? Getting comfortable with conversations about race and ethnicity in psychotherapy. *Professional Psychology: Research and Practice, 34*(3), 278–286. https://doi.org/10.1037/0735-7028.34.3.278

Causadias, J. M. (2020). What is culture? Systems of people, places, and practices. *Applied Developmental Science, 24*(4), 310–322. https://doi.org/10.1080/10888691.2020.1789360

Celinska, D., & Swazo, R. (2016). Curriculum designs in counselor education programs: Enhancing counselors-in-training openness to diversity. *Journal of Counselor Preparation & Supervision, 8*(3), 4.

Chan, C. D., Cor, D. N., & Band, M. P. (2018). Privilege and oppression in counselor education: An intersectionality framework. *Journal of Multicultural Counseling and Development, 46*(1), 58–73. https://doi.org/10.1002/jmcd.12092

Chung, R. C.-Y., & Bemak, F. P. (2012). Social justice counseling: The next steps beyond multiculturalism. Sage Publications, Inc.

Constantine, M. G. (2007). Racial microaggressions against African American clients in cross-racial counseling relationships. *Journal of Counseling Psychology, 54*(1), 1–16. https://doi.org/10.1037/0022-0167.54.1.1

Craig, G. (2002). Poverty, social work and social justice. *The British Journal of Social Work,* 669-682. https://www.jstor.org/stable/23716487

Crenshaw, K. W. (2017). *On intersectionality: Essential writings.* The New Press.

Curtis-Boles, H., & Bourg, E. (2010). Experiences of students of color in a graduate-level diversity course. *Training and Education in Professional Psychology, 4*, 204–212. http://dx.doi.org/10.1037/a0017985

Dameron, M. L., Camp, A., Friedmann, B., & Parikh-Foxx, S. (2020). Multicultural education and perceived multicultural competency of school counselors. *Journal of Multicultural Counseling and Development, 48*(3), 176-190. https://doi.org/10.1002/jmcd.12176

Danso, R. (2018). Cultural competence and cultural humility: A critical reflection on key cultural diversity concepts. *Journal of Social Work, 18*(4), 410–430. https://doi.org/10.1177/1468017316654341

Dessel, A., Levy, D. L., Lewis, T. O., McCarty-Caplan, D., Jacobsen, J., & Kaplan, L. (2019). Teaching note—challenges in the classroom on LGBTQ topics and Christianity in social work. *Journal of Social Work Education, 55*(1), 202–210. https://doi.org/10.1080/10437797.2018.1513879

Dessel, A. B., & Rodenborg, N. (2017). An evaluation of intergroup dialogue pedagogy: Addressing segregation and developing cultural competency. *Journal of Social Work Education, 53*(2), 222–239.

Dorn-Medeiros, C. M., Christensen, J. K., Lértora, I. M., & Croffie, A. L. (2020). Relational strategies for teaching multicultural courses in counselor education. *Journal of Multicultural Counseling and Development, 48*(3), 149-160. https://doi.org/10.1002/jmcd.12174

Eagan, M. K., Jr., Hurtado, S., Chang, M. J., Garcia, G. A., Herrera, F. A., & Garibay, J. C. (2013). Making a difference in science education: The impact of undergraduate research programs. *American Educational Research Journal, 50*(4), 683–713. https://doi.org/10.3102/0002831213482038

Estrada, D., Poulsen, S., Cannon, E., & Wiggins, M. (2013). Orienting counseling students toward multiculturalism: Exploring privilege during a new student orientation. *The Journal of Humanistic Counseling, 52*(1), 80-91. https://doi.org/10.1002/j.2161-1939.2013.00034.x

Fisher-Borne, M., Cain, J. M., & Martin, S. L. (2015). From mastery to accountability: Cultural humility as an alternative to cultural competence. *Social Work Education, 34*(2), 165–181. https://doi.org/10.1080/02615479.2014.977244

Fondacaro, K. M., & Harder, V. S. (2014). Connecting cultures: A training model promoting evidence-based psychological services for refugees. *Training and Education in Professional Psychology, 8*(4), 320–327. https://doi.org/10.1037/tep0000071

Galán, C. A., Bekele, B., Boness, C., Bowdring, M., Call, C., Hails, K., ... & Yilmaz, B. (2021). A call to action for an antiracist clinical science. *Journal of Clinical Child & Adolescent Psychology, 50*(1), 12-57. https://doi.org/10.1080/15374416.2020.1860066

Gómez, J. M. (2022). Diversity wanted! Utilizing transdisciplinary scholarship on structural inequality to educate psychology graduate students. *Teaching of Psychology, 0*(0). https://doi.org/10.1177/00986283211061687

Greene, J. H., & Flasch, P. S. (2019). Integrating intersectionality into clinical supervision: A developmental model addressing broader definitions of multicultural competence. *Journal of Counselor Preparation and Supervision, 12*(4), 14. https://repository.wcsu.edu/jcps/vol12/iss4/14

Greenwald, A. G., Dasgupta, N., Dovidio, J. F., Kang, J., Moss-Racusin, C. A., & Teachman, B. A. (2022). Implicit-bias remedies: Treating discriminatory bias as a public-health problem. *Psychological Science in the Public Interest, 23*(1), 7-40. https://doi.org/10.1177/15291006211070781

Grills, C., Cooke, D., Douglas, J., Subica, A., Villanueva, S., & Hudson, B. (2016). Culture, racial socialization, and positive African American youth development. *Journal of Black Psychology, 42*(4), 343–373. https://doi.org/10.1177/0095798415578004

Grzanka, P. R., Gonzalez, K. A., & Spanierman, L. B. (2019). White supremacy and counseling psychology: A critical–conceptual framework. *The Counseling Psychologist, 47*(4), 478–529. https://doi.org/10.1177/0011000019880843

Gutierrez, D. (2018). The role of intersectionality in marriage and family therapy multicultural supervision. *American Journal of Family Therapy, 46*(1), 14–26. https://doi.org/10.1080/01926187.2018.1437573

Hage, S. M., Miles, J. R., Lewis, J. A., Grzanka, P. R., & Goodman, L. A. (2020). The social justice practicum in counseling psychology training. *Training and Education in Professional Psychology, 14*(2), 156–166. https://doi.org/10.1037/tep0000299

Hall, C. C. I. (2014). The evolution of the revolution: The successful establishment of multicultural psychology. In F. T. L. Leong, L. Comas-Díaz, G. C. Nagayama Hall, V. C. McLoyd, & J. E. Trimble (Eds.), *APA handbook of multicultural psychology, Vol. 1: Theory and research* (pp. 3–18). American Psychological Association. https://doi.org/10.1037/14189-001

Hansen, H., Braslow, J., & Rohrbaugh, R. M. (2018). From cultural to structural competency— training psychiatry residents to act on social determinants of health and institutional racism. *JAMA psychiatry, 75*(2), 117-118. https://doi.org/10.1001/jamapsychiatry.2017.3894

Haskins, N. H., & Singh, A. (2015). Critical race theory and counselor education pedagogy: Creating equitable training. *Counselor Education and Supervision, 54*(4), 288–301. https://doi.org/10.1002/ceas.12027

Hicks, Q. S., Hammond, B. M., Winters, R. L., & Boersma, J. M. (2019). Identifying the influence of factors on the quality of critical reflection: Framing, frequency, and feedback. *Journal of Applied Learning and Teaching, 2*(1), 7-15. https://doi.org/10.37074/jalt.2019.2.1.2

Hoke, M. M., & Robbins, L. K. (2011). Continuing the cultural competency journey through exploration of knowledge, attitudes, and skills with advanced practice psychiatric nursing students: An exemplar. *Nursing Clinics, 46*(2), 201-205. https://doi.org/10.1016/j.cnur.2011.02.004

Holyoak, D., Johnson, W., Garcia, M., & Johnson, D. J. (2020). I see you: Addressing diversity in clinical practice through therapist's way-of-being. *Contemporary Family Therapy, 42*(2), 131-139. https://doi.org/10.1007/s10591-019-09519-4

Hook, J. N., Davis, D., Owen, J., & DeBlaere, C. (2017). Cultural humility: Engaging diverse identities in therapy. *American Psychological Association*. https://doi.org/10.1037/0000037-000

Huey, S. J., Jr., Tilley, J. L., Jones, E. O., & Smith, C. A. (2014). The contribution of cultural competence to evidence-based care for ethnically diverse populations. *Annual Review of Clinical Psychology, 10*, 305–338. https://doi.org/10.1146/annurev-clinpsy-032813-153729

Imperato, C., Schneider, B. H., Caricati, L., Amichai-Hamburger, Y., & Mancini, T. (2021). Allport meets internet: A meta-analytical investigation of online intergroup contact and prejudice reduction. *International Journal of Intercultural Relations, 81*, 131-141. https://doi.org/10.1016/j.ijintrel.2021.01.006

Jernigan, V. B. B., Hearod, J. B., Tran, K., Norris, K. C., & Buchwald, D. (2016). An examination of cultural competence training in US medical education guided by the tool for assessing cultural competence training. *Journal of Health Disparities Research and Practice, 9*(3), 150.

Jones, J. M., Begay, K. K., Nakagawa, Y., Cevasco, M., & Sit, J. (2016). Multicultural counseling competence training: Adding value with multicultural consultation. *Journal of Educational & Psychological Consultation, 26*(3), 241–265. https://doi.org/10.1080/10474412.2015.1012671

Jones, J. M., & Lee, L. H. (2021). Multicultural competency building: A multi-year study of trainee self-perceptions of cultural competence. *Contemporary School Psychology, 25*(3), 288-298. https://doi.org/10.1007/s40688-020-00339-0

Killian, T., & Floren, M. (2020). Exploring the relationship between pedagogy and counselor trainees' multicultural and social justice competence. *Journal of Counseling & Development, 98*(3), 295-307. https://doi.org/10.1002/jcad.12324

Kirmayer, L. J., Kronick, R., & Rousseau, C. (2018). Advocacy as key to structural competency in psychiatry. *JAMA psychiatry, 75*(2), 119-120. https://doi.org/10.1001/jamapsychiatry.2017.3897

Koç, V., & Kafa, G. (2019). Cross-Cultural Research on Psychotherapy: The Need for a Change. *Journal of Cross-Cultural Psychology, 50*(1), 100–115. https://doi.org/10.1177/0022022118806577

Kormos, J., Csizér, K., & Iwaniec, J. (2014). A mixed-method study of language-learning motivation and intercultural contact of international students. *Journal of Multilingual and Multicultural Development, 35*(2), 151-166. https://doi.org/10.1080/01434632.2013.847940

Kumpfer, K. L., Alvarado, R., Smith, P., & Bellamy, N. (2002). Cultural sensitivity and adaptation in family-based prevention interventions. *Prevention Science, 3*(3), 241–246. https://doi.org/10.1023/A:1019902902119

Lakes, K., López, S. R., & Garro, L. C. (2006). Cultural competence and psychotherapy: Applying anthropologically informed conceptions of culture. *Psychotherapy: Theory, Research, Practice, Training, 43*(4), 380–396. https://doi.org/10.1037/0033-3204.43.4.380

Lawrence, S. M., & Tatum, B. D. (1997). Teachers in transition: The impact of antiracist professional development on classroom practice. *Teachers College Record, 99*(1), 162-178. https://doi.org/10.1177/016146819709900110

Lee, E. K. O. (2014). Use of avatars and a virtual community to increase cultural competence. *Journal of Technology in Human Services, 32*(1-2), 93-107. https://doi.org/10.1080/15228835.2013.860364

Lehman, D. R., Chiu, C.-y., & Schaller, M. (2004). Psychology and culture. *Annual Review of Psychology, 55*, 689–714. https://doi.org/10.1146/annurev.psych.55.090902.141927

Lenes, E., Swank, J. M., Hart, K. A., Machado, M. M., Darilus, S., Ardelt, M., Smith-Adcock, S., Rockwood Lane, M., & Puig, A. (2020). "Color-conscious multicultural mindfulness training in the counseling field": Erratum. *Journal of Counseling & Development, 98*(3), 345. https://doi.org/10.1002/jcad.12328

Leong, F. T. L., Pickren, W. E., & Vasquez, M. J. T. (2017). APA efforts in promoting human rights and social justice. *American Psychologist, 72*(8), 778–790. https://doi.org/10.1037/amp0000220

Lin, L., Stamm, K., & Christidis, P. (2018, February 1). How diverse is the psychology workforce? *Monitor on Psychology, 49*(2). https://www.apa.org/monitor/2018/02/datapoint

Lombardo, P. A. (1985). Three generations, no imbeciles: New light on Buck v. Bell. *NYUL Rev., 60,* 30.

Lorelle, S., Atkins, K., & Michel, R. (2021). Enhancing social justice and multicultural counseling competence through cultural immersion: A guide for faculty. *Journal of Counselor Preparation & Supervision, 14*(1), 4.

Lusk, M., Terrazas, S., & Salcido, R. (2017). Critical cultural competence in social work supervision. *Human Service Organizations: Management, Leadership & Governance, 41*(5), 464-476. https://doi.org/10.1080/23303131.2017.1313801

Mackelprang, R. W., & Salsgiver, R. O. (1996). People with disabilities and social work: historical and contemporary issues. *Social Work,* 7-14. https://www.jstor.org/stable/23718240

Mallett, R., Hagen-Zanker, J., Slater, R., & Duvendack, M. (2012). The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness, 4*(3), 445-455. https://doi.org/10.1080/19439342.2012.711342

Mathis, W., Cyrus, K., Jordan, A., & Rohrbaugh, R. (2019). Introducing a structural competency framework for psychiatry residents: drawing your neighborhood. *Academic Psychiatry, 43*(6), 635-638. https://doi.org/10.1007/s40596-019-01077-z

McDowell, T., Goessling, K., & Melendez, T. (2012). Transformative learning through international immersion: Building multicultural competence in family therapy and counseling. *Journal of Marital and Family Therapy, 38*(2), 365-379. https://doi.org/10.1111/j.1752-0606.2010.00209.x

McLeod, K., Moore, R., Robinson, D., Ozkul, D., Ciftci, S., & Vincent, K. (2020). Using the pluriverse concept to critique Eurocentrism in education. *Journal of Applied Learning & Teaching, 3*(SI1), 30-39. https://doi.org/10.37074/jalt.2020.3.s1.4

Merta, R. J., Stringham, E. M., & Ponterotto, J. G. (1988). Simulating culture shock in counselor trainees: An experiential exercise for cross-cultural training. *Journal of Counseling & Development, 66,* 242–245. http://dx.doi.org/10.1002/j.1556-6676.1988.tb00858.x

Metzl, J. M., & Hansen, H. (2018). Structural competency and psychiatry. *JAMA Psychiatry, 75*(2), 115-116. https://doi.org/10.1001/jamapsychiatry.2017.3891

Misra, S., & McMahon, G. (2006). Diversity in higher education: The three Rs. *Journal of Education for Business, 82*(1), 40–43. https://doi.org/10.3200/JOEB.82.1.40-43

Moradi, B., & Grzanka, P. R. (2017). Using intersectionality responsibly: Toward critical epistemology, structural analysis, and social justice activism. *Journal of Counseling Psychology, 64*(5), 500–513. https://doi.org/10.1037/cou0000203

Mosher, D. K., Hook, J. N., Captari, L. E., Davis, D. E., DeBlaere, C., & Owen, J. (2017). Cultural humility: A therapeutic framework for engaging diverse clients. *Practice Innovations, 2*(4), 221–233. https://doi.org/10.1037/pri0000055

Nagy, G. A., LeMaire, K. L., Miller, M. L., Bhatt-Mackin, S., & Railey, K. (2022). Training and education to advance multicultural mental health-care delivery (the "TEAM mental health-care delivery model"): A pilot evaluation of outcomes, acceptability, and feasibility. *Training and Education in Professional Psychology, 16*(1), 67–77. https://doi.org/10.1037/tep0000347

North, C. E. (2006). More than words? Delving into the substantive meaning(s) of "social justice" in education. *Review of Educational Research, 76*(4), 507–535. https://doi.org/10.3102/00346543076004507

Parikh-Foxx, S., Grimmet, M., & Dameron, M. L. (2020). Examining the use of digital storytelling and immersion in a multicultural counseling course. *Journal of Multicultural Counseling and Development, 48*(3), 137-148. https://doi.org/10.1002/jmcd.12173

Platt, J. J. (2012). A Mexico City–based immersion education program: Training mental health clinicians for practice with Latino communities. *Journal of Marital and Family Therapy, 38*(2), 352–364. https://doi.org/10.1111/j.1752-0606.2010.00208.x

Presseau, C., Luu, L. P., Inman, A. G., & DeBlaere, C. (2019). Trainee social justice advocacy: Investigating the roles of training factors and multicultural competence. *Counselling Psychology Quarterly, 32*(2), 260–274. https://doi.org/10.1080/09515070.2018.1476837

Radford, J. P. (1991). Sterilization versus segregation: Control of the 'feebleminded', 1900–1938. *Social Science & Medicine, 33*(4), 449-458. https://doi.org/10.1016/0277-9536(91)90327-9

Ratts, M. J., Singh, A. A., Nassar-McMillan, S., Butler, S. K., & McCullough, J. R. (2016). Multicultural and Social justice counseling competencies: Guidelines for the counseling profession. *Journal of Multicultural Counseling and Development, 44*(1), 28–48. https://doi.org/10.1002/jmcd.12035

Remedios, J. D., & Snyder, S. H. (2018). Intersectional oppression: Multiple stigmatized identities and perceptions of invisibility, discrimination, and stereotyping. *Journal of Social Issues, 74*(2), 265–281. https://doi.org/10.1111/josi.12268

Ridley, C. R., Sahu, A., Console, K., Surya, S., Tran, V., Xie, S., & Yin, C. (2021). The process model of multicultural counseling competence. *The Counseling Psychologist, 49*(4), 534–567. https://doi.org/10.1177/0011000021992339

Río, M. R. S. D. (2017). Teaching at the intersections: Liberatory and anti-oppressive pedagogical praxis in the multicultural counseling classroom as a Queer Puerto Rican educator. *Feminist Teacher, 27*(2-3), 90-105. https://doi.org/10.5406/femteacher.27.2-3.0090

Robinson, M. A., Cross-Denny, B., Lee, K. K., Werkmeister Rozas, L. M., & Yamada, A. M. (2016). Teaching note—Teaching intersectionality: Transforming cultural competence content in social work education. *Journal of Social Work Education, 52*(4), 509-517. https://doi.org/10.1080/10437797.2016.1198297

Rogers, M. R., & O'Bryon, E. C. (2014). Multicultural training models and curriculum. In F. T. L. Leong, L. Comas-Díaz, G. C. Nagayama Hall, V. C. McLoyd, & J. E. Trimble (Eds.), *APA handbook of multicultural psychology, Vol. 2. Applications and training* (pp. 659–679). American Psychological Association. https://doi.org/10.1037/14187-037

Rohrbaugh, R. M., Bromage, B., Encandela, J. A., Spell, V. T., Williamson, B., & Mathis, W. S. (2019). Allying with our neighbors to teach structural competence: The Yale Department of Psychiatry Structural Competency Community Initiative (YSCCI). In H. Hansen, H. & J. Metzl (Eds.), *Structural competency in mental health and medicine.* Springer, Cham. https://doi.org/10.1007/978-3-030-10525-9_13

Rosenthal, L. (2016). Incorporating intersectionality into psychology: An opportunity to promote social justice and equity. *American Psychologist, 71*(6), 474–485. https://doi.org/10.1037/a0040323

Roysircar, G., Gard, G., Hubbell, R., & Ortega, M. (2005). Development of counseling trainees' multicultural awareness through mentoring English as a Second Language students. *Journal of Multicultural Counseling and Development, 33*,17–36. http://dx.doi.org/10.1002/j.2161-1912.2005.tb00002.x

Seaton, E. K., Gee, G. C., Neblett, E., & Spanierman, L. (2018). New directions for racial discrimination research as inspired by the integrative model. *American Psychologist, 73*(6), 768–780. https://doi.org/10.1037/amp0000315

Shannonhouse, L. R., Myers, J. E., & Barrio Minton, C. A. (2018). Cultural immersion in counselor education: Trends, prevalence, and common components. *Journal of Multicultural Counseling and Development, 46*(4), 283-296. https://doi.org/10.1002/jmcd.12115

Sheely-Moore, A. I., & Kooyman, L. (2011). Infusing multicultural and social justice competencies within counseling practice: A guide for trainers. *Adultspan Journal, 10*(2), 102-109. https://doi.org/10.1002/j.2161-0029.2011.tb00129.x

Stabb, S. D., & Reimers, F. A. (2013). Competent poverty training. *Journal of Clinical Psychology, 69*(2), 172–181. https://doi.org/10.1002/jclp.21956

Stanhope, V., Tuchman, E., & Sinclair, W. (2011). The implementation of mental health evidence-based practices from the educator, clinician and researcher perspective. *Clinical Social Work Journal, 39*(4), 369-378. https://doi.org/10.1007/s10615-010-0309-y

Stoecker, R., & Tryon, E. A. (Eds.). (2009). *The unheard voices: Community organizations and service learning*. Temple University Press.

Sudak, D. M., DeJong, S. M., Bailey, B., & Rohrbaugh, R. M. (2020). Training psychiatrists to achieve mental health equity. *Psychiatric Clinics of North America, 43*(3), 555–568. https://doi.org/10.1016/j.psc.2020.05.003

Sue, D.W., Bernier, J., Durran, M., Feinberg, L., Pedersen, P., Smith, E., & Vasquez-Nuttall, E. (1982). Position paper: Multicultural counseling competencies. *Counseling Psychologist,10,* 45–52. https://doi.org/10.1177/0011000082102008

Sue, S., Zane, N., Hall, G. C. N., & Berger, L. K. (2009). The case for cultural competency in psychotherapeutic interventions. *Annual Review of Psychology, 60,* 525–548. https://doi.org/10.1146/annurev.psych.60.110707.163651

Tasca, C., Rapetti, M., Carta, M. G., & Fadda, B. (2012). Women and hysteria in the history of mental health. *Clinical practice and epidemiology in mental health: CP & EMH, 8,* 110. https://doi.org/10.2174/1745017901208010110

Thibeault, D. (2019). Understanding Indigenous culture through service learning and cultural immersion. *Journal of Social Work Education, 55*(3), 476-488. https://doi.org/10.1080/10437797.2019.1593901

Tievsky, D. L. (1988). Homosexual clients and homophobic social workers. *Journal of Independent Social Work, 2*(3), 51–

62. https://doi.org/10.1300/J283v02n03_06

Toporek, R. L., & Worthington, R. L. (2014). Integrating service learning and difficult dialogues pedagogy to advance social justice training. *The Counseling Psychologist, 42*(7), 919–945. https://doi.org/10.1177/0011000014545090

Toraif, N., Augsberger, A., Young, A., Murillo, H., Bautista, R., Garcia, S., ... & Gergen Barnett, K. (2021). How to be an antiracist: Youth of color's critical perspectives on antiracism in a youth participatory action research context. *Journal of Adolescent Research, 36*(5), 467-500. https://doi.org/10.1177/07435584211028

Tormala, T. T., Patel, S. G., Soukup, E. E., & Clarke, A. V. (2018). Developing measurable cultural competence and cultural humility: An application of the cultural formulation. *Training and Education in Professional Psychology, 12*(1), 54–61. https://doi.org/10.1037/tep0000183

Trinh, N. H., Jahan, A. B., & Chen, J. A. (2021). Moving from cultural competence to cultural humility in psychiatric education. *Psychiatric Clinics, 44*(2), 149-157. https://doi.org/10.1016/j.psc.2020.12.002

Truong, M., Paradies, Y., & Priest, N. (2014). Interventions to improve cultural competency in healthcare: A systematic review of reviews. *BMC Health Services Research, 14*(1), 1-17. https://doi.org/10.1186/1472-6963-14-99

Venner, K. L., & Verney, S. P. (2015). Motivational interviewing: Reduce student reluctance and increase engagement in learning multicultural concepts. *Professional Psychology: Research and Practice, 46*(2), 116–123. https://doi.org/10.1037/a0038856

Waite, R., & Hassouneh, D. (2021). Structural competency in mental health nursing: Understanding and applying key concepts. *Archives of Psychiatric Nursing, 35*(1), 73-79. https://doi.org/10.1016/j.apnu.2020.09.013

Whaley, A. L., & Davis, K. E. (2007). Cultural competence and evidence-based practice in mental health services: A complementary perspective. *American Psychologist, 62*(6), 563–574. https://doi.org/10.1037/0003-066X.62.6.563

Williams, D. R., & Etkins, O. S. (2021). Racism and mental health. *World Psychiatry, 20*(2), 194. https://doi.org/10.1002/wps.20845

Yamada, A.-M., & Brekke, J. S. (2008). Addressing mental health disparities through clinical competence not just cultural competence: The need for assessment of sociocultural issues in the delivery of evidence-based psychosocial rehabilitation services. *Clinical Psychology Review, 28*(8), 1386–1399. https://doi.org/10.1016/j.cpr.2008.07.006

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Sociodemographic factors and teaching method preferences among university academics: Implications for effective curriculum implementation

Benjamin Ayua Ambe[A] — A Department of Environmental Education, University of Calabar, Nigeria

Cassidy Etta Agbor[B] — B Department of Environmental Education, University of Calabar, Nigeria

Idongesit Emmanuel Ephraim[C] — C Department of Environmental Education, University of Calabar, Nigeria

Okon Eyo Eneyo[D] — D Department of Environmental Education, University of Calabar, Nigeria

Edung Etim Asuquo[E] — E Department of Environmental Education, University of Calabar, Nigeria

Ime Edet Sam[F] — F Department of Environmental Education, University of Calabar, Nigeria

Michael Obun Etan[G] — G Department of Environmental Education, University of Calabar, Nigeria

Anthony Ntol Ngban[H] — H Department of Educational Foundations, University of Calabar, Nigeria

Achi Ndifon Bekomson[I] — I Department of Environmental Education, University of Calabar, Nigeria

Melvina Nkemdilim Amalu[J] — J Department of Environmental Education, University of Calabar, Nigeria

Keywords

Academic faculty;
curriculum implementation;
gender;
lesson presentations;
teaching experience;
teaching methods.

Correspondence

ayuambe3@gmail.com [A]

Article Info

Received 25 June 2023
Received in revised form 28 August 2023
Accepted 30 August 2023
Available online 4 September 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.25

Abstract

This research examined sociodemographic factors and teaching methods preferences among university academics: implications for effective curriculum implementation. This study employed a quantitative survey design; 400 academics were sampled. A questionnaire was used to obtain data; descriptive statistics and chi square analysis were used to test research hypotheses. The percentage of academics who prefer various teaching techniques during lessons for efficient curriculum implementation differs significantly; there is no meaningful connection between gender, academic faculty, years of classroom instruction, and their preference for teaching methods. The project method, followed by experimentation and demonstration methods, which are more constructivist and allow students to participate in their classes actively, were recommended as tools for academics to use more frequently. Despite these outcomes, individual differences must be respected. Regardless of gender, it is recommended that institutions regularly hold professional development seminars and training sessions, encourage multidisciplinary collaboration among educators, and enhance mentoring programmes and platforms for less experienced educators. The intersection of variables, including gender, faculty type, and teaching experience, should be taken into account in a comprehensive approach to pedagogical enhancement. Institutions can be aware of the changing requirements and preferences of educators by establishing channels for academic feedback on teaching techniques and preferences.

Introduction

Teaching is an art that transcends the passing of knowledge from one person to another. It embodies the moulding of the receivers' character, attitudes, knowledge, belief systems, and personalities. Academics teaching in higher institutions are called lecturers or professors. Amadioha (2017), Ambe and Agbor (2014) observed that teaching is a significant aspect of curriculum implementation. It is not done haphazardly but follows laid-down strategies and methods. This study sought to investigate sociodemographic considerations of university academics and teaching method preferences during lesson presentations for effective curriculum implementation.

This research draws from the *constructivist learning theory*, an educational philosophy emphasizing active and experiential learning, where learners construct their knowledge and understanding through meaningful interactions with their environment (Bada & Olusegun, 2015). Rooted in the works of Jean Piaget and Lev Vygotsky, constructivism suggests that learners actively participate in the learning process. In line with this theory, teaching methods aim to create engaging and collaborative learning experiences, such as problem-based, project-based, and inquiry-based instruction. These methods enable learners to investigate, question and gain understanding, fostering critical thinking, creativity, and the application of knowledge in real-world contexts. By promoting student agency, interaction, and reflection, constructivist teaching methods provide a learner-centred approach that empowers students to construct their understanding, leading to deeper and more meaningful learning outcomes (O'Neill, 2014).

In contemporary higher education, effectively implementing curricula ensures quality learning outcomes. As universities strive to provide comprehensive education, it is essential to consider the sociodemographic characteristics of university academics and their preferences for teaching methods. Understanding the influence of sociodemographic factors on teaching method preference is critical for designing inclusive and effective curricula that cater to the diverse needs of faculty members and students.

Currently, limited research examines the relationship between sociodemographic considerations of university academics and their teaching method preference for effective curriculum implementation. While there are numerous ways to teach, including lectures and conversations, group work, and technology-enhanced learning, it remains unclear how academics' sociodemographic factors, including age, gender, educational background, years of teaching experience, and cultural background, influence their preference for specific teaching approaches.

The lack of comprehensive investigation into sociodemographic considerations and teaching method preference presents a significant gap in our understanding of curriculum implementation in higher education. Moreover, the current literature must include comprehensive research on the relationship between sociodemographic considerations of university academics and their teaching methods' preference for effective curriculum implementation. With this knowledge, curriculum developers and educators may be able to align instructional strategies with faculty members' characteristics and preferences, hindering the creation of an engaging learning environment. Rigorous empirical research is needed to address this gap and inform evidence-based decision-making processes in curriculum design and faculty development initiatives.

Additionally, this research can contribute to uncovering the underlying patterns and dynamics; such research can inform evidence-based decision-making processes in curriculum design and faculty development initiatives by enhancing the overall quality of higher education and promoting a deeper comprehension of the variables influencing effective curriculum implementation in diverse academic contexts. By conducting in-depth investigations in this area, educational stakeholders can foster an inclusive and learner-centred environment that maximizes the potential of both faculty members and students.

This study investigates sociodemographic factors and teaching method preferences among university academics: Implications for effective curriculum implementation. Specifically, the researchers seek to find out whether:

1. There is a difference in the proportion of academics preferring different teaching methods during lessons for effective curriculum implementation.

2. There is an association between gender and teaching methods' preference among academics during lessons for effective curriculum implementation.

3. There is an association between the academics' faculty and teaching method preference during lessons for effective curriculum implementation.

4. Any Association exists between the academics' years of teaching experience and teaching method preference during lessons for effective curriculum implementation.

Research hypotheses

1. The percentage of academics who prefer various teaching techniques during lessons for efficient curriculum implementation does not differ significantly.

2. No association exists between gender and teaching methods preference among academics during lessons for effective curriculum implementation.

3. There is no association between the academics' faculty and teaching method preference during lessons for effective curriculum implementation.

4. There is no association between the years of academic teaching experience and teaching method preference for effective curriculum implementation.

Literature review

Amadioha (2017) defined teaching as the impartation of unknown knowledge to the learner; it is a process of getting learners educated. Skills are not contagious; they are methodically transmitted from person to person in an organized setting, either formally or informally. Ambe and Agbor (2014) noted that a professionally competent teacher must be able to prepare for his lesson, plan the lesson and present the lesson systematically to achieve the stated objectives. Ephraim et al. (2022) noted that to attain goals in their respective disciplines for successful learning, teachers must create lesson plans and instructional materials and use the right teaching techniques.

As Ambe and Agbor (2014) noted, seasoned educators draw from a broader and more complex body of information than upcoming ones. Clotfelter et al. (2010) show that over 20 years of experience is more effective than no experience; nonetheless, they contrast in efficacy with teachers who have five years of experience. Earlier studies by Kim and Seo (2018) and Kim et al. (2019) discovered a strong correlation between teachers' professional effectiveness and their expertise. Moreover, the likelihood of an educator being productive increases with the length of service in a college. The Illinois State Board of Education (ISBE, 2002) and Emmanue and Ambe (2014) argued that a trained educator is skilled in all academic disciplines and exhibits subject-matter expertise.

An instructor's ability to effectively convey the substance of the course relies on their comprehension of that subject's structure (Ambe & Agbor, 2014). Also, an instructor's perspective of proper instruction follows from their ability to do so. Ambe and Agbor (2014) argued that a lecturer's ability to effectively convey the substance of a field rests on their comprehension of the dynamics of that field. This, in turn, leads to the impression of an instructor as having appropriate teaching abilities.

There are several categories of knowledge that seasoned teachers pick up, including knowledge of the subject's fundamental concepts, often known as subject area knowledge (Niemelä & Tirri, 2018; Mupa & Chinooneka, 2015). Knowing how to make a subject interesting and understandable is known as pedagogical content knowledge, among other things.

Nwogu and Esobhawan (2014) observed that teaching involves practical communication skills, abilities for handling classroom operations and effective instructional approaches. In a classroom situation, a teacher, a lecturer, or an academic utilizes various teaching methods while executing any lesson. Teaching methods refer to the variety of styles, techniques, and ways the teacher uses to expound a lesson to students. The University of Buffalo (2022) clarified that the more general approaches to assisting students in meeting their learning objectives are known as teaching strategies. These strategies could be student-centred, teacher-centred, or technology-centred.

Shah and Udgaonkar (2018), Ambe and Onnoghen (2018), and Bhat (2017) found that teachers' gender does not significantly influence teaching effectiveness. This means, therefore, that irrespective of the gender of the teacher, as long as they have the requisite professional competence, they are prepared to use appropriate teaching methods and adequately deliver the lesson, effective learning will take place. Amadioha (2017) highlights the importance of a recurrent teaching method, focusing on a major activity relevant to all disciplines in any teaching-learning setting. According to Al-Rawi (2013), a teacher uses a teaching style as a framework to arrange and carry out instructional strategies and tasks for the achievement of school or educational goals. Sikaleya (2022) observed that there are over fifty educational methodologies in practice, and the educator must use effective instructional strategies in the classroom to accomplish the teaching objective. Among the myriad of teaching methods available to any academic are the following, which these researchers purposely selected for this study.

Lecture method of teaching

Sikaleya (2022) and Al-Rawi (2013) see the lecture as an oral instructional technique. According to the authors, its advantage is that it gives the teacher total control of the lesson and makes them active participants, while the students are primarily passive; it saves time. Amadioha (2017) sees the lecture as a presentation method that involves chalk and talk, as the teacher is the primary participant; they explain points, express opinions, give students new ideas, and occasionally write on the board. According to the authors, its advantage is that a large class is taught relatively quickly.

Kapur (2020) sees the lecture as the most comprehensively used pedagogical method; it is the oldest teaching method, and academics use it extensively. Alaagib et al. (2019) observed that one of the most popular forms of instruction in medical education is the lecture. In an article, the researchers taught students using the problem-based method and the traditional lecture. In the end, the learners were distributed a test and a questionnaire. The students' attention ($P = 0.002$) and participation ($P = 0.003$) were higher in the problem-based lecture technique than in the traditional lecture (Alaagib et al., 2019). The awareness of the learning objectives did not significantly change between the problem-based lecture and conventional lectures.

In schools, the lecture is a common teaching strategy and a primary method (Bala et al., 2017; Noel et al., 2015). Abdulbaki et al. (2018a), while stressing the demerits of the lecture, noted that in any discipline of education, including nursing education, lecturing puts students in a receptive role rather than as active participants, thus hampering learning. In schools, the lecture is a common teaching strategy. The authors noted further that lectures are not successful at altering attitudes or values, teaching manual dexterity, or teaching higher-order cognitive abilities like application, analysis, synthesis, or evaluation (Abdulbaki et al., 2018a).

The present mode of instruction in medical schools is the lecture, yet lectures by themselves are inadequate for fostering cooperative learning and skill development (Dharmambal & Anavarathan, 2021). Viswanathan and Viswanathan (2017) noted that lectures are the instructional technology in a teacher's repertoire. According to Sadeghi et al. (2014), lectures are a quick, easy, and affordable way to introduce large topics to numerous populations of students. The authors researched two teaching methods and concluded that students preferred the mixed-learning approach over lectures. As a result, it is asserted that the lecture is teacher-centred, with the lecturer spending most of the lecture interacting with learners who may be listening passively.

Demonstration method

Demonstration as a teaching method requires the lecturer to practicalize whatever they are teaching the students. By using body language, gestures, postures, and facial expressions to illustrate a point during a lecture, this technique is known as a demonstration (Hussain, 2020). Mohammed et al. (2016) opined that demonstration generates interest, presents ideas and concepts more clearly, offers direct experiences, and reinforces learning. Learners can see, hear, and perhaps even experience an actual incident. Hajar et al. (2021) are of the view that the demonstration method is an approach for delivering learning information that involves showing pupils a particular procedure, circumstance, or object that is being examined, whether it be actual or made-up. This method is frequently combined with vocal comments.

Al-Rawi (2013) claims that the demonstrative teaching approach is successful at imparting scientific laboratory experimentation and tool use abilities. Omotayo and Adedeji (2020) posited that the demonstration teaching style entails demonstrating a special procedure or talent to the participants. While stressing the advantages of the demonstration method, Eze and Nwaukwa (2019) observed that it helps make links between facts and how they apply in real life, it may increase student attention and help them remember information better. A significant disadvantage of this method is that it is not child-centred; it may cause the slow students to be dragged at the speed of the fast learners. There is a limited activity for students; they merely observe the demonstrator (teacher) with little active participation. Time is usually challenging for the demonstration method (Hussain, 2020; Eze & Nwaukwa, 2019; Mohammed et al., 2016).

Discussion method

The discussion method is learner-centred, where the students are active discussion group members. The discussion method is a two-way communication between participants where ideas are shared between students with the moderation of the teacher or one of the students knowledgeable in the subject of discussion. Abdulbaki et al. (2018b) noted that the discussion process is not merely controlled by one individual presentation, as in the lecture. Ying (2020) observed that important learning outcomes for students are produced through discussion methods. while the discussion method is important, it focuses more on student engagement and learning than teaching and improves self-confidence and eloquence among the learners (Ying, 2020). This method cannot be used for all topics. Extroverted students may take over the discussion at the expense of introverted ones.

Experimentation method

In an experimental teaching method, investigations are involved in which hypotheses are scientifically tested. A straightforward and entertaining framework for introducing students to quantitative social research is provided via experiments (Soares et al., 2016). According to Soares et al. (2016), this lesson plan could be used as a guide to teach students how to conduct more difficult research. The method is best used with advanced learners, of which higher institution students are a part. When the experimental teaching technique (ETM) and the teacher-centred traditional teaching method were compared for knowledge and understanding levels, it was found that the experimental teaching approach performed better (Chingala, 2020).

The website holah.karoo.net (n.d) records that the experimental method is the preferred mode of instruction. Moreover, the experiment is a form of causal analysis often performed in the laboratory. It allows precise control of variables, can be replicated, and yields quantitative data. Its disadvantage lies in the fact that behaviour in the laboratory is narrow and artificial. We may have field experiments or natural experiments.

Anderson and McLean (2018) noted that teaching experimentation is a series of lessons where researchers test their hypotheses in steps, and students learn and reason. Soares et al. (2016) lamented that regarding experimentation, the teaching of science in schools needed to be applied appropriately by teachers to carry the learners along. Although experimentation arouses students' interest in learning, experimentation is time-consuming. With many students in higher institutions, completing course outlines within semesters might be impossible if other teaching methods are not involved.

Project method

According to Knoll (2014), one of the common teaching strategies is the project method, which is frequently explored under the titles of project work, project strategy, and task-based learning. The project method is widely used in various educational fields, but regrettably, it still needs greater importance in regular education. In the words of Kolodziejski and Przybysz-Zaremba, (2017), the project technique is frequently employed when instructing college students. It helps students master their intellect, abilities, moral habits, and experimental abilities. Every level of development in education uses the project method extensively. Yet, for it to be applied, the instructor in charge of overseeing its execution must have the necessary skills and knowledge.

The project method enables new approaches because learners can ask questions that awaken their curiosity. The constructivist educational paradigm, which has the creative activity of the person or group as its cornerstone, is the bedrock upon which the project method is based. Prtljaga and Veselinov (2017) argued that the project method should be used in classrooms because it helps to improve student participation and reinforces understanding and mental activity in the classroom.

Methodology

In order to collect quantitative information on the sociodemographic features of the university academics (such as gender, faculty, years of teaching experience), as well as their preferences for different teaching styles, the researchers employed a quantitative survey research technique. This is in line with Loeb et al.'s (2017) argument that researchers use analytical and data visualization methods to transform raw data into useful findings for intended audiences.

The population of the study draws from the 2,867 academic staff of the University of Calabar, Calabar-Nigeria and the University of Cross River State (UNICROSS), all in South-South, Nigeria. The University of Calabar, from records of the Human Resources Directorate, has 2,410 academic staff, while UNICROSS, from records of the academic planning unit, has 457 academic staff. A multistage sampling procedure was adopted to obtain a sample of the study. In the first stage, the researchers sampled five faculties using the hat-and-draw method, three faculties (Education, Science, and medical sciences) were sampled from UNICAL out of 16, and two from UNICROSS (Social Science and Arts) out of eight faculties.

In the second stage, the researchers chose nine departments using a basic random selection technique (30% of 30) as sample departments for the study. Thirdly, the complete faculty in each sampled department was chosen using a purposive sampling technique. Questionnaires were administered to staff in their offices who agreed to participate in the study. 400 academics were sampled, making it 13.95% of the population.

Instrumentation

A four-item structured questionnaire that elicited information on gender, years of teaching experience, faculty of the respondents, and the lecturer's preference for six teaching methods was used to obtain data. Gender was categorized into male(1) and female (2). Years of teaching experience were categorized into three possible answers: 1-10 years (young academics) was scored one point, 11-20 years (intermediate academics) was scored two points, and 21 years and above (mature academics) was scored three points. The lecture approach received a score of one, demonstration two points, discussion three points, experimentation four points, project method five points, and a combination of at least two methods was scored six points. For the faculty of the respondents, education was scored one point, social science two points, arts three points, science four points, and medical sciences five points. The data analysis used descriptive statistics like frequency counts, percentages, and the Chi-square technique of data analysis. Microsoft Excel was used. The ethics committee of the University of Calabar and the University of Cross River State gave written approval for this research study to be conducted. Findings were expressed in tables and bar charts.

Results

The results of the analyses are presented hypothesis by hypothesis. Table 1 shows the demographic characteristics of the participants.

Hypothesis 1: The percentage of academics who prefer various teaching techniques during lessons for efficient curriculum implementation does not differ significantly. Data from item 4 of the instrument was computed and subjected to descriptive analysis of simple percentages and the Chi-Square analysis technique to test hypothesis one, as shown in Figure 1.

In Figure 1, the observed preference count shows that 103 (25.8%) academics prefer using the lecture method during their lessons, while 39 academics (a mere 9.8%) prefer using the project teaching method. 74 academics (18.5%) preferred combining at least two teaching methods in their lessons. Whereas 73 (18.3%) of Academics make more use of the demonstration method of teaching, 55 (13.8%) and 56 (14%) prefer discussion and experimentation methods, respectively.

A further test was conducted for the difference in the proportion of teachers preferring six different commonly used teaching methods, Chi-Square $\chi 2$ (5, N = 400) = 36.4, p=7.755. Since the calculated chi-square value of 36.4 is higher than the critical value of 11.07, the negative statement is rejected; therefore, it is concluded that the percentage of academics who prefer various teaching techniques during lessons for efficient curriculum implementation differ significantly. Looking at Figure 1, we see a high preference for the lecture method and the slightest preference for the project method. Here, there is a highly significant difference in the proportion of academics using different teaching methods during lessons.

Table 1: Demographic characteristics of the study participants. N=400.

S/N	VARIABLE	FREQUENCY	PERCENTAGE (%)
1	**GENDER**		
	Male	193	48.25
	Female	207	51.75
2	**ACADEMICS FACULTY**		
	Education	125	31.25
	Social science	129	32.25
	Arts	78	19.5
	Sciences	42	10.2
	Medical sciences	26	6.5
3	**YEARS OF TEACHING EXPERIENCE**		
	Young academics	169	42.25
	Intermediate academics	195	48.75
	Mature academics	36	9

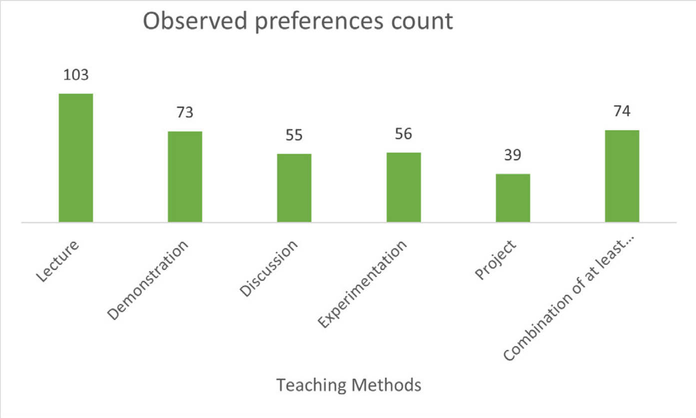


Figure 1: Bar chart showing observed teaching methods preference counts.

Hypothesis 2: No association exists between gender and teaching method preference among academics during lessons for effective curriculum implementation. Data from item 1 of the instrument were computed and subjected to the Chi-Square analysis technique to test hypothesis 2, as shown in Table 2.

A test of independence comparing the gender of Academics with their preference for teaching methods was performed. χ2 (5, *N* = 400) = 6.89, *p*=.229; at .05 level of significance; since the p-value is greater than 0.05, there is strong evidence to fail to reject the null hypothesis. The conclusion is that there is no quantitatively significant relationship between academic preference for teaching methods and gender.

It is evident from Table 2 that more male academics (57 or 14.25%) prefer lectures as compared to female academics (46 or 11.5%). More female academics from the study sample make use of a combination of at least two teaching methods (47 or 11.75%), the project (20 or 5%) and the demonstration method (40 or 10%). On the other hand, a slightly higher number of male academics prefer the discussion method (28 or 7%) and experimentation (29 or 7.25%) as against 27 (6.75%) and, again, 27 (6.75%) for female academics.

Table 2: Synopsis of chi-square study of the ratio of the association between gender and teaching method preference among academics during lessons for effective curriculum implementation; n=400.

	Gender			
Teaching method	**Male (n = 193) n (%)**	**Female (n = 207) n (%)**	**Total (n = 400) n (%)**	χ^2 **(p-value)**
Lecture	57 (14.25)	46 (11.5)	103 (25.8)	
Demonstration	33 (8.25)	40 (10)	73 (18.3)	
Discussion	28 (7)	27 (6.75)	55 (13.8)	11.07 (.229)
Experimentation	29 (7.25)	27 (6.75)	56 (14)	
Project	19 (4.75)	20 (5)	39 (9.8)	
Combination	27 (6.75)	47 (11.75)	74 (18.5)	
Total	**193 (48.25)**	**207 (51.75)**	**400 (100)**	

Not Significant at 0.05 α level; df = 5; χ^2- Cal.=6.89; χ^2-critical = 11.07; p-value= .229.

Hypothesis 3: There is no association between academics' faculty and teaching methods preference during lessons for effective curriculum implementation. Data from item 2 of the instrument were computed and subjected to the Chi-Square analysis technique to test hypothesis 3, as shown in Table 3.

From Table 3, a test of independence was calculated comparing academics' faculty with their preference for various teaching methods. χ2 (20, *N* = 400) = 24.7, *p*=.213; p>0.05l; the alpha level is less than the calculated p-value, there is strong evidence to fail to reject the null hypothesis. We conclude that there is no relationship between academic faculty and their choice of different teaching approaches. It can be gleaned from Table 3 that most academics belonging to the social science faculty prefer to use the lecture method in their lessons. At the same time, arts and education faculty academics are next in that category.

A more significant number of the respondents from the faculty of education prefer the demonstration method followed by academics from social science, Arts, Science, and Medical in that decreasing order. Another category of teaching method with a high response pattern is the Combination of at least two teaching methods; here, out of a total of 74 academics who use this method, Education and Social Science Academics both tally with 25 Academics each showing their preference for this approach. Overall, the project method is the least preferred among academics, with 39 respondents out of 400 preferring this method. Closely following is the Experimentation method, with only 56 respondents out of 400 Academics sampled for the study.

Table 3: Synopsis of chi-square study of the ratio of the association between academics' faculty and teaching method preferences.

	Academics' faculty						
Teaching method	**Education (n = 125) n (%)**	**Social Science (n =129) n (%)**	**Arts (n = 78) n (%)**	**Sciences (n = 42) n (%)**	**Medical (n = 26) n (%)**	**Total (n = 400) n (%)**	χ^2 **(p-value)**
Lecture	23(5.75)	45 (11.25)	23 (5.75)	5 (1.25)	7 (1.75)	103 (25.8)	
Demonstration	27(6.75)	21 (5.25)	14 (3.5)	7 (1.75)	4 (1)	73 (18.3)	
Discussion	22 (5.5)	11 (2.75)	12 (3)	7 (1.75)	3 (0.75)	55 (13.8)	31.41 (.213)
Experiments	15 (3.75)	17 (4.25)	11 (2.75)	6 (1.5)	7 (1.75)	56 (14)	
Project	13 (3.25)	10 (2.5)	8 (2)	6 (1.5)	2 (0.5)	39 (9.8)	
Combination	25 (6.25)	25 (6.25)	10 (2.5)	11 (2.75)	3 (0.75)	74 (18.5)	
Total	**125 (31.25)**	**129 (32.25)**	**78 (19.5)**	**42 (10.5)**	**26 (6.5)**	**400 (100)**	

Not Significant at 0.05 α level; df = 20; χ^2- Cal.=24.7; χ^2-critical = 11.07; p-value= .213.

Hypothesis 4: There is no association between years of academic teaching experience and teaching methods preference for effective curriculum implementation. Data from item 3 of the instrument were computed and analysed using the Chi-Square method to test this assumption, as shown in Table 4. For this research, young academics or early-career academics refer to individuals who have been teaching for one to ten years; intermediate academics or mid-career academics encompass those with teaching experience ranging from 11 to 20 years; and mature academics or seasoned academics include those who have engaged in teaching for 21 years or more.

As shown in Table 4, a test of independence was calculated by comparing academics' years of teaching experience with their preference for various teaching methods, χ2 (10, *N* =

400) = 7.91, p=.638; p>0.05. Since the alpha level is less than the calculated p-value, there is strong evidence to fail to reject the null hypothesis. It is inferred that academics' years of teaching experience and their preferences for different teaching approaches do not statistically correlate.

It is evident from the data that more intermediate academics (52 or 13%) prefer the lecture method of teaching, followed by young academics (40 or 10%). In the same vein, 37 (9.25%) of intermediate academics preferred the demonstration method of teaching, while young academics followed closely with 31 (7.75%). 39 (9.75%) of young academics prefer a combination of more than two teaching methods, followed by intermediate academics (27 or 6.75%). 31 intermediate academics (7.75%) preferred the discussion method, followed by young academics (20 or 5%). Only 4 (1%) of mature academics preferred the discussion method of teaching for their classes. Almost an equal number of young academics (25 or 6.25%) and intermediate academics (26 or 6.5%) preferred the experimentation method, while only 5 (1.25%) chose it. However, 22 (5.5%) intermediate academics preferred the project method, followed by 14 (3.50%) of young academics. These variations in choices by various shades of academics go to show that academics' years of teaching experience and their preferences for different teaching approaches do not statistically correlate.

Table 4: Synopsis of chi-square study of the ratio of the association between years of academics' teaching experience and teaching method preference.

	Teaching experience				
Teaching method	**Young Academics (n = 169) n (%)**	**Intermediate Academics (n = 195) n (%)**	**Mature Academics (n = 36) n (%)**	**Total (n = 400) n (%)**	**χ^2 (p-value)**
Lecture	40 (10)	52 (13)	11 (2.75)	103 (25.8)	
Demonstration	31 (7.75)	37 (9.25)	5 (1.25)	73 (18.3)	
Discussion	20 (5)	31 (7.75)	4 (1)	55 (13.8)	18.31 (.638)
Experimentation	25 (6.25)	26 (6.5)	5 (1.25)	56 (14)	
Project	14 (3.50)	22 (5.5)	3 (0.75)	39 (9.8)	
Combination	39 (9.75)	27 (6.75)	8 (2)	74 (18.5)	
Total	**169 (42.25)**	**195 (48.75)**	**36 (9)**	**400 (100)**	

Not Significant at 0.05 α level; df = 10; χ^2- Cal.=7.91; χ^2-critical = 18.31; p-Value= .638.

Discussion

The analysis of hypothesis 1 shows that the percentages of academics who prefer various teaching techniques during lessons for efficient curriculum implementation differ significantly. This might be because academics who participated in the study are professional teachers who have undergone some basic training in teaching methods. It is, therefore, easy for them to switch from one method to another. Academics prefer diverse teaching methods for efficient curriculum implementation, benefiting higher education by promoting student-centred approaches, engagement, comprehension, critical thinking, and creativity, fostering a dynamic learning environment and equipping students with necessary skills. This finding agrees with Nwogu and Esobhawan (2014), who observed that teaching involves practical communication skills, effective teaching strategies, and classroom management techniques. In a classroom situation, a teacher, a lecturer, or an academic utilizes various teaching methods while executing any lesson.

The finding of this study also agrees with the thoughts of Sikaleya (2022), who counted over fifty teaching methods in practice for teachers to use in education. The findings of this study, however, disagree with the observation of Dharmambal and Anavarathan (2021), who stated that the present mode of instruction in medical schools is the lecture, yet lectures by themselves are inadequate for fostering cooperative learning and skill development.

The results of hypothesis 2 reveal that there is no quantitatively significant relationship between academic preference for teaching methods and gender. A plausible explanation of the finding is that teaching itself is an art; it requires the teacher to apply the appropriate methods in the lesson, notwithstanding the teacher's gender. This finding agrees with the findings of Ambe and Onnoghen (2018), who, in their study, found that teachers' gender has no significant influence on teaching effectiveness. Therefore, regardless of the gender of the teacher, if they have the requisite professional competence and are prepared to use appropriate teaching methods and adequately deliver the lesson, effective learning will take place.

Positive results for teaching and learning in higher education can be seen in the lack of a quantitatively significant association between academic preference for teaching techniques and gender. It highlights the significance of fair and inclusive educational practices, supports creative teaching methods, and promotes an atmosphere in which all students can flourish and realise their full potential. Higher education institutions can provide a more engaging and encouraging learning environment for students of both genders by focusing on instructional effectiveness and personalised approaches.

The results of hypothesis 3 show no statistically significant association between academics' faculty and their preference for various teaching methods. This finding agrees with the Illinois State Board of Education (ISBE, 2002), which argued that a trained educator is skilled in all academic disciplines and exhibits subject-matter expertise. The finding also agrees with Ambe and Agbor (2014), who argued that a teacher needs to have a broad and liberal education, strong topic knowledge, sound teaching techniques, an understanding of child psychology, and knowledge of societal variables impacting students who attend school.

The analysis of hypothesis 4 shows no statistically significant association between academics' years of teaching experience and their preference for various teaching methods. This finding disagrees with Ambe and Agbor (2014), who noted that seasoned educators draw from a broader and more complex body of information than upcoming ones. The finding, however, agrees with Niemelä and Tirri (2018) and Mupa and Chinooneka (2015), who argued that there are several categories of knowledge that seasoned teachers pick up, including knowledge of the subject's fundamental concepts, often known as subject area knowledge. Knowing how to make a subject interesting and understandable is known as pedagogical content knowledge, among other things. This finding does not agree with Clotfelter et al. (2010), who show that over 20 years of experience is more effective than no experience; nonetheless, they contrast in efficacy with teachers who have five years of experience.

Conclusions

The study examined sociodemographic factors and teaching method preferences among university academics, highlighting the significance of this information for effective curriculum implementation. The findings indicate notable variations in teaching technique preferences among academics, challenging the idea of uniform teaching approaches in higher education. The study found no significant correlation between gender and academic preference for teaching techniques, indicating that academics' preferred teaching approaches are not significantly influenced by their gender. Academic faculty type did not significantly influence their choice of teaching strategies, suggesting consistent preferences across positions and disciplines. Years of teaching experience did not significantly correlate with academics' preferences for different teaching methods. We must keep in mind, though, that the phrase 'no statistically significant relationship' does not imply that there is, in fact, no relationship at all; rather, it indicates that the study did not uncover enough evidence to establish a meaningful relationship based on the selected statistical criteria.

Recommendations

The adoption of teaching methods that promote constructivist learning, such as the project method, experimentation, and demonstration methods, is encouraged for academics to enhance student engagement and active participation in the classroom. Gender and faculty type do not significantly influence teaching method preferences, but it is crucial to recognize and respect individual differences. Offering professional development workshops and training sessions can enhance pedagogical skills, regardless of gender. Promoting interdisciplinary collaboration among educators can lead to innovative methods across disciplines. Mentoring programs and platforms for less experienced educators can contribute to a well-rounded teaching environment. A holistic approach to pedagogical enhancement should consider the intersection of factors like gender, faculty type, and teaching experience. Educational institutions should promote flexibility and adaptability in curriculum design and delivery, allowing educators to experiment with different techniques and adjust methods based on student feedback. Establishing mechanisms for academic feedback on teaching methods and preferences can help institutions stay attuned to educators' evolving needs and preferences.
To ensure the effective implementation of these recommended teaching techniques, university authorities should implement monitoring mechanisms for instructional practices. One approach to achieving this is through the incorporation of ICT-based learning management systems, which can help track and assess the integration of prescribed and appropriate teaching methods into academics' lessons. This proactive approach to monitoring can support continuous improvement in teaching practices and contribute to a more enriching educational experience for students.

References

Abdulbaki, K., Suhaimi, M., Alsaqqaf, A., & Jawad, W. (2018a). The impact of using the lecture method on teaching English at university. *European Journal of Education Studies, 4*(5), 285-302. doi: 10.5281/zenodo.123887

Abdulbaki, K., Suhaimi, M., Alsaqqaf, A., & Jawad, W. (2018b). The use of the discussion method at university: Enhancement of teaching and learning. *International Journal of Higher Education, 7*(6), 118-128.

Alaagib, N. A., Musa, O. A., & Saeed, A. M. (2019). Comparison of the effectiveness of lectures based on problems and traditional lectures in physiology teaching in Sudan. *BMC Medical Education, 19*(1), 1-8.

Al-Rawi, I. (2013). Teaching methodology and its effects on quality learning. *Journal of Education and Practice, 4*(6), 100-105.

Amadioha, S. W. (2017). *Essential teaching methods for education and behavioural sciences*. Sabcos Printing and Publishers.

Ambe, B. A. & Agbor, C. A. (2014). Assessment of teacher's professional competence, teacher's teaching experience on the implementation of environmental education curriculum in cross river state, Nigeria. *Journal of Educational and Social Research, 4*(6), 521-529.

Ambe, B. A., & Onnoghen, N. U. (2018). Teachers' gender and self-concept as predictors of environmental education curriculum implementation in cross river state schools, Nigeria. *Multidisciplinary Journal of Research and Development Perspective, 7*(1), 36-44.

Anderson, V. L., & McLean, R. A. (2018). *Design of experiments: A realistic approach.* CRC Press.

Bada, S. O., & Olusegun, S. (2015). Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education, 5*(6), 66-70.

Bala, P., Kaur, T., & Kaur, M. (2017). Study on effectiveness of lecture and smart class method of teaching on academic achievements among upper primary school students. *International Letters of Social and Humanistic Sciences, 76,* 25-29.

Bhat, R. L. (2017). A study of teaching effectiveness of prospective teachers in relation to stream and gender. *Amity International Journal of Teacher Education, 3*(1), 1-6.

Chingala, C. (2020). *Exploring the teaching of geometric and spatial reasoning in early childhood education in selected primary schools in Shibuyunji district* (Doctoral dissertation, The University of Zambia).

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources, 45*(3), 655-681.

Dharmambal, C. & Anavarathan S. (2021). Comparison of seminar and lecture as teaching method and student perception on the seminar and lecture in medical education. *International Journal of Research in Dermatology, 7*, 188-191.

Emmanue, I. A. & Ambe, B. A. (2014). 'Influence of teachers, professional qualification, and area of specialisation on the implementation of environmental education curriculum in cross river state–Nigeria. In *International Conference on Chemical, Environment & Biological Sciences (CEBS)* (Vol. 914120, pp. 155–160). http://dx.doi. org/10.15242/IICBE C

Ephraim, I. E., Ambe, B. A., Udumo, B. O., Onnoghen, N. U., Egbonyi, E. E., Omoogun, R. M., ... & Okon, H. P. (2022). A survey of virtual instructional competence among tourism teachers in secondary schools. Calabar South local Government Area, Nigeria. *Education and Information Technologies, 27*(8), 11307-11323.

Eze, T. I., & Nwaukwa, V. E. (2019). Effectiveness of demonstration method in entrepreneurship education in tertiary institutions in south Nigeria. *NAU Journal of Technology and Vocational Education, 3*(1), 44-53.

Hajar, I., Umanailo, M. C. B., Umanailo, R., Indriani, D. E., Handayani, N., & Nursyifa, A. (2021). The effectiveness of demonstration method in learning poetry for grade X students of SMA Negeri Sawa. In *Proceedings of the 11th annual international conference on industrial engineering and operations management* (pp. 6198-6208).

Holah.Karoo. (n.d). *The experimental method.* http://www.holah.karoo.net/experimental_method.htm.

Hussain, M. A. (2020). Effectiveness of demonstration method to teach the abstract concepts to the children between the age of six to ten. Experimental research. *International Journal of Education (IJE), 8*(2), 23-32. DOI: 10.5121/ije.2020.8203 23

Illinois, State Board of Education (ISBE). (2002). *Illinois professional teaching standard* (2nd ed.). http://www.westga.edu/

Kapur, R. (2020). *Lecture method: The comprehensively used method.* Researchgate.net/publication/345893936_lecture_method_comprehensively_used_pedological_method.

Kim, K. R., & Seo, E. H. (2018). The relationship between teacher efficacy and students' academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal, 46*(4), 529-540.

Kim, L. E., Jörg, V., & Klassen, R. M. (2019). A meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. *Educational Psychology Review, 31*, 163-195.

Knoll, M. (2014). Project method. In D. C. Phillips (Ed.), *Encyclopedia of educational theory and philosophy* (pp. 665-669). Sage.

Kołodziejski, M., & Przybysz-Zaremba, M. (2017). Project method in educational practice. *University Review, 11*(4), 26-32.

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). *Descriptive analysis in education: A guide for researchers.* Department of Education, Institute of Education Sciences, National Centre for Education Evaluation and Regional Assistance.

Muhammad, A. U., Bala, D., & Ladu, K. M. (2016). Effectiveness of demonstration and lecture methods in learning concept in Economics among secondary school students in Borno state, Nigeria. *Journal of Education and Practice, 7*(12), 51-59.

Mupa, P., & Chinooneka, T. I. (2015). Factors contributing to ineffective teaching and learning in primary schools: Why Are schools in decadence?. *Journal of Education and Practice, 6*(19), 125-132.

Niemelä, M. A., & Tirri, K. (2018). Teachers' knowledge of curriculum integration: A current challenge for Finnish subject teachers. In Y, Weinberger & Z, Libman (Eds.), *Contemporary pedagogies in teacher education and development* (pp. 119-132).

Noel, M. D., Daniels, F. M., & Martins, P. (2015). The future of lecture method as a teaching strategy in community nursing education. *Journal of Family Medicine & Community Health, 2*(8), 1-4.

Nwogu, U. J., & Esobhawan, B. I. (2014). Teacher quality and quantity: Implications for the implementation of universal basic education in Nigeria. *Academic Research International, 5*(1), 199-208.

O'Neill, S. A. (2014). Mind the gap: Transforming music engagement through learner-centred informal music learning. *The Recorder: Journal of the Ontario Music Educators' Association, 56*(2), 18-22.

Omotayo K. A., & Adedeji, F. T. (2020). Effects of demonstration teaching method on senior secondary school students' academic performance in Biology in Ekiti State, Nigeria. *Social Science Education Journal (SOSCED-J), 4*(2), 15-21.

Prtljaga, S., & Veselinov, D. (2017). The influence of the project method on the achievement of young learners in the field science and social studies. *Research in Pedagogy, 7*(2), 254-264. DOI: 10.17810/2015.63

Reference. (2020, March 26). *What are some advantages of the discussion method?* https://www.reference.com/world-view/advantages-discussion-method-c4d1bd5e9758cc19.

Sadeghi, R., Sedaghat, M. M., & Ahmadi, F. S. (2014). Comparison of the effect of lecture and blended teaching methods on students' learning and satisfaction. *Journal of Advances in Medical Education & Professionalism, 2*(4), 146-150.

Shah, S. R., & Udgaonkar, U. S. (2018). Influence of gender and age of teachers on teaching: Students perspective.

International Journal of Current Microbiology and Applied Sciences, *7*(1), 2436-2441.

Sikaleya, G. (2022). *50 methods of teaching.* Independent. academia.edu/GRACESIKALEYA.

Soares, B. C., Castelhano de Campos, M. E., Thomaz, J. R., Pereira, G. C & Roehrs, R. (2016). The importance of experimentation in the teaching of sciences to elementary school. *Revista Monografias Ambientais - REMO Revista do Centro do Ciências Naturais e Exatas, 15*(2), 1-17. DOI:10.5902/2236130827003

University of Buffalo. (2022). *Curriculum, assessment, and teaching transformation.* Buffalo.edu/catt/develop/design/teaching-methods.html

Visuanathan, M., & Visauanathan, K. V. (2017). Effectiveness of didactic lecture in imparting knowledge in medical education. *International Journal of Medical Science and Education. An Official Publication of Association for Scientific and Medical Education (ASME),4*(3), 247-251.

Ying, J. (2020). The Importance of the discussion method in the undergraduate business classroom. *Humanistic Management Journal*, *5*(2), 251-278.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Engaging students through messaging applications in foreign language learning

Patricia Lorenz[A]

Lydia König[B]

[A] Senior Lecturer and Fellow, Ridge View Residential College, National University of Singapore

[B] Transmedia Writer, Science Centre Singapore

Keywords

Facilitation;
foreign language acquisition;
messaging apps;
online learning;
peer-to-peer learning;
peer mentor;
Social Capital Theory;
Social Exchange Theory;
theory of instinctive information-sharing behaviour.

Correspondence

plorenz@nus.edu.sg [A]

Article Info

Received 20 June 2023
Received in revised form 2 October 2023
Accepted 2 October 2023
Available online 3 October 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.30

Abstract

This study explores undergraduate students' experiences and perceptions of group discussions facilitated by eStudentMentors in WhatsApp or Telegram groups in the context of German language learning. eStudentMentors are senior peers who volunteer to support novice language learners. The study was conducted with 159 students enrolled in German Language Level 1 and German Language Level 2 at a university in Singapore. Each of the language classes comprises 15-21 students and are assigned an eStudentMentor to facilitate online learning and discussions in a dedicated WhatsApp or Telegram group. With considerations to information-sharing behaviour, the study examines how receptive students are to online information-sharing in this context. As such, a combined quantitative and qualitative online survey questionnaire was used to collect data, with survey questions examining information-sharing behaviour governed by personality traits and based on the Social Exchange Theory, Social Capital Theory, and theory of instinctive information-sharing behaviour. Overall, 55.97% of the students who completed the survey used the WhatsApp or Telegram groups created by their eStudentMentor to ask or answer questions and share information at least once throughout the semester. In comparison, the other 44.13% were completely inactive in their chat groups. This analysis examined the reasons behind this group of students' inactivity and found that a major factor for individual inactivity was the overall inactivity of the chats, which makes the active facilitation of the eStudentMentor a crucial element for success. The lack of social bonds appears to be another main reason for inactivity. Additional factors are class size, unfamiliarity with their classmates, fear of judgement, feeling awkward, having negative assumptions, and low commitment levels. The findings counter the Social Exchange Theory (SET) propositions, as the greater benefits of information sharing are overshadowed by the seemingly trivial cost. Yet, these costs in the form of social perceptions and pressures appear to accumulate and collectively outweigh the rationally perceived benefits to the users.

Introduction

This study explores undergraduate students' experiences and perceptions of group discussions facilitated by eStudentMentors in WhatsApp or Telegram groups. In this technological age, many universities across the globe are adopting e-learning to complement traditional teaching and learning methods (Belias et al., 2013; Kasraie & Kasraie, 2010). In comparison to traditional teaching and learning methods, modern methods that incorporate the use of technology, social media, and the internet are more attractive to students and have improved student performance (Sirbu et al., 2014). Hardy et al. (2023) also observed that well-constructed online learning can contribute to the learners' wellbeing and foster a sense of connection in a learner community.

eStudentMentors are senior learners who volunteer to guide level 1 (LG5001) and level 2 (LG5002) German learners, who learn German at a Common European Framework of Reference for Languages (CEFR) A1 Level, through a dedicated chat group in their learning journey (Common European Framework of Reference for Languages, 2023). The eStudentMentors are recruited from higher level courses in which students learn German at a CEFR B1 level. The study was conducted with a cohort size of 159 students enrolled in German Language Level 1 (LG5001) and German Language Level 2 (LG5002) at a Singaporean university. Each German language class, comprising 15-21 learners, was assigned an eStudentMentor to facilitate and encourage class discussions in a WhatsApp or Telegram group. The students enrolled in the classes were free to use the platform to engage in further learning outside classroom time; be it to clear doubts or share resources. With considerations to information sharing behaviour an online survey comprising quantitative Likert scale questions and qualitative questions was developed. The study aims to examine whether peer mentors (eStudentMentors) are able to extend learning beyond the classroom and if it is widely used and accepted by students. The study also examines how receptive students are to stimulated online information sharing in a group chat, and if it positively affects students' overall learning experience and performance.

Blended learning

Blended learning combines traditional face-to-face (FTF) instruction with technology-mediated instruction, while emphasising the central role of technology (Bonk & Graham, 2005). Essentially, the technological part takes place outside the FTF teaching location, subject to individual students' own time, space and pace (Hockly, 2018). A survey done in 2020 found that 94% of lecturers thought blended learning to be one of the more active approaches compared to just traditional FTF learning (Widyasari et al., 2020). Such interactive teaching approaches benefit students by improving their understanding and increasing their subjective learning gains (Alonso et al., 2011). Kember (2010) has also noted that blended learning increases students' learning productivity and improves their communication skills by enabling active participation and constructive communication (Gecer & Dag, 2012; Sirbu et al., 2014). Blended learning facilitates prompt feedback and increased accessibility to both information technology and human resources (Poon, 2013). Based on the previously recorded benefits, this study begins with a promising projected outcome for students through the use of modern, interactive, and technological learning.

Stein and Graham (2020) argue that in order to choose the technological aspects of blended learning, teachers should focus less on the technology itself and instead focus on how it enables learning. Studies have pointed out some disadvantages of technology itself, such as some platforms being user-unfriendly or facing overwhelming technical difficulties, which hinder learning (Szadziewska & Kujawski, 2017). Hence, for the purpose of this study, a medium thought to be low-maintenance and user-friendly was social media—a platform that is readily available and familiar to students in a Singapore context. Social networking sites, such as Facebook and Instagram, are one of the most visited sites on the internet. Along with their open interface and popularity, these sites enable effective communications, diminish social barriers and are deemed a good medium to facilitate learning (McCarthy, 2010). Chat groups allow students to discuss, clear doubts, and share information in real-time. Students are also no strangers to these online platforms, which have long been successfully employed to support university students in language learning (Conroy, 2010). In one study by Shih (2011), the online platform enabled students to continue to learn English as a foreign language outside of class time and thereby provided more flexibility in the learning process, enabling learner-centric interactions and practice that helped students acquire new knowledge. However, a disadvantage that has been pointed out is that students are not motivated or disciplined enough to put aside time outside the classroom to engage in learning (Tosun, 2015).

Messaging applications, such as WhatsApp and Telegram, have long been identified to be a large part of the technological trends after the web-based social networks (Cetinkaya, 2017). It is also argued that WhatsApp surpasses social networking sites, such as Facebook, in terms of educational purposes, connectivity, and ease of usage (Rani et al., 2019) since it is swifter and more efficient to facilitate real-time communication (Kamel et al., 2016). WhatsApp, in particular, is one of the most used mobile communication tools in the world (Statista, 2016). Based on the application's flexible and informal nature, WhatsApp has been proven to facilitate seamless and informal learning (Annamalai, 2018). This communication tool has also been proven to support learning outside of class time by increasing students' interest and motivation, creating a sense of belonging, and enabling peer support, as well as information sharing among students (Cetinkaya, 2017; Raiman et al., 2017). Previous studies have also demonstrated that messaging applications have helped enhance the university experience (Nitza & Roman, 2016). Thus, the use of WhatsApp or Telegram was adopted over social networking sites for this research, with WhatsApp being used in the earlier implementations of eStudentMentor run group chats and Telegram being predominantly used in consecutive runs. The Singaporean students' growing preference for Telegram is driven by the ability to create groups based on user names only, thus avoiding having to save each chat member's number in the phone.

The use of eStudentMentors

Studies have also shown that student mentorship enhances participation, engagement, and acquisition of knowledge, contributing to a more satisfying learning experience and academic performance (Bhatia et al., 2016). Deri (2022) notes that students are better equipped to succeed and to reach their learning goals if they witness someone else succeeding, which highlights the role model function of eStudentMentors recruited from higher-level language learners. According to Snowden and Hardy's (2013) findings, having a student mentor eases a mentee into a social environment and helps him or her to become more comfortable with their peers, which later enables them to reflect better on their thinking. Indeed, a mentorship programme can contribute to learning in light of the fact that students consider continuous access to an instructor, or a figure with more knowledge than them to be vital in the context of blended learning (Martinez-Caro & Campuzano-Bolarin, 2011).

This study goes a step further by replacing the 'faculty figure' with a fellow student, albeit from a higher level, to create an even more informal and accessible learning environment. Informality is important as students perceive it to be more amiable, which helps to combat detached interactions between faculty and student (Dumford & Miller, 2018). Thus, this study combines the use of peer mentorship on top of the communication application in hopes that it will boost learning outcomes. The study is conducted in a foreign language learning environment at the university level that has implemented blended learning through a flipped classroom approach. As such, total beginners to German language learning are tasked to familiarise themselves with grammatical content prior to the FTF classes, with the aim to engage students in more seamless and self-directed learning, where students do not feel the necessity to set aside specific time to contribute to their learning while being able to familiarise themselves with new materials at their own pace. Hence, the eStudentMentors are engaged in order to support and motivate the A1 level learners in this self-directed learning process.

The eStudentMentors were recruited from higher-level German language learners to act as role models and to provide a seemingly informal platform for learning beyond the classroom. All eStudentMentors were volunteers, who were neither paid nor rewarded in any other form. While each eStudentMentor was in charge of one class of learners, the department head supported the peer mentors in a chat in order to clarify queries or provide advice in case complex questions were asked in the individual chats. Yet, in order to maintain the informal and highly individual nature of the various chat groups, no specific routine or schedule was implemented. Thus, the eStudentMentors displayed varying levels of engagement and motivational strategies in their respective group chats. The eStudentMentors were also left to decide whether they wanted to facilitate the group chat on WhatsApp or Telegram and in the context of this study, no differentiation was made between WhatsApp or Telegram chat groups.

Methodology

After the implementation of the eStudentMentor facilitated chat groups, it was found that only very few group chats were active throughout the semester. The need to identify the causes for this inactivity was deemed important. Consequently, a Qualtrics survey was conducted online, using questions to collect qualitative and quantitative data in order to understand and evaluate undergraduate students' experiences and perceptions of group discussions facilitated by eStudentMentors in WhatsApp or Telegram groups.

All 159 students enrolled in the CEFR A1 courses were encouraged to participate in the survey. Surveyed students were assured that participation in the survey is optional, and that survey data is anonymised. Surveyed students had 12 weeks to use the group chats for their learning prior to the online survey being conducted. The survey structure and questions used in this study were adapted from Wang and Chan's (2011) survey, examining information-sharing behaviour governed by personality traits and based on the Social Exchange Theory (SET), Social Capital Theory (SCT), and the theory of instinctive information-sharing behaviour (Widen-Wulff, 2014).

The online survey in this study consisted of 12 sections and a total of 53 questions, 39 of which were Likert scale questions, with choices ranging from 'strongly disagree' to 'strongly agree' on a 6-point scale, and from 'never' to 'daily' on a 5-point scale. The former was used to gather how agreeable students were to the learning conditions, while the latter measured the frequency with which students engaged with the set medium. 13 of the other questions in the survey were open-ended, and one was a yes-no question. Each of the 12 sections terminated with one of the open-ended questions, which were deemed critical in understanding the reasons behind students' agreements, or lack thereof, with the learning platform. (A copy of the questionnaire can be found in Appendix A.)

The 53 questions were categorised into 12 segments, with each segment covering a different aspect of the theories that govern information-sharing behaviour. The first segment elicited responses on general group chat behaviour. The second and third segments explored the intrinsic benefits of Social Exchange Theory and the theory of instinctive information-sharing. The theory of instinctive information-sharing behaviour posits that it is natural for individuals to share information with other people (Fehr et al., 2008). Segments four to eight focus on the Social Exchange Theory under the consideration of cost, as well as intrinsic and extrinsic motivation. The Social Exchange Theory (SET) examines the motivations behind human social interactions and considers the costs and rewards of social behaviours (Homans, 1974). Lastly, the final segments evaluate pro-sharing norms and dimensions in the Social Capital Theory. The Social Capital Theory (SCT) also governs information-sharing behaviour. SCT examines social capital resources, which, according to Putnam (2000), arise from individual relationships and the benefits that they entail. By employing theories that govern information-sharing behaviour, this analysis seeks to understand the reasons behind the German language learners' behaviour in their eStudentMentor-facilitated WhatsApp or Telegram groups.

Findings

Out of 159 students, 93.7% fully completed the survey. Mean values are used in this study to analyse Likert scale data; a smaller mean value indicates a more positive response and vice versa. Values less than 3.5 are considered positive, while values above and equal to 3.5 are negative. All graphics illustrating the results of Likert scale data questions indicate the number of students choosing the respective answers.

55.97% (81 out of 145 students) of the students who completed the survey used the WhatsApp or Telegram groups created by their eStudentMentor to ask or answer questions and share information pertaining to the study of the German language at least once throughout the semester. Students who were active in their chat groups understood the purpose of the chat groups and used them to share resources, exchange ideas, seek or provide clarifications and support their peers. The other 44.13% (64 out of 145 students), however, were completely inactive in their chat groups.

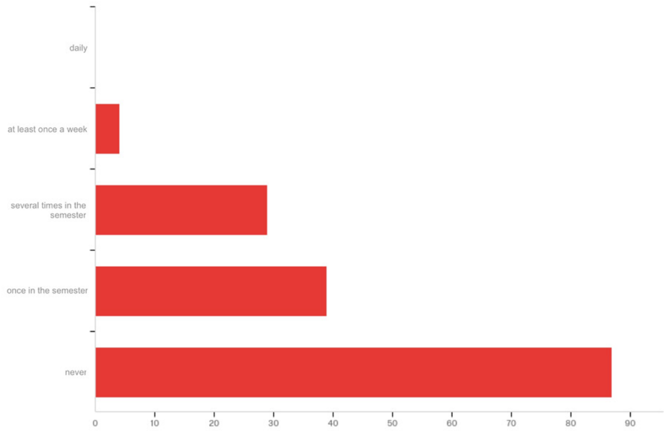


Figure 1: Q1.1. I have asked questions in my WhatsApp/ Telegram group.

The theory of instinctive information-sharing behaviour suggests the natural inclination of individuals to share information with others (Warneken & Tomasello, 2007), which was largely true for the students who participated in this study. In stark contrast to the recorded inactivity, 88.27% (128 out of 145 students) surveyed felt that it was natural for them to share information with other students, 90.34% (131 out of 145 students) reported that when they learned something new, they wished to share it with other students, and 87.58% (127 out of 145 students) preferred to share information with other students rather than keep it to themselves. Clearly, more than half of the students who were completely inactive in their chat groups made a conscious decision to go against their instinctive desire to share information with their peers.

92.41% of surveyed students (134 out of 145 students) agreed that active participation in information sharing via their chat groups benefited their learning. 88.27% (128 out of 145 students) agreed that asking questions in their

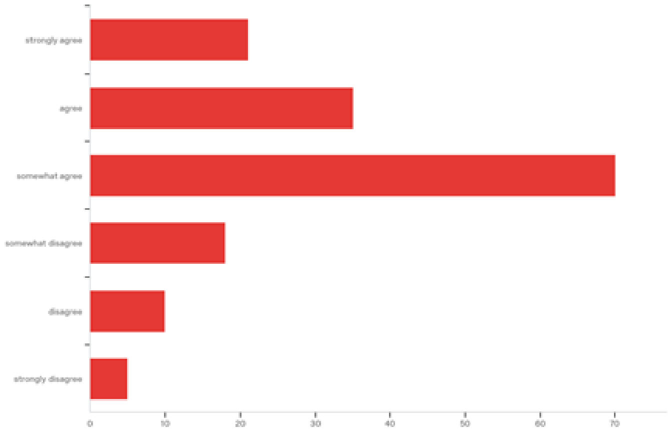


Figure 2: Q2.1. I enjoy sharing information with others in our WhatsApp/Telegram group.

chat groups helped with their learning, and 89.65% (130 out of 145 students) agreed that answering questions in their chat groups helped with their learning. Yet again, this understanding stands in contrast to the actual chat group participation rate.

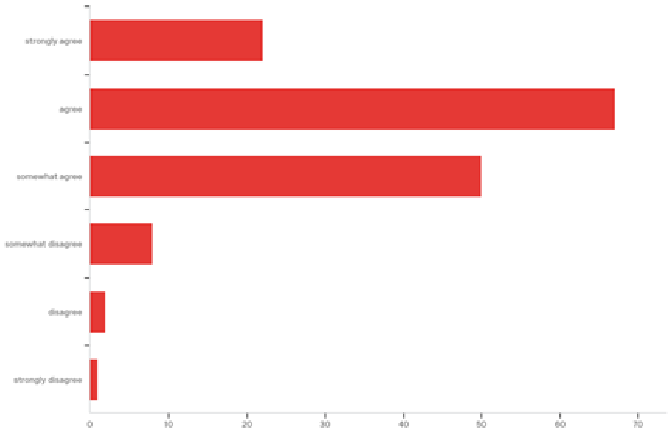


Figure 3: Q8.1 Active participation in the information sharing benefits my learning.

Despite being informed at the beginning of the semester that their eStudentMentors would only be facilitating the discussions rather than providing them with answers, 90.45% (142 out of 145 students) expected their eStudentMentors rather than their peer learners to answer their questions when surveyed.

Surprisingly, only 17.24% (25 out of 145 students) thought that sharing information in their group chats made them lose the knowledge that made them stand out with respect to other students. 15.17% (22 out of 145 students) thought that sharing information in their chat groups made them lose power over knowledge that no one else had, and 13.10% (19 out of 145 students) thought that sharing information in their chat groups made them lose their unique value. Out of those who did not utilise the chat group, 20.31% (13 out of 64 students) thought they would be disadvantaged in any of the above mentioned reasonings that would derive from information sharing.

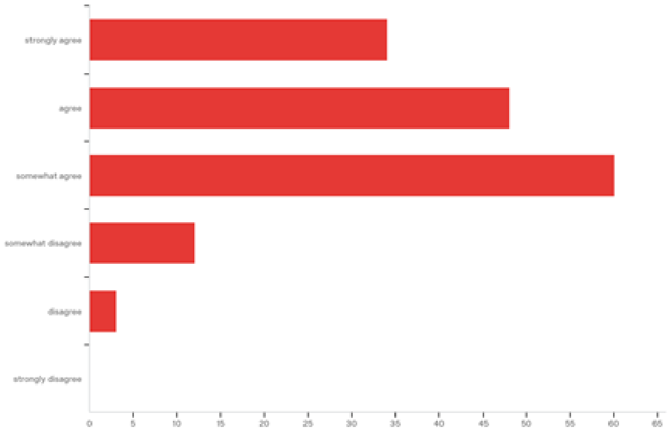


Figure 4: Q4.3. When I ask questions in my WhatsApp/ Telegram group, I expect my eStudentMentor to answer my questions.

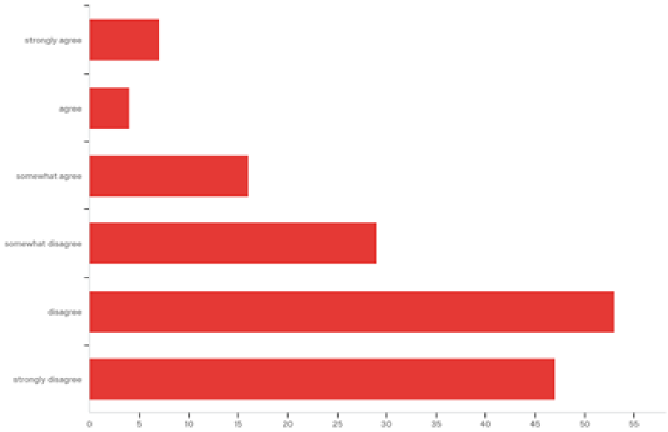


Figure 5: Q5.1. Sharing Information in our WhatsApp/ Telegram group makes me lose my knowledge that makes me stand out with respect to others.

Despite a large number of students not using the eStudentMentor facilitated group chats, 97.24% of surveyed students (141 out of 145 students) created a separate small group chat (Team Chats) with their oral assessment team members. Students are grouped in teams of three for the final group assessment, and the survey demonstrated that Team Chats are more frequently used than the eStudentMentor chat groups.

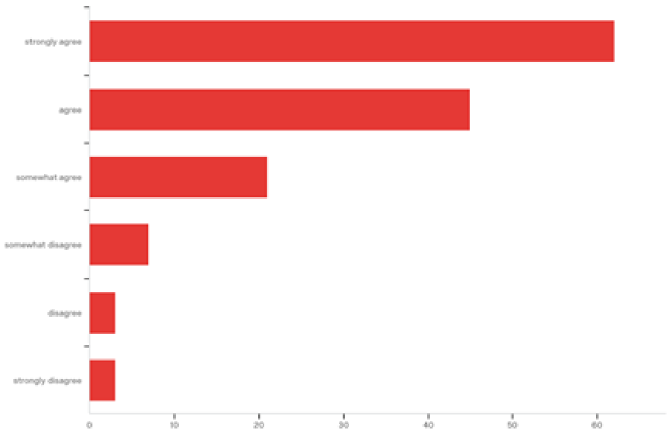


Figure 6: Q12.2. I ask my questions in my Team Chats rather than in my eStudentMentor chat group.

Analysis and discussion

This analysis seeks to understand the reasons behind this group of students' inactivity in their eStudentMentor facilitated chat groups, as 44.13% of students (64 out of 145 students) were completely inactive throughout the semester. Most eStudentMentors made some attempts to keep the chat groups active by posting subject-relevant questions or information about events, yet hardly any managed to initiate extensive discussion. Thus, despite the natural inclination of surveyed students to share information with others, more than half of the ones who were completely inactive in their group chats made a conscious decision to go against their instinctive desire to share information with their peers. The Social Exchange Theory provides a possible explanation for this behaviour.

The Social Exchange Theory (SET) consists of various propositions that influence information-sharing behaviour, the following of which are applicable to this study: value, success and rationality. The success proposition is determined by the likelihood or success rate in obtaining information; the value proposition focuses on the value of actions; the rationality proposition is determined by the success rate, value of action, and reasoning behind the action (Homans, 1974). The open-ended questions revealed that one reason that stopped some students from participating in their group chats was the lack of activity. Due to the chat group being quiet and with few asking questions, students did not expect any information-sharing, as evident below:

Not many people asking/answering questions.
It is too quiet in the group.
There isn't much activity in the group in the first place.

This lack of activity meant that students were less likely to be rewarded for asking questions or sharing resources in their chat groups. As Homans (1974) points out, if an action were proved to be successful, one would repeat the action in the future and vice versa. This ties in with the success and rational propositions of SET, which suggest that students were less willing to engage in discussions in their WhatsApp groups if they were less likely to be rewarded for it and that they had rationalised the fact that the lack of activity meant a lack of information.

Wang (2013) explains that one is likely to partake in information sharing if it generates more rewards than punishments. These considerations form the basis of the following equation by Molm (1997), which serves to predict social behaviours:

Behaviour (Profits) = Rewards of Interaction – Costs of Interaction

Applied to this study, the above equation takes the following form:

Likelihood of Utilising Chat Group = Rewards of Utilisation – Costs of Utilisation

In this study, 'utilisation' refers to asking and answering questions or sharing useful links and images related to the learning of the German language in chat groups. Costs refer to how a student's action affects or takes away something from them. Rewards are the benefits students obtain from utilising the chat groups. Divided by intrinsic and extrinsic motivational factors, these benefits form the basis of motivation theory.

An intrinsic motivational factor arises from the pleasure an individual derives from engaging in an activity, which makes the means (in itself) an end (Deci & Ryan, 2010). According to the survey results of this study, the mean value of students having enjoyed sharing information through the chat group is 2.85. Generally, students feel good from sharing or helping a fellow classmate or contributing to the 'greater good' of their class. Interestingly, however, 100% of those who had never used the WhatsApp or Telegram group agreed that they enjoyed sharing information. For this group of students, the pleasure that they could derive from sharing information in their chat groups appeared insufficient to motivate them to utilise the chat groups.

An extrinsic motivational factor, on the other hand, can be derived from a utilitarian value that results from a perceived overall usefulness (Davis, 1989; Limone et al., 2019). In this study, students attributed the perceived overall usefulness of their group chats to refining their understanding of the course's syllabus, clarifying their doubts and misconceptions, and improving their proficiency in the German language. According to the survey results, there is a mean value of 2.36 for students' perception that sharing information in their group chats benefited their learning. Yet, while the understanding of the benefits of asking questions stands in contrast to the actual chat group participation rate, many students could explain the ways in which sharing information benefited them in the open-ended questions.

Students realised that they could not only learn from mistakes they were unaware of, but from their peers' mistakes as well. The community presence made it possible to bounce off and clarify from each other's doubts. Additionally, the different explanations or thoughts from other students could bring about a wider range of perspectives for an individual student. The explanations, in turn, are good for revision, and is also a form of learning for the explainer as well. One student noted that it was especially helpful that others were asking the questions that shy students were too afraid of. Overall, these reasons emphasise the understanding of the benefits of the eStudentMentor facilitated chat group. These reasons provided by the students in the open-ended questions were as follows:

People could help to correct your mistakes (that you did not know).
Teaching is a form of learning.
Explanations from people may bring about different perspectives on issues and everyone learns better.
We learn things that we might not already know. Learn from other people's mistakes.
The forum helps make me aware of doubts others have concerning the content. Any discussion on those queries will benefit my learning experience as well.
Asking questions will help to clarify my doubts and similar doubts that others may have but choose not to ask about. Answering questions helps to clarify the doubts of others and gives everyone an opportunity to learn if I make a mistake.
It allows me to revise concepts I already learnt and trying to think of new ways to better explain it improve my understanding of the concept.

Again, 93.75% (60 out of 64 students) of those who did not use their chat group at all understood that engaging with the group in any way would benefit their learning. Yet, they consciously chose not to use their chat groups despite knowing that it could benefit their learning.

Another extrinsic motivational factor is reciprocity, which is fuelled by one's expectation of receiving help in return for helping other people (Connolly & Thorn, 1990; Hung et al., 2011; Kollock, 1999). This motivational factor applies to the students who participated in this study since they understood that communication was not one-way where they could take without giving and that it helped contribute to the group's 'greater good', as evident in the explanations they provided:

Because if I help others, I expect others to reciprocate, if not it's just one-way traffic.
I help you, you help me, let's make this world a better place.

In fact, 77.24% (112 out of 145 students) who participated in this study expected their peers to discuss their questions in the chat groups, and 84.71% (123 out of 145 students) expected their peers to answer their questions in the group chats. Some students clarified that it was precisely due to the purpose of the chat groups that they had such expectations while other students explained that their expectations arose from their perceptions that some of their peers were smarter than them and hence would be able to provide the solutions to their questions, or perhaps they would do so purely out of politeness or a desire to reciprocate favours. Yet, 90.45% of surveyed students (142 out of 145 students) expected their eStudentMentors to answer their questions instead of taking responsibility for their own learning by participating actively in group discussions. Some students even blamed the eStudentMentors for not providing them with the answers they wanted and deemed both the eStudentMentors and the chat groups 'quite useless'. On the other hand, since the eStudentMentors existed outside the realm of competition within each class, some students reasoned that the eStudentMentors were in a much better position to provide answers to students' questions. Such reasoning resonates with the value proposition of SET as they may no longer see value in the group due to the eStudentMentors not meeting expectations. Another possible explanation for this might be a lingering dependency on a perceived higher authority within a peer-learning setting.

In theory, participating actively in their chat groups would also reward students by improving their image, considering that they would appear to be more knowledgeable than their peers. Thus, the survey measures students' perceptions with respect to image. On one hand, 46.20% (67 out of 145 students) agree that those who share knowledge have more prestige, while 37.93% (55 out of 145 students) agree that answering questions made them look smart. On the other hand, however, 24.17% (35 out of 145 students) agree that asking questions would make them look stupid. Although the majority seem unfazed by how their interactions with the chat group affect their image, there are still a handful who do take their image into consideration. For these students, the act of sharing and answering is taken positively while the act of asking is taken negatively. Responding to enquiries boosts their image. Asking, however, according to the following responses, scares them as they do not want to sound 'stupid'. The fear of appearing 'stupid' surfaced several times in the answers to the open-ended question in question block pertaining to extrinsic motivations. This is considered a cost in utilising the chat group.

Afraid to sound stupid
Slightly scared to ask a stupid question
Asking is part of learning. However, if we get a sarcastic reply to our question, we may feel stupid.

From the last response above, the students' attention towards another's sarcastic reply also highlights how students discern their peer's perceptions. The response suggests that students perceive there is a consensus that their peers also subscribe to, whereby asking is a call for humiliation. Out of those who were completely inactive in their chat groups, 31.25% (20 out of 64 students) follow the same mindset. Their fear of sounding or being made to feel stupid prevented them from making the best use of the chat groups that their eStudentMentors created to help them with their learning of the German language. For the remaining 68.75% (40 out of 64 students), their image does not appear to be the key factor for disengaging from the group chat. Only 10.34% (15 out of 145 students) were able to overcome their fear of ruining their image, and consequently, reap the benefits of using their chat groups.

Another possible cost of utilising the chat groups is students losing their knowledge power. According to Wang (2013), some people may withhold information if they believe it would benefit them more as opposed to sharing it. Yet only a very small number of students that participated in this study were reluctant to participate in their chat groups and preferred to keep information to themselves instead, due to the fear of losing their competitive advantage. The percentage of students who were reluctant to participate in their group chats due to their competitive mindset is surprisingly low in the context of a highly competitive Singapore. One of the chief explanations to this result might be the emphasis on collaborative learning. This resulted in the implementation of fixed learner teams from the beginning of the semester, which are the basis for all in-class group activities and team assignments. These teams consist of three students and are created on the principle of diversity in terms of gender and field of study. Some students also explained that they were more willing to help their peers when they were not graded on a Bell Curve when they could be certain that they would lose nothing by helping their peers.

According to Nahapiet and Ghoshal (1997), there are three dimensions to the Social Capital Theory (SCT), namely, structural, cognitive and relational (see Table 1).

According to SCT, in an ideal situation, an individual would perceive collective obligation, trust, and assimilate into the norm of the collective if he or she identified strongly with the collective (Lewicki & Bunker, 1996). Applied to this study, it would mean a strong relational dimension between an individual student and the student's entire chat group, hence the student having good chemistry with other classmates. Structurally, a student would be able to build good relationships and can influence classmates when

Table 1. Dimensions of Social Capital Theory.

	Structural	Cognitive	Relational
Focus	Individual's ability to build strong relationships with other individuals in a social network; how well an individual could influence the people around them.	Shared meaning or understanding within a group of people	Strength of the connection between the individual and the collective.

necessary. Cognitively, the student would share a common goal and understanding with the rest of the class. This cannot be validated by this study, as only about half the students felt that having chat groups served to nurture a sense of commitment by enabling them to interact, help one another, share important information, learn together as a class and build rapport outside of class. 48.97% (71 out of 145 students) thought their chat groups fostered group spirit; 50.35% (73 out of 145 students) thought their chat groups nurtured a sense of belonging; and 48.97% (71 out of 145 students) reported feeling a sense of loyalty towards their chat groups. 55.17% (80 out of 145 students) reported that they would feel a loss if their chat groups were no longer available. For better ease of analysis, 92 students had agreed with at least one of the abovementioned statistics. Out of the 92, 38 students did not use the group chat at all. They may have perceived that engagement would improve relations and class dynamics yet persisted in ignoring the group.

Personality traits, including openness to experience, extraversion, neuroticism, conscientiousness and agreeableness, may have been the cause to such perceptions (Matzler et al., 2008). Some students thought that they were not in a good position to help their peers due to self-doubts and insecurities, while other students were afraid of disturbing or annoying other students with their messages. Another group of students preferred to seek clarifications from their German lecturers directly. Others preferred face-to-face interactions and would rather discuss problems in person. Some students also prioritised other modules and commitments over the elective German language course. One student had difficulties expressing his or her question in words. A few students were independent, self-reliant learners who preferred to solve problems or source answers on their own rather than seek help from other students. Such reasonings could explain how students maintain a positive perception towards the chat group even if they did not use it at all. Indeed, some students pointed out that they were only more confident to share information with their peers when they were certain of the information that they were going to share, when the information was important, and when their peers requested for the information.

Many students were also intimidated by the size of the chat groups and felt shy, awkward and uncomfortable discussing problems with a larger group of students that included all students of their class, many of whom they were unfamiliar with. For these students, an easy workaround was to create separate groups with the members of their respective student teams within the class. These smaller chat groups will be referred to as Team Chats in this analysis, to avoid confusing them with the larger group chats created by the eStudentMentors.

92.41% (134 out of 145 students) mostly used Team Chats to organise meetings and prepare for team assignments. Students also reported feeling more comfortable discussing problems in these comfort zones that they created for themselves. 90.78% of the students that participated in this study asked more questions in their Team Chats than in their eStudentMentor facilitated chat groups. 88.27% (128 out of 145 students) asked and answered more questions in their Team Chats than in their chat groups, and 86.20% (125 out of 145 students) preferred using Team Chats to eStudentMentor facilitated chat groups to discuss problems. According to the strength-of-strong tie proposition of SCT, students were more likely to share information with the ones whom they had bonded with. Students felt closer to the small collaborative learning teams they were assigned to, because they interacted far more with their team members than with their classmates, as evident:

I'm closer to my team than my class as a whole. And the discussions on my team chat can be further discussed after class or when the opportunity arises.
I prefer Team Chat cos there's more familiarity and also we know each other's strengths and weaknesses better.
I prefer discussions in my Team Chat because, with fewer members, the pressure to reply is greater so I can be assured that someone will reply to my query faster.

As a result, they found it easier, less awkward and more comfortable to discuss questions with their team members than with the entire German class. Since there were only three people in the Team Chats, students were also less afraid of being judged for asking 'stupid questions'.

On the other hand, 41.37% (60 out of 145 students) preferred discussing problems in their chat groups instead because they were hosted by eStudentMentors. One student correlated students' willingness to participate in collaborative learning in their chat groups with having a friendly eStudentMentor to facilitate the discussions. According to the strength-of-weak-tie proposition of SCT (Lin, 2001), these students may have recognised that their Team Chat would be limited in knowledge since they are equal in level, thus driving them to look into the chat group where they have access to their eStudentMentor, one who has higher levels of knowledge. Additionally, eStudentMentors played a large role in the formation of pro-sharing norms in the chat groups by sharing exciting events that were related to the learning of German language and by replying promptly to students' messages. 84.92% (123 out of 145 students) agreed that their eStudentMentors encouraged information-sharing in the chat groups. The social environment in the chat groups was also healthy as 96.58% (140 out of 145 students) felt encouraged to respect other members while 95.21% (138 out of 145 students) felt open to conflicting views in the chat groups. Overall, students should have felt safe to engage in the group. Yet, it cannot be dismissed that some students were still concerned of being judged by their peers, as previously mentioned. This is highlighted by one team that explained they tried to reach a consensus in their Team Chat before seeking clarifications in their eStudentMentor-facilitated chat group. According to the rationality proposition of SET, students chose either Team Chats or chat groups based on that they believed had a higher probability of attaining the results that they wanted.

Conclusion

While the analysis shows the benefits of using chat groups facilitated by eStudentMentors to enhance collaborative learning, it also reveals the issues that prevented a notable number of students from making the best use of their chat groups to help with their language learning. Although the students recognised the value and benefits of the eStudentMentor-facilitated chats, a considerable number of students still chose not to partake in it due to their perceptions. A major factor was the overall relative inactivity of the chats, which makes the active role of the eStudentMentor a crucial element for success. Further studies would need to be conducted in order to determine whether certain engagement strategies would automatically result in more active group chats, as this study provided no evidence for specific desirable engagement strategies. A second main reason for the lack of active participation in the chats was identified as the lack of social bonds and unfamiliarity with their classmates. Further factors were fear of judgement and low commitment levels, which trickle down to additional problems, such as feeling awkward or having negative assumptions. These, in turn, hinder communication and the building of social bonds. These findings debunk SET propositions, in as much as the benefits of information sharing, though certainly greater, are diminished by a cost that should, in fact, be trivial. In the end, the seemingly small cost in the form of social perceptions and pressures accumulate and outweigh the rationally perceived benefits which students knew they could acquire.

References

Alonso, F., Manrique, D., Martinez, L., & Vines, J. M. (2011). How blended learning reduces underachievement in higher education: An experience in teaching computer sciences. *IEEE Transactions on Education, 54*(3), 471-478. https://doi.org/10.1109/TE.2010.2083665

Annamalai, N. (2018). How do we know what is happening in WhatsApp: A case study investigating pre-service teachers' online activity. *Malaysian Journal of Learning and Instruction, 15*(2), 207-225. https://doi.org/10.32890/mjli2018.15.2.8

Belias, D., Sdrolias, L., Kakkos, N., Koutiva, M., & Koustelios A. (2013). Traditional teaching methods vs. teaching through the application of information and communication technologies in the accounting field: Quo vadis? *European Scientific Journal, 9*(28), 73-101. https://doi.org/10.19044/esj.2013.v9n28p%25p

Bhatia, K., Takayesu, J. K., & Nadel, E. S. (2016). A novel mentorship programme for residents integrating academic development, clinical teaching and graduate medical education assessment. *Perspectives on Medical Education, 5*(1), 56–59. https://doi.org/10.1007/s40037-015-0236-2

Bonk, C. J., & Graham C. R. (2005). *The handbook of blended learning: Global perspectives, local designs.* Pfeiffer.

Castellanos, J., Gloria, A. M., Besson, D., & Harvey L. O. C. (2016). Mentoring matters: Racial ethnic minority undergraduates' cultural fit, mentorship, and college and life satisfaction. *Journal of College Reading and Learning, 46*(2), 81-98. https://doi.org/10.1080/10790195.2015.1121792

Cetinkaya, L. (2017). The impact of WhatsApp use on success in education process. *International Review of Research in Open and Distributed Learning, 18*(7), 59-74. https://doi.org/10.19173/irrodl.v18i7.3279

Common European Framework of Reference for Languages. (CEFR). (2023). *The CEFR levels*. The Council of Europe. https://coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Connolly, T., & Thorn, B. K. (1990). Discretionary databases: Theory, data, and implications. In J. Fulk & W. Steinfield (Eds.), *Organizations and communication* (pp. 219-233). Sage Publications.

Conroy, M. A. (2010). Internet tools for language learning: University students taking control of their writing. *Australasian Journal of Educational Technology, 26*(6), 861-882. https://doi.org/10.14742/ajet.1047

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-341. https://doi.org/10.2307/249008

Deci, E. L., & Ryan, R. M. (2010). Intrinsic motivation. In I. B. Weiner & W. E. Craighead (Eds), *The Corsini Encyclopedia of psychology*. https://doi.org/10.1002/9780470479216.corpsy0467

Deri, C. E. (2022). Social learning theory and academic writing in graduate studies. *Journal of Applied Learning & Teaching, 5*(1), 20-26. https://doi.org/10.37074/jalt.2022.5.s1.4

Dumford, A. D., & Miller A. L. (2018). Online learning in higher education: Exploring advantages and disadvantages for engagement. *Journal of Computing in Higher Education 30*, 452-465. https://doi.org/10.1007/s12528-018-9179-z

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature, 454*(7208), 1079-1083. https://doi.org/10.1038/nature07155

Gecer, A., & Dag, F. (2012). A blended learning experience. *Educational Sciences: Theory and Practice, 12*(1), 438-442.

Hardy, A., Ma, Y., Toettenborg, A., Mccaughan, L., Perkins, R., Andersen, L. L., & Ooi C. S. (2023). The role of online tourism education and its impact on student wellbeing during a 'COVID-pause. *Journal of Applied Learning & Teaching, 6*(1), 93-102. https://doi.org/10.37074/jalt.2023.6.1.16

Hockly, N. (2018). Blended learning. *ELT Journal, 72*(1), 97-101. https://doi.org/10.1093/elt/ccx058

Homans, G. C. (1974). *Social behaviour: Its elementary forms*. Jovanovich.

Hung, S. Y., Durcikova, A., Lai, H. M., & Lin W. M. (2011). The influence of intrinsic and extrinsic motivation on individuals' knowledge sharing behavior. *International Journal of Human-Computer Studies, 69*(6), 415-427. https://doi.org/10.1016/j.ijhcs.2011.02.004

Kamel, B. M. N., Giustini, D. M., & Wheeler, S. (2016). Instagram and WhatsApp in health and healthcare: An overview. *Future Internet, 8*(3), 37. https://doi.org/10.3390/fi8030037

Kasraie, N., & Kasraie E. (2010). Economies of e-learning in the 21st century. *Contemporary Issues in Education Research, 3*(10), 57-62. https://doi.org/10.19030/cier.v3i10.240

Kember, D., McNaught, C., Chong, F. C., Lam, P., & Cheng, K. F. (2010). Understanding the ways in which design features of educational websites impact upon student learning outcomes in blended learning environments. *Computers in Education, 55*(3), 1183–1192. https://doi.org/10.1016/j.compedu.2010.05.015

Kollock, P. (1999). The economies of online cooperation: Gifts and public goods in cyberspace. *Communities in Cyberspace,* 220-239.

Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations* (pp. 114-139). Sage Publications. https://doi.org/10.4135/9781452243610

Limone, P., Sinatra M., Tanucci, G., & Monacis, L. (2019). The utilitarian vs hedonic teacher acceptance of ICT use. *Turkish Online Journal of Distance Education, 20*(4), 1-10. https://doi.org/10.17718/tojde.640495

Lin, N. (2001). *Social capital: A theory of social structure and action*. Cambridge University Press.

Martinez-Caro, E., & Campuzano-Bolarin, F. (2011). Factors affecting students' satisfaction in engineering disciplines: Traditional vs. blended approaches. *European Journal of Engineering Education, 36*(5), 473–483. https://doi.org/10.1080/03043797.2011.619647

Matzler, K., Renzl, B., Müller, J., Herting, S., & Mooradian, T. A. (2008). Personality traits and knowledge sharing. *Journal of Economic Psychology, 29*, 301-313. https://doi.org/10.1016/j.joep.2007.06.004

McCarthy, J. (2010). Blended learning environments: Using social networking sites to enhance the first year experience. *Australasian Journal of Educational Technology, 26*(6), 729-740. https://doi.org/10.14742/ajet.1039

Molm, L. D. (1997). *Coercive power in social exchange*. Cambridge University Press. https://doi.org/10.1017/CBO9780511570919

Nahapiet, J., & Ghoshal, S. (1997). Social capital, intellectual capital and the creation of value in firms. *Academic of Management Proceedings, 1997*(1), 35-39. https://doi.org/10.5465/ambpp.1997.4980592

Nitza, D., & Roman, Y. (2016). WhatsApp messaging: Achievements and success in academia. *International Journal* *of Higher Education, 5*(4), 255-261. https://doi.org/10.5430/ijhe.v5n4p255

Poon, J. (2013). Blended learning: An institutional approach for enhancing students' learning experiences. *Journal of Online Learning and Teaching, 9*(2), 271-288.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon and Schuster.

Raiman, L., Antbring, R., & Mahmood, A. (2017). WhatsApp messenger as a tool to supplement medical education for medical students on clinical attachment. *BMC Medical Education, 17*(7), 7. https://doi.org/10.1186/s12909-017-0855-x

Rani, P., Tayal, D. K., & Bhatia, M. P. S. (2019). SNA using user experience. *2019 International Conference on Machine Learning, Bid Data, Cloud and Parallel Computing (COMITCon)*, 125-128. https://doi.org/10.1109/COMITCon.2019.8862174

Shih, R. (2011). Can web 2.0 technology assist college students in learning english writing? integrating facebook and peer assessment with blended learning. *Australasian Journal of Educational Technology, 27*(5), 829-845. https://doi.org/10.14742/ajet.934

Sirbu, C. C., Tonea, E., Iancu, T., Pet, E., & Popa, N. D. (2015). Aspects concerning the usage of modern methods for teaching – learning – evaluation in universities. *Procedia – Social and Behavioral Sciences, 182*, 550-554. https://doi.org/10.1016/j.sbspro.2015.04.777

Snowden, M., & Hardy, T. (2013). Peer mentorship and positive effects on student mentor and mentee retention and academic success. *Widening Participation and Lifelong Learning*, 76-92. https://doi.org/10.5456/WPLL.14.S.76

Statista. (2016). *Statistics and market data on mobile internet & apps*. http://www.Statista.com

Stein, J., & Graham, C. R. (2020). *Essentials for blended learning: A standards-based guide* (2nd ed.). Routledge.

Szadziewska, A., & Kujawski, J. (2017). Advantages and disadvantages of the blended-learning method used in the educational process at the faculty of management at the university of gdansk, in the opinion of undergraduate students. *Proceedings of ICERI2017 Conference*, 3938-3946. https://doi.org/10.21125/iceri.2017.1051

Tosun, S. (2015). The effects of blended learning on efl students' vocabulary enhancement. *Procedia—Social Behavioral Sciences, 199*, 641-647. https://doi.org/10.1016/j.sbspro.2015.07.592

Wang, D., & Chan, H. C. (2011). A theory of instinctive information sharing. *PACIS 2011 Proceedings*, 206. https://aisel.aisnet.org/pacis2011/206

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*(3), 271-294. https://doi.org/10.1111/j.1532-7078.2007.tb00227.x

Widen-Wulff, G. (2014). *The challenges of knowledge sharing in practice: A social approach.* Chandos Publishing Limited.

Widyasari, R., Humaira, M. A., Tambunan, A. R. S., Rachmadtullah, R., & Samsudin, A. (2020). Using blended learning approach (BLA) in inclusive education course: A study investigating teacher students' perception. *International Journal of Emerging Technology in Learning (iJET), 15*(2), 72-85. https://doi.org/10.3991/ijet.v15i02.9285

Appendix A

Segment	Theory	Questions
1	Not Applicable General Group Chat Behaviour	1.1 I have asked questions in my WhatsApp/Telegram group 1.2 I have answered questions in my WhatsApp/Telegram Group 1.3 How often do you share questions in your WhatsApp/Telegram Group? 1.4 Why have you been active/ inactive in your WhatsApp/Telegram group?
2	Intrinsic Benefit in Social Exchange Theory	2.1 I enjoy sharing information with others in our WhatsApp/Telegram group 2.2 It feels good to help someone else by sharing my knowledge in our WhatsApp/Telegram group 2.3 I enjoy helping others when they ask questions in our WhatsApp/Telegram group. 2.4 Why do you enjoy/ dislike helping others in your WhatsApp/Telegram group
3	Theory of Instinctive Information Sharing	3.1 It is natural for me to share information with others. 3.2 When I learn something new, I wish to share it with others. 3.3 I prefer sharing information with people rather than keeping it to myself. 3.4 In which way does it come natural/ unnatural for you to share information with others?
4	Social Exchange Theory	4.1 When I ask questions in my WhatsApp/Telegram group, I expect my peer learners to discuss my questions 4.2 When I ask questions in my WhatsApp/Telegram group, I expect my peer learners to answer my questions 4.3 When I ask questions in my WhatsApp/Telegram group, I expect my mentor to answer my questions 4.4 Why do you (not) expect help from others in your WhatsApp/Telegram group?
5	Cost in Social Exchange Theory	5.1 Sharing information in our WhatsApp/Telegram group makes me lose my knowledge that makes me stand out with respect to others. 5.2 Sharing information in our WhatsApp/Telegram group makes me lose my knowledge that no one else has. 5.3 Sharing information in our WhatsApp/Telegram group makes me lose my unique value. 5.4 Why does sharing information in your WhatsApp/Telegram group help you get an advantage/ make you lose your advantage in terms of knowledge?
6	Extrinsic Benefit in Social Exchange Theory	6.1 Students who share our knowledge in our WhatsApp/Telegram group have more prestige than those who do not. 6.2 Sharing information in our WhatsApp/Telegram group improves others' respect of me. 6.3 Asking questions in our WhatsApp/Telegram group makes me look stupid. 6.4 Answering questions in our WhatsApp/Telegram group makes me look smart.
		6.5 In which way does sharing information in your WhatsApp/Telegram group improve your image/ make others lose respect of you?
7	The Ego and Intrinsic Motivation in Social Exchange Theory	7.1 Sharing information in our WhatsApp/Telegram group makes me feel important. 7.2 Sharing information in our WhatsApp/Telegram group increases my self-esteem. 7.3 Sharing information in our WhatsApp/Telegram group makes me feel needed. 7.4 In which way does sharing information in your WhatsApp/Telegram group make you feel better/ worse about yourself?
8	Extrinsic Motivation in Social Exchange Theory	8.1 Active participation in the information sharing benefits my learning. 8.2 Asking questions in our WhatsApp/Telegram group helps my learning. 8.3 Answering questions in our WhatsApp/Telegram group helps my learning. 8.4 In which way does sharing information in your WhatsApp/Telegram group help your learning?
9	Relational Dimension in Social Capital Theory; Pro-sharing Norms	9.1 Our eStudentMentor encourages us to ask questions in our WhatsApp/Telegram group. 9.2 Our eStudentMentor encourages us to post answers in our WhatsApp/Telegram group. 9.3 Our eStudentMentor encourages discussion in our WhatsApp/Telegram group. 9.4 How does your eStudentMentor encourage you to share information in your WhatsApp/Telegram group?

10	Structural Dimension in Social Capital Theory	10.1 Members are encouraged to respect other members in our WhatsApp/Telegram group. 10.2 Members are encouraged to be open to conflicting views in our WhatsApp/Telegram group. 10.3 In which way are members encouraged to respect others and their views in your WhatsApp/Telegram group?
11	Construct of Commitment and Cognitive Dimension in Social Capital Theory	11.1 I would feel a loss if our WhatsApp/Telegram group was no longer available. 11.2 I feel a sense of loyalty towards our WhatsApp/Telegram group. 11.3 Our WhatsApp/Telegram group created a group spirit. 11.4 Our WhatsApp/Telegram group nurtured a sense of belonging. 11.5 How has your WhatsApp/Telegram group nurtured/ failed to nurture a sense of commitment?
12	Propositions of Social Capital Theory	12.1 Do you have a WhatsApp/Telegram chat group with your Oral-Team members (Team Chat)? 12.2 I ask more questions in my Team Chat than in my WhatsApp/Telegram group. 12.3 I answer more questions in my Team Chat than in my WhatsApp/Telegram group. 12.4 I prefer using the Team Chat for discussion of problems or questions.
		12.5 I mostly use my Team Chat for organisational matters (i.e. when and where to meet, etc.) 12.6 I prefer discussions in the WhatsApp/Telegram group, because it has more members and generates more responses and ideas. 12.7 I prefer discussions in the WhatsApp/Telegram group, because it is hosted by an eStudentMentor. 12.8 Why did you prefer discussions in the WhatsApp/Telegram group/ Team Chat?

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : **2591-801X**

Content Available at : **http://journals.sfu.ca/jalt/index.php/jalt/index**

Metamorphosis of a teacher educator: A journey towards a more critical self

Jaber Kamali[A]

[A] *Ibn Haldun University Teacher Training Center, Istanbul, Turkey*

Keywords

Autoethnography;
critical pedagogy;
critical teacher education;
narrative analysis.

Abstract

Critical teacher education emerged as a response to the liberal, hegemonic, and power-oriented world that affected teacher education as well. Albeit widely discussed, moving towards becoming this type of teacher educator is neither easy nor fast. This autoethnographic narrative study describes my journey as a teacher educator from a non-critical, product-oriented, passive teacher educator to a more critical, process-oriented, active teacher educator who learns, questions, relearns, and unlearns. The data are gathered from different sources of my personal portfolio, including training diaries, field notes, memories, feedback, and observation. The findings of the study reveal the underlying factors that shape our thoughts, beliefs, and practices and how we can gain voice and agency and transform into critical teacher educators.

Correspondence

jaber.kamali@ihu.edu.tr [A]

Article Info

Received 2 June 2023
Received in revised form 29 June 2023
Accepted 12 July 2023
Available online 18 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.8

Introduction

The influence of power and hegemony are traceable in education, where the theories are mostly created in the West and are practised worldwide. The field of education is filled with political consideration (Akbari, 2008) in choosing what, how, and whom to teach, and this influences the educational life of us. It becomes more critical when it comes to teacher education as teacher educators play a crucial role in the teachers' preparation and professional development (Richards & Schmidt, 2010). As a response to this power-oriented educational context, some movements came into existence, which we can call critical pedagogy (Freire, 1979). Critical pedagogy posits that the traditional form of education, which oftentimes reinforces current power principles and maintains social inequality, should be replaced by a more liberatory form that is able to empower students by hearing their voices and promoting social change. This concept has gained momentum in education hitherto and entered the field of applied linguistics by Pennycook (2001) when he relates the "micro relations of applied linguistics to macro relations of society" (p. 2). Critical teacher education, as the offspring of critical pedagogy, gained momentum by "loss of equity, economic and social justice and the polarisation of the labour force" (Hill, 2007, p. 210) as a result of the liberal and neoliberal impact on education. The literature on critical teacher education is rich, encompassing a wide range of concepts such as social justice teacher education (McDonald, 2008), teacher agency (Tao & Gao, 2021), reflective teacher observation (Javahery & Kamali, 2023), reflective practice (Farrell, 2019), and critical reflection (Bassot, 2023), to name a few. These concepts assist the researchers in exploring various aspects of criticality in teacher education, and contributions to this concept are multimodal.

Hawkins and Norton, (2009), reviewing the literature on critical teacher education, attributed some principles to this type of teacher education. First, it is context-specific. They state that "teacher educators drew on their cultural and historical knowledge of the context and the students in order to work innovatively with teacher-learners" (p. 36). The second principle is being responsive to learners. In this principle, "language teacher educators took into account their knowledge of their teacher learners' languages, cultures, desires and histories, and connected learning to the backgrounds and experiences students brought to the learning environment" (p. 36). The third principle which is dialogic engagement deals with respecting learners voices by stating that "language teacher educators used collaborative dialogue to construct and mediate meanings and understandings… to promote reflection among participants, and to link explicit critical awareness of social justice issues to educational practices" (p. 36). Reflexivity is the next principle which means an "insightful analysis of what occurred, and how they might use what they have learned … to re-design future possibilities" (p. 36). The last principle, i.e. praxis, is "… integrating theory and practice in the interests of educational and social change" (p. 36).

One of the qualitative methods to explore critical teacher education is autoethnography. Autoethnography, as a form of ethnography, was widely discussed by Denzin and Lincoln (2005). They made the first attempts to apply the concept to qualitative research. Later, Canagarajah (2012) defined autoethnography by dividing it into its components, namely *auto, ethno*, and *graphy*. He defined *auto* as a type of research "conducted and represented from the point of view of the self, whether studying one's own experiences or those of one's community" (p. 260). *Ethno*, in his thoughts, highlights how culture and society (re)shape in connection with personals. Lastly, *graphy* as a form of "writing is not only the means of disseminating one's knowledge and experiences; there is an emphasis on the creative resources of writing, especially narrative, for generating, recording, and analyzing data" (p. 260). Different scholars have employed and approved the advantages of autoethnography to examine an emic insider view in teacher education (e.g., López-Gopar, et al., 2022; Weng & Troyan, 2023; Yazan, 2019; Yazan et al., 2023). This approach became prominent on the ground that I was gathering data on myself where I employed a meta-awareness of my actions which not only helped me conduct research but also influenced and informed my actions. Furthermore, as reflection can help educators to have a better understanding of their environment, their interactants, and themselves, this study uses a self-reflective autoethnographic approach to examine a change that has occurred in the author during his journey since 2006 as a teacher educator and how a shift towards a more critical teacher educator has emerged.

Teacher metamorphosis (Kamali, 2014, 2021) was characterized as a type of shift from a one-role approach to a teacher to a package of other roles such as assessor, material developer, iconoclast, and the like. It was introduced as a remedy to some criticisms of post-method pedagogy (Kumaravadivelu, 2001). Kumaravadivelu (2006) attributed three principles to post-method, which seem to be the cornerstones for critical pedagogy where teaching is possible, particular, and practical. Therefore, in this post-method era, teachers take a more participatory approach to teaching and, by giving voice and agency to students, bring their life into the classroom. This is the teachers' duty, then, to design specific content which is context-specific.

This study, in the same vein, portrays the metamorphosis of myself as a teacher educator to explore the underlying agenda of this journey. To depict the shift in myself from a passive recipient of knowledge to a critical teacher educator, I provide narratives from my reflective notes and diaries, field notes, memories, colleagues' and trainees' feedback, and observation that I have collected in all these years. Therefore, the research question this study tries to provide an answer to is as follows: How did I transform into a more critical teacher educator?

The data gathered from 325 pages of electronic and handwritten reflective notes (from my colleagues and trainees); 145 pages of my own diaries and reflective notes (handwritten and electronic); trainees' feedback on 63 training classes (handwritten and electronic through email and Google forms); and almost 80 hours of classroom observations (both teachers and trainers), out of which almost six hours were transcribed. These have been gathered since 2007.

In order to analyze the data, deductive thematic analysis (Braun & Clarke, 2006) was employed since the data were codified according to the predetermined mode of critical teacher education (Hawkins & Norton, 2009). The emerged sub-themes from the data were categorized into the critical teacher education themes: context-specificity, responsiveness to learners, dialogic engagement, reflexivity, and praxis. Then, the most critical subthemes, which I believed made a significant impact on each theme, were selected and analyzed. For example, in the first theme, i.e. context-specificity, there were different sub-themes gathered from my diary, classroom observation, and reflective notes. The one that I considered the most influential in my journey towards my critical self, however, was my experience with the reading-aloud technique in reading classes, for it was the most remarkable thought-provoking encounter with a concept that I discovered was not always applicable in my own context.

The transformation journey

From a context-negligent to a context-aware teacher educator

My first training diary in June 2009 read as follows:

> Today, I had a great time teaching the trainees how to teach reading. They were very cooperative and learned a lot. Mohadeseh told me that she didn't know that reading aloud is wrong in the class. She said she used it in her classes, and since the students liked it, she thought it could be a good idea. I also taught them the steps in teaching reading as … and I think now they can teach reading well.

As the diary shows, I adopted a one-size-fits-all approach in which I dictated the steps of teaching reading, not leaving any room for alternative lesson shapes. In effect, the sheer example of the prescriptive approach is traceable in the prohibition of the reading-aloud technique in the class. There was a time that reading a text aloud was assumed to make boredom for the students and since it is for the purpose of reading comprehension, not reading for pronunciation, I presumed it was wrongdoing in the class. Reading Ur's (2016) book "100 Teaching Tips", I realized there were numerous benefits for this technique in language classes. As Ur (2016) pointed out, by reading aloud, "you intuitively use appropriate prosody: group the words into sense-patterns, insert pauses in the right places, and add appropriate intonation. This is what clarifies meaning" (p. 161). It is also "easy to stop every now and again to explain new words as you feel necessary. And you can keep an eye on your students, pick up any expressions of incomprehension and respond as necessary" (Ur, 2016, p. 61).

Therefore, training in one of the in-house training sessions for an institution in Iran in September 2018, I designed and delivered a session on reading-aloud techniques and their benefits. I asked the teachers to apply these techniques in their classrooms and complete a reflective note about this based on their own experience. One of the trainees completed the reflective note as follows:

> When I was reading the text to the students, I saw some of them smiling, and at the end, they told me that they loved my accent and they wanted to speak like me. I think reading aloud could make an emotional bond in us, and I could make them more motivated to learn English.

With the advent of movements against linguistic imperialism (Mackenzie, 2022; Philipson, 1992) more awareness was raised towards more critical concepts such as English as a lingua franca (ELF) (Jenkins, 2006, 2007) or world Englishes (Kachru, 1992) in which accents are considered a sign of marginalization by which native speakers of English gain privilege. Nonetheless, this reflective note clearly revealed a cultural peculiarity. Unlike the ELF movement, you may hear that students feel more emotionally connected to teachers with more native-like accents because they see their dreams in assimilation into the Western society where there are more resources, facilities, and quality of life. This is evident in some countries with more social or political challenges; however, in some countries in which patriotism is practiced and valued, you may never hear this sentence (Westheimer, 2006). This is not to confirm or appreciate native accents but to affirm the emotional bonds which can be created differently in different cultures.

What I learned from the experience is that believing and preaching what you have learned in one book or heard somewhere and extending it to all contexts can be counter-productive. This extract showed that the technique opened a new door for the students, without which it would never happen. Context-specificity, then, can act as a source of motivation and inspiration that a top-down, prescriptive form of teacher education can never do.

From a fixed-plan to a trainee-responsive teacher educator
Until recently, I have tried to have well-thought, fixed lesson plans for my training sessions which were strictly followed during the session, and deviations from them were sins that should never be committed. It is in line with the literature on reflection to be prepared for a session (reflection-for-action) (Farrell, 2018); however, what I have neglected was that most of this literature discusses plan B as long as we deal with the unpredictability of life and human being as our interlocutors. It was in December of 2017 when I was responsible to train a group of teachers at the Ministry of Education to teach at state schools. Since they were assigned to teach young learners, I designed a session on chants and songs in which I prepared numerous resources they could use in their classes to feel students more motivated and teach them English sounds and rhymes. However, at the beginning of the session, one of my trainees pointed out that they were not allowed to use music (especially English) in their classes due to parents' and schools' resistance to these songs. She said that she had already tested it in her internship, and some of the pupils' parents warned her if she did it again, they would not let their children come to the school. Then, I was in a dilemma of continuing my prepared lesson plan, which I spent hours preparing or shifting to what they wanted to work on, the plan which was applicable to their immediate context. I decided to continue with my fixed lesson plan; however, that session was one of the least

favourite sessions of the course, as the end-of-the-course survey indicated. The reason was explained by one of the trainees in the comment section of the survey.

> I know that our teacher [trainer] did his best to tell us how to integrate songs in our sessions, but it was not helpful since it is not in our hands to use them here. We just follow the rules because if we don't, we'll be fired. So, what is the use of it when we can't use it? I know songs can make students motivated but I can't use them. I really preferred a session about how to deal with students' parents instead of that session. I want to know how to deal with their orders and how I can convince them about something I do in the class and I know it is true.

The extract clearly evidences Kumaravadivelu's (1994) parameter of particularity by which he argued that language pedagogy should be relevant to a group of learners. The influential role of context in language learning is acknowledged by different scholars (e.g., Barkhuizen, 2008; Moranski & Zalbidea, 2022) who claim that context is an inseparable part of any language learning milieu without which language is meaningless, nonsense, and hard to remember. This experience has also added to the existing literature on the use of context in teacher education (e.g., Adonious, 2013; Bax, 1997; Zhao, 2022) by asserting that context is even more critical in teacher education as it might make teacher education inapplicable to teaching. This experience suggested the significance of the role that the trainees and their context can play in the content of the session. From then on, I attempted to analyze the trainees of a program and their social, cultural, and educational backgrounds to avoid any similar experiences.

From an authoritative to a dialogic teacher educator

In conceptualizing anti-Machiavellian teacher education as a type of critical teacher education, I argued that Machiavellian teacher education is one that prefers fear to love (Kamali, 2022). In this education, it is beneficial for both groups if trainees are afraid of the trainer. Unlike that, an anti-Machiavellian approach prioritizes love and posits that this phenomenon can inspire and develop teachers professionally. Sources of teachers' beliefs are abundant, one of which is our learning experience which has a profound impact on how we teach and train (Karaca & Uysal, 2023). In the first years of my training career, being surrounded by my learning experience in a formal teaching context in Iran, where classes were places for practising teachers' authority, I was excessively obsessed with the belief that authority (of which people are afraid) can earn me respect, dominance, and dignity. However, I was proved wrong by co-training with one of my colleagues between the years 2017 and 2019. Seeing the trainees' evaluation surveys at the end of the course, I was always thrilled by how popular he was. Therefore, I decided to observe him. He was generous enough to accept, and I observed a series of his sessions to find out the reasons for his popularity. His training sessions were fun and gamified in which the trainees had voices and could contribute freely, criticize sharply, and disagree fully. It was odd for me because I considered it a sign of failure when a trainee disagreed with me. For him, however, it was a moment he could reflect and convince others to do. The following extract is from one of his sessions in June 2018 on teaching through text, where he demonstrated a mini-lesson on the grammar of present simple and present continuous using the PPP model (Presentation, Practice, Production) (see Anderson, 2017). He started by talking about his daily routine and ended with a discussion about a favourite place. In the evaluation part of the demo, one of his trainees disagreed with it, saying that he did not follow the same context in his demo.

> Trainer: Ok, now, tell me how you found the session.
>
> Trainee 1: It was really good, but I think you didn't keep the context in the session. In the preparation phase (referring to the PPP model), you discussed the daily routine, and in production, it was changed to a favorite place.
>
> Trainer: Great, very good point. Do you agree with her (addressing the class)?
>
> Trainee 2: Yes, but I think sometimes it is good to change the context to add more variety to the class. You know, such as a fresh start.
>
> Trainer: I do agree. So, now we have two opinions. 1. We have to keep the context 2. It's better to change it. Go back into your groups and discuss the advantages and disadvantages of each.

The extract evidently showed that not only the trainer was not defensive and did not try to justify his actions, but he also seized the opportunity to promote reflection in the classroom. This is in line with magic moments introduced by Harmer (2017), who argues that these are off-plan moments in the classroom. Having observed his sessions and the way he gave voice and agency to his trainees and exploited the magic moments of the sessions, I decided to run more dialogic sessions where I welcomed contributions and valued opinions. The implication of this observation for my training sessions was that I added one part to all my tasks for teacher education with the name "my suggestions". Adding this column, I ask the trainees to openly discuss their opinions and add to the existing body of knowledge in that field. This is how I can promote dialogue and interaction in the training sessions.

From a passive to a reflective teacher educator

Active learning is defined as peoples' "initiatives and responsibilities for their own progress" (Niemi, 2002, p. 763), which can be synonymous with agency. On the contrary, and based on this definition, passive learning can be characterized as learning in which the learner does not have any responsibility except for being the recipient of knowledge. As a novice teacher educator, my sources of knowledge were the books I read about teaching, the sessions I observed, and performing them as they were without reflecting upon their benefits and demerits for my context, my trainees, and society. I have de facto heard about reflection all the time; however, I did not have it in my own training toolkit. I really cannot name a single moment as a turning point in this journey that suddenly changed me from a recipient to a reflective educator; however, some elements were remarkable. One of these elements was the love of writing. I cannot recall the exact time it flourished in me; nevertheless, I have been always mesmerized by Persian literature and wrote some poems, albeit not professionally. Writing journal diaries about my teaching and training helped me to reflect and re-reflect on what I had done in my classes. One of the clearest reflective moments for me occurred in one of my journal entries when I watched my own sessions of training in January 2014.

> Today, a trainee asked me about "withitness" and wanted to know how she could apply it in her classes. OH MY GOODNESS. I hadn't heard that before. It was embarrassing. At first, I tried to think and made up a response based on my discretion, but it was unsuccessful. Although then I honestly said "I don't know what it is, I search and let you know", I think I will lose their trust in me. I don't feel well now.

As the diary reads, I noticed a gap in my knowledge. Writing about this moment in my class filled me with a desire to find a way to solve it. It was then I asked some of my colleagues and surfed the net. Searching the net, I came across an article by Alan Waters (1998) about monkey management that argued that problems should be solved at the lowest organizational level. Applying it to English Language Teaching (ELT), he provided clear examples of answering students' questions by proposing the question to the class before providing the answer to it. By doing so, students will become responsible for their own learning. Besides, they do not lose their trust and confidence in you as a teacher. I applied this to teacher education and could add it to my toolkit as a teacher educator. Now, if trainees ask me a question I have no answer to, I am more confident and have more strategies to use since, at first, I pose the question to the class and ask them to find the answer to their questions, then I provide one form of teachers' scaffolding (Tajeddin & Kamali, 2020) for them.

Reflection in the form of journal writing not only could provide me with the answer to my questions but also could persuade me to include it in my professional life because it can enhance professional development and establish a culture of thinking and reflecting. I have found the answer to the question I was asked (withitness); however, the learning that occurred from this experience was deeper than that specific word; that is, I learned how to answer my students' questions and how to reflect on the learning moments of my classroom

From a theory-oriented to a practice-oriented teacher educator

Bridging the gap between theory and practice has always been the ultimate goal of theoreticians and practitioners. Although everyone in the field of education claims to do so, there is a big gap between these two agendas. I can remember my first training course in an institution in Iran I was employed as a director of studies after three years of teaching in 2007. One of my main responsibilities was to run teacher training courses for the teachers of that institution. Not taking any formal training and not having enough experience in teacher training, I read the book "learning teaching" (Scrivener, 2010) and taught it in a very trainer-fronted environment. I was not informed about the wrong approach I took at that time because Iran is a high-power distance country (Hofstede et al., 2010), in which people do not challenge their authority. Based on Maslow's hierarchy of competence (Franz et al., 2018), one can only move to conscious competence if s/he becomes aware of his/her incompetence. Not being challenged and aware of my incompetence, I continued the way I was training for a while. As I have already mentioned, although I cannot call a point in my professional life when a radical change occurred in my approach to teacher training, an incident had a dramatic impact on my view on training sessions. I applied for a teacher training course in 2016, where I got familiar with the concept of process and content (McGrath, 1997) of training sessions. I learned that due to the nature of the teaching and training career, which is performance-based, trainees should not only learn what to teach but how to teach. Being interested in the concept, I read some literature regarding it (e.g., DelliCarpini, 2009; Woodward, 2003) and found out that teachers need to see how to teach rather than being told. One of the conclusions of McGrath's (1997) paper was a gist of what I have learned:

> If, in training (trainers), we use only those categories of process or process options with which participants are already familiar, we cannot expect them to use other processes in their own teaching. We may even dull their interest in their own learning (p. 172).

As the quote suggests, using different processes in transferring the content can add to teachers' teaching repertoire by which they can run more interesting, engaging, learner-friendly sessions; otherwise, teachers may acquire explicit knowledge about the language which cannot be transferred well to their learners.

Conclusion

The present study depicted my transformation as a teacher educator towards a more active role in my society, the one which reinforced more criticality in me where I asked more questions, reflected more frequently, and provided more opportunities for student teachers to have voices and agency, and criticized more constructively (Figure 1). To be a more critical self, I have suffered pain and experienced failures. However, the key to success in this journey was and is consistency.

My journey illustrates my metamorphosis as a teacher educator. This is significant on the ground that teacher educators are considered powerful agents who dictate the methods, approaches, and techniques to teachers. This is the raison d'être for this group's attitudes, beliefs, and ideologies to be under-researched. Nonetheless, this group is also influenced by social, cultural, political, and ideological

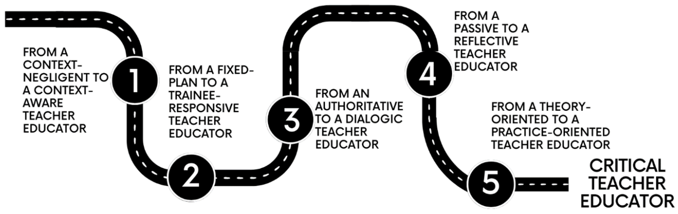


Figure 1. My journey towards a more critical teacher educator.

macro policies which shape and reshape their competence and performance (Hallett, 2010; Maaranen et al., 2019). This is noteworthy that the change is continuous and cannot be attributed to a specific moment; however, the moments that inspire, revolutionize, and challenge us are not few in this journey. It is our duty, then, to embrace them, reflect upon them, and accept them to bring about the change in us and therefore, move towards a better society.

Self-reflective autoethnographies, albeit being recognized as critical tools for teacher education (e.g., Canagarajah, 2012; Yazan, 2019) are not highlighted in teacher educators' learning per se. Therefore, the results of this study aid teacher educators to apply autobiographical self-reflection to their own context by collecting information about the critical moments in their classes, reflecting upon those events and drawing their own map of becoming a more critical teacher educator. This is also applicable for teachers to evaluate their own development and transformation to a more critical self. These reflective maps provide valuable schematic realizations of the teachers' and teacher educators' journey, when, why, and how the changes happen during this journey, and where they are heading in their journeys.
It should be also noted that "narratives are shaped by and imply an analysis of experience" (Canagarajah, 2012, p. 261) which can make it very personal and hard to generalize. Nevertheless, this limitation can turn into a benefit as it can encourage other researchers, teachers, and teacher educators to conduct it on themselves, their trainees, or their students in their own context and map their own experiential, reflective journey.

References

Adonious, M. (2013). Preparing teachers: The importance of connecting contexts in teacher education. *Australian Journal of Teacher Education, 38*(8), 47-60. https://doi.org/10.14221/ajte.2013v38n8.7

Akbari, R. (2008). Postmethod discourse and practice. *TESOL Quarterly, 42*(4), 641-652. https://doi.org/10.1002/j.1545-7249.2008.tb00152.x

Anderson, J. (2017). A potted history of PPP with the help of ELT Journal. *ELT Journal, 71*(2), 218-227. https://doi.org/10.1093/elt/ccw055

Baratta, A. (2018). "I speak how I speak:" A discussion of accent and identity within teachers of ELT. In B. Yazan and N. Rudolph (Eds.), Criticality, teacher identity, and (in)equity in English language teaching. *Educational Linguistics, 35,* 163–178. Springer. https://doi.org/10.1007/978-3-319-72920-6_9

Barkhuizen, G. (2008). A narrative approach to exploring context in language teaching. *ELT Journal, 62*(3), 231-239. https://doi.org/10.1093/elt/ccm043

Bassot, B. (2023). *The reflective practice guide: An interdisciplinary approach to critical reflection.* Taylor & Francis.

Bax, S. (1997). Roles for a teacher educator in context-sensitive teacher education. *ELT Journal, 51*(3), 232-241. https://doi.org/10.1093/elt/51.3.232

Boonsuk, Y., Wasoh, F., & Waelateh, B. (2023). Whose English should be talked and taught? Views from international English teachers in Thai higher education. *Language Teaching Research, 0*(0). https://doi.org/10.1177/13621688231152194

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101. http://dx.doi.org/10.1191/1478088706qp063oa

Canagarajah, A. S. (2012). Teacher development in a global profession: An autoethnography. *TESOL Quarterly, 46*(2), 258-279. https://doi.org/10.1002/tesq.18

DelliCarpini, M. (2009). Enhancing cooperative learning in TESOL teacher education. *ELT Journal, 63*(1), 42-50. https://doi.org/10.1093/elt/ccn016

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research.* Sage.

Farrell, T. S. (2018). *Reflection-"as"-Action in ELT. ELT Development Series.* TESOL Press.

Farrell, T. S. (2019). *Reflective practice in ELT.* London, UK: Equinox.

Franz, HW., Kaletka, C., Pelka, B., Sarcina, R. (2018). Growing experience: From unconscious incompetence to unconscious competence. In H. Franz, C. Kaletka , B. Pelka and R. Sarcina (Eds.) *Building leadership in project and network management: A facilitator's toolset,* (pp. 209-226). Springer.

Freire, P. (1979). *Pedagogy of the oppressed.* Bloomsbury Publishing.

Hallett, F. (2010). Do we practice what we preach? An examination of the pedagogical beliefs of teacher educators. *Teaching in Higher Education, 15*(4), 435-448. https://doi.org/10.1080/13562517.2010.493347

Harmer, J. (2017). *Essential teacher knowledge*. Pearson.

Hawkins, M., & Norton, B. (2009). Critical language teacher education. In A. Burns & J.C. Richards (Eds.), *Cambridge guide to second language teacher education*, (pp. 30-39). Cambridge University Press.

Hill, D. (2007). Critical teacher education, new labour, and the global project of neoliberal capital. *Policy Futures in Education, 5*(2), 204-225. https://doi.org/10.2304/pfie.2007.5.2.204

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures et organisations: Nos programmations mentales*. Pearson Education France.

Javahery, P., & Kamali, J. (2023). Teachers' personality types and their attitude toward receiving and employing postobservation feedback. *Psychology in the Schools*. https://doi.org/10.1002/pits.22900

Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly, 40*(1), 157-181. https://doi.org/10.2307/40264515

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford University Press.

Kachru, B. B. (1992). World Englishes: Approaches, issues and resources. *Language teaching, 25*(1), 1-14. https://doi.org/10.1017/S0261444800006583

Kamali, J. (2014). Post method survival. *Procedia-Social and Behavioral Sciences, 98*, 824-829. https://doi.org/10.1016/j.sbspro.2014.03.487

Kamali, J. (2021). Teachers' beliefs about teacher metamorphosis. *MEXTESOL Journal, 45*(1), 1-8. https://www.mextesol.net/journal/index.php?page=journal&id_article=23348

Kamali, J. (2022). Conceptualizing an anti-Machiavellian approach in second language teacher education. *The Teacher Trainer Journal. 36*(1), 35-37. https://edition.pilgrimsttj.com/html5/reader/production/default.aspx?pnum=35&edid=047b880c-9431-4bea-a017-dec3241e3021&isshared=true

Karaca, M., & Uysal, H. H. (2023). An investigation into L2 writing teacher beliefs and their possible sources. *Assessing Writing, 56*, 100710. https://doi.org/10.1016/j.asw.2023.100710

Kumaravadivelu, B. (1994). The postmethod condition: (E)merging strategies for second/foreign language teaching. *TESOL Quarterly, 28*(1), 27-48. https://doi.org/10.2307/3587197

Kumaravadivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly, 35*(4), 537-560. https://doi.org/10.2307/3588427

Kumaravadivelu, B. (2006). TESOL methods: Changing tracks, challenging trends. *TESOL Quarterly, 40*(1), 59-81. https://doi.org/10.2307/40264511

López-Gopar, M. E., Sughrua, W. M., & Huerta Cordova, V. (2022). The journey of a critical-oriented ELT curriculum and the identities of teacher educators: a collaborative and analytic autoethnography. *Teachers and Teaching,* 1-14. https://doi.org/10.1080/13540602.2022.2062733

Maaranen, K., Kynäslahti, H., Byman, R., Jyrhämä, R., & Sintonen, S. (2019). Teacher education matters: Finnish teacher educators' concerns, beliefs, and values. *European Journal of Teacher Education, 42*(2), 211-227. https://doi.org/10.1080/02619768.2019.1566317

Mackenzie, L. (2022). Linguistic imperialism, English, and development: implications for Colombia. *Current Issues in Language Planning, 23*(2), 137-156. https://doi.org/10.1080/14664208.2021.1939977

McDonald, M. A. (2008). The pedagogy of assignments in social justice teacher education. *Equity & Excellence in Education, 41*(2), 151-167. https://doi.org/10.1080/10665680801943949

McGrath, I. (1997). *Feeding, leading, showing, throwing: Process choices in teacher training and trainer training. Learning to train: Perspectives on the development of language teacher trainers* (pp. 163-172). Hemel.

Moranski, K., & Zalbidea, J. (2022). Context and generalizability in multisite L2 classroom research: The impact of deductive versus guided inductive instruction. *Language Learning, 72*(S1), 41-82. https://doi.org/10.1111/lang.12487

Niemi, H. (2002). Active learning—a cultural change needed in teacher education and schools. *Teaching and teacher education, 18*(7), 763-780. https://doi.org/10.1016/S0742-051X(02)00042-2

Pennycook, A. (2021). *Critical applied linguistics: A critical re-introduction*. Routledge.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford University Press.

Pole, C., & Morrison, M. (2003). *Ethnography for education*. McGraw-Hill Education (UK).

Richards, J. C., & Schmidt, R. W. (2010). *Longman dictionary of language teaching and applied linguistics*. Routledge.

Scrivener, J. (2005). *Learning teaching*. Macmillan.

Tajeddin, Z., & Kamali, J. (2020). Typology of scaffolding in teacher discourse: Large data-based evidence from second language classrooms. *International Journal of Applied Linguistics, 30*(2), 329-343. https://doi.org/10.1111/ijal.12286

Tajeddin, Z., & Pakzadian, M. (2020). Representation of inner, outer and expanding circle varieties and cultures in global ELT textbooks. *Asian-Pacific Journal of Second and Foreign Language Education, 5*(1), 1-15. https://doi.org/10.1186/s40862-020-00089-9

Tao, J., & Gao, X. A. (2021). *Language teacher agency*. Cambridge University Press.

Ur, P. (2016). *Penny Ur's 100 Teaching Tips: Cambridge* *handbooks for language teachers*. Cambridge University Press.

Waters, A. (1998). Managing monkeys in the ELT classroom. *ELT Journal, 52*(1), 11-18. https://doi.org/10.1093/elt/52.1.11

Weng, Z., & Troyan, F. J. (2023). Developing language teacher identity through photo-elicitation in TESOL teacher education. *TESOL Journal,* e714. https://doi.org/10.1002/tesj.714

Westheimer, J. (2006). Politics and patriotism in education. *Phi Delta Kappan, 87*(8), 608–620. https://doi.org/10.1177/003172170608700817

Woodward, T. (2003). Loop input. *ELT Journal, 57*(3), 301-304. https://doi.org/10.1093/elt/57.3.301

Yazan, B. (2019). Toward identity-oriented teacher education: Critical autoethnographic narrative. *TESOL Journal, 10*(1), e00388. https://doi.org/10.1002/tesj.388

Yazan, B., Penton Herrera, L. J., & Rashed, D. (2023). Transnational TESOL practitioners' identity tensions: A collaborative autoethnography. *TESOL Quarterly, 57*(1), 140-167. https://doi.org/10.1002/tesq.3130

Zhao, Y. (2022). New context, new teachers, and new teacher education. *Journal of Technology and Teacher Education, 30*(2), 127-133. https://www.learntechlib.org/primary/p/221169/

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Exploring the synergistic effects of combining design thinking and project-based learning in a blended course

João Alberto Arantes do Amaral[A]

Luciano Gamez[B]

[A] *Assistant Professor, Department of Tecnology in Educational Design, Federal University of São Paulo, Brazil*

[B] *Associate Professor, Department of Tecnology in Educational Design, Federal University of São Paulo, Brazil*

Keywords

Design thinking;
project-based learning;
synergistic effects.

Correspondence

joao.arantes@unifesp.br [A]

Article Info

Received 23 June 2023
Received in revised form 17 August 2023
Accepted 22 August 2023
Available online 22 August 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.21

Abstract

This article presents an account of our experience in designing and implementing a course titled "Integration Project 3", in which we combined design thinking and project-based learning approaches. The course, conducted between March and June 2023, involved twelve undergraduate students from the Department of Technology in Educational Design at the Federal University of São Paulo situated in São Paulo, Brazil. Our research objective aimed to elucidate the synergistic effect of combining both approaches. Employing a mixed-method research design, we collected data from the students' project website, from focus group activities and from a questionnaire. We conducted a systemic analysis utilizing causal loop diagrams. The findings of our study are as follows: 1) The implementation of design thinking methodology in addressing challenges faced by community partners not only motivated the students to learn but also facilitated the development of their problem-solving skills. 2) The integration of project-based learning and design thinking methodologies engendered the development of students' project management skills and facilitated the application of acquired knowledge across various academic disciplines, thereby promoting interdisciplinary learning. 3) The students' determination to work on real-life project tasks was influenced positively by their motivation to learn and negatively by the stress due to real-life project constraints.

Introduction

The Department of Technology in Educational Design at the Federal University of São Paulo, known as TEDE, offers a two-and-a-half-year undergraduate course that focuses on developing instructional designers. Throughout the course, there are four capstone courses, one for each semester, called "Integration Projects". These courses aim to bridge the knowledge gained from supportive courses taken by students during each semester.

In "Integration Project 1", the objective is to encourage students to work on projects related to networked open learning experiences, such as designing and evaluating Massive Open Online Courses (MOOCs). In "Integration Project 2", students undertake projects that involve non-formal education, in collaboration with NGOs and public institutions.

The primary goal of "Integration Project 3" is to provide students with a practical opportunity to apply conceptual frameworks in designing educational experiences within the formal education context. This includes projects with schools at fundamental and high school levels. Lastly, "Integration Project 4" challenges students to collaborate with corporations in designing educational courses. These capstone disciplines serve as a way for students to integrate their learning across different areas and apply their knowledge and skills in practical settings, aligning with the overall objective of the instructional design program at TEDE.

In this article, we analyze "Integration project 3", developed in the first semester of 2023 and delivered to twelve students. The students were presented with a challenge: to develop educational products or initiatives, including courses, games, and more, specifically designed for Early Childhood Education Center Paulistinha, a public K-8 educational institution, our institutional partner (hereafter referred to as Paulistinha). To encourage collaboration and promote a harmonious working environment, the students were divided into teams of four. Each team was given the freedom to choose a single project from a carefully curated list (Table 1) provided by the principal of Paulistinha.

The course spanned a duration of 15 weeks and was implemented using a blended learning methodology, combining both online and face-to-face components. The students engaged in only three physical meetings with the professors throughout the course, with the majority of the activities conducted online. However, the students were granted autonomy to arrange additional in-person meetings with the institutional partner as necessary, facilitating the resolution of queries or enabling the testing of the prototypes developed by the students.

The course preparation

The Integration Project was led by two professors who assumed the responsibility of guiding the course. A month prior to its commencement, these professors proactively initiated contact with the principal of the school, articulating the overarching objectives of "Integration Project 3" and inquiring about any specific educational requirements the institution may have had. Subsequently, the principal promptly established communication channels with the school's teachers, proposing four distinct project themes aligned with the creation of educational products and initiatives (as summarized in Table 1).

Table 1. The project's themes.

Project theme	Description
Paulistinha's portal project	The task assigned to the students involved the creation of a comprehensive online portal that would serve as an informational hub for the school, encompassing detailed insights into the school's infrastructure, facilities, as well as highlighting the diverse range of activities undertaken by its students.
Paulistinha's memory project	The students were tasked with the mission of creating a dynamic online portal that would present the rich history of the school, spanning its inception in the year 1950 to the present day. This portal would serve as a captivating platform, incorporating a compelling blend of archival photographs, descriptive narratives, and engaging videos, all meticulously curated to capture and present the enduring legacy and evolution of the institution over the years.
Paulistinha's web radio project	The students have been entrusted with the task of developing a captivating web radio platform, which would serve as an innovative and dynamic medium for disseminating news and hosting podcasts meticulously curated by the school's teachers. This project would provide a unique opportunity for the school's students to harness their creative talents.
Paulistinha´s diversity Project	The students have been assigned the responsibility of creating a comprehensive digital repository of multimedia resources, encompassing diverse materials related to issues of diversity. This repository, consisting of texts, videos, and audios, is intended to be utilized by the teachers at school in their respective courses.

The professors informed the principal that the projects would be developed over a span of 12 weeks. Throughout the project's duration, students would engage with the school principal and teachers, seeking clarification about the products and initiatives to be developed. The professors underscored the significance of the school administrators providing prompt responses to the students' queries, to ensure that the progress of their projects is not impeded. The school representatives were requested to respond to the students' inquiries, ensuring that their responses were provided by the end of the week in which the questions were posed.

The professors leading the Integration Project proceeded to design a course that integrated project-based learning and design thinking approaches. The students would work in teams, with each team focusing on one of the project's themes. Subsequently, the professors shared their course concepts with their colleagues within the department, encouraging collaboration and seeking input for further improvement. The professors actively engaged in a constructive dialogue, asking probing questions and providing valuable suggestions to enhance the course design. Upon receiving the departmental approval, the two professors established a virtual learning environment using Moodle.

The course schedule was thoughtfully structured to enable students to follow MIT Teaching System Lab's (MITx, 2019) six-stage framework (Discover, Focus, Imagine, Prototype, Try, Reflect and Share) while working on their projects.

To document their progress, each team of students was mandated to develop a project website, furnishing a weekly report of their undertakings. Consequently, four student websites were established (one for Paulistinha's portal project, one for Paulistinha's memory project, one for Paulistinha's web radio project, and one for Paulistinha's diversity project). Additionally, they were encouraged to reflect upon their learning experiences on their project's website, documenting any challenges encountered and the solutions they employed to overcome them.

The course delivery

The course delivery was structured into clear and sequential phases, starting with the Discover and Focus phases, which spanned the initial month. During this period, student teams conducted visits to schools and engaged in meetings with principals, teachers, and students. Through interviews and collaborative activities, these interactions provided valuable insights into the context and challenges they would encounter. Furthermore, questionnaires were distributed to the teachers, facilitating the establishment of project boundaries to ensure minimal overlap between projects.

At the conclusion of the Discover and Focus phases, the students participated in a face-to-face meeting with the professors, where they presented their insights gained during this phase. Subsequently, the course transitioned into the Imagine and Prototype phases, which lasted for one month. Within this phase, student teams engaged in intensive brainstorming sessions, fostering the generation of diverse solutions to the identified problems. These efforts culminated in the development of prototypes.

Following the Imagine and Prototype phases, the course progressed into the Try, Reflect, and Share phases, which extended over a duration of two months. During these phases, the students actively tested their prototypes with the involvement of students and teachers of Paulistinha. Based on the feedback received, the students refined their prototypes, leading to the finalization of the products that were eventually delivered to the teachers at school.

Theoretical review

Design thinking is a systematic problem-solving approach that incorporates empathy, collaboration, and iteration as essential elements (Brown, 2008). It can be developed by following a method with defined stages. However, there is no consensus among researchers regarding the precise number or naming of stages within the design thinking process (Arantes do Amaral et al., 2023). Scholars propose different frameworks, including five-stage (Plattner et al., 2010), six-stage (Lewrick et al., 2018), and even eight-stage processes (Mueller-Roterberg, 2018). They are all very similar; they all begin with one stage that aims to understand and define the problem and the people who have this problem. Then, it progresses to figure out different solutions to that problem, followed by the development and testing of prototypes that could be used to solve the problem. These prototypes may evolve into a final product or service that solves the problem. Although these approaches share similarities, we chose to adopt the MIT Teaching System Lab's (MITx, 2019) six-stage framework for design thinking in our course. This framework consists of the following stages: discovery, focus, imagine, prototype, try, and reflect and share (Figure 1).

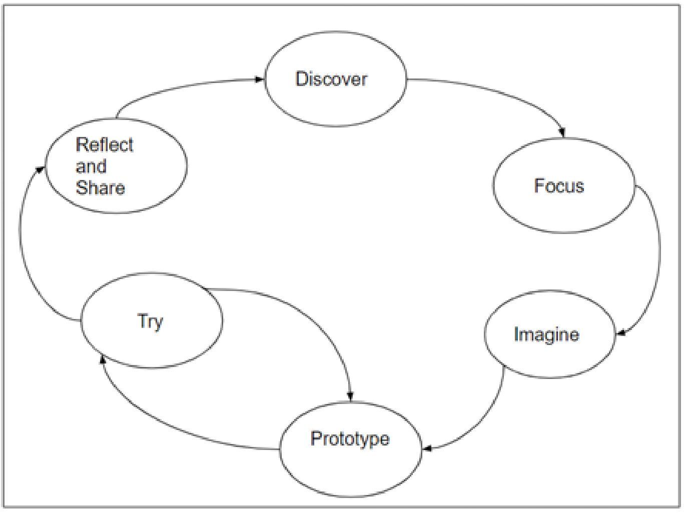


Figure 1. MIT Teaching System Lab's six-stage framework (based on MITx, 2019).

We determined that this particular framework would be more accessible and comprehensible for our students, enabling them to effectively understand and follow the design thinking process. The objective of the discovery stage is to develop an understanding of the problem and the individuals affected by it, commonly referred to as users. Designers employ various methods such as interviews, focus group activities, questionnaires, and field observations to gain a clear understanding of the contextual nuances. They immerse themselves in the users' context to glean insights. The discovery stage is also known by other scholars as empathize stage (Wolniak, 2017) or understand stage (Lewrick et al., 2018) .

During the focus stage, designers define the problem statement and identify key challenges. Other researchers call this stage as "define stage" (Wolniak, 2017). In the imagine stage, designers explore multiple possibilities to solve the problem and select the most viable option. The imagine stage is also referred as "ideate stage" by other scholars (Sándorová, 2020). The prototype and try stages are closely connected. Designers create prototypes, which can take the form of sketches, models, storyboards, software, or any other device, and subsequently subject them to rigorous testing. These two phases are highly interactive and often involve active participation from the users. Prototypes can be refined or discarded based on feedback received during these processes. The prototype and try stages are also referred as "create stage" by other researchers (Luchs, 2015).

During the reflection and sharing process, designers contemplate the solution, and the process followed. If the prototype successfully addresses the problem, the design process concludes. However, if further refinement is required, designers may initiate additional six-stage cycles to iterate and improve their design solution.

Contemporary educational institutions, including K-12 schools (Li & Zhan, 2022), universities (Arantes do Amaral et. al, 2022), and graduate schools (Kurokawa, 2013), have increasingly embraced the integration of design challenges and interdisciplinary projects into their curricula across various grade levels. This pedagogical approach enables students to apply design thinking methodologies and cultivate their ability to generate innovative solutions for real-world problems. By engaging in hands-on activities and collaborative ideation, students are afforded opportunities to identify problems, develop prototypes, and rigorously test their ideas. These projects encompass a diverse array of themes and objectives. For instance, students may undertake endeavors focused on formulating sustainable solutions for issues within their school or community, such as the establishment of recycling initiatives or the implementation of energy-efficient practices (Dotson et al., 2020; Earle & Leyva-de la Hiz, 2021). Alternatively, they may participate in initiatives aiming to promote inclusivity (Ballenger & Sinclair, 2020), collaborate with community organizations (Ramos et al., 2016) or businesses (Glen et al., 2015), engage in STEAM (Science, Technology, Engineering, the Arts and Mathematics) projects (Ananda et al., 2023), redesign courses (Acharya et al., 2021), or partake in humanitarian design endeavors, among other possibilities.

Designing and delivering a course centered on design thinking presents several challenges. Firstly, it necessitates a paradigm shift from a traditional teacher-centered approach to a student-driven model, thereby requiring a corresponding transformation in the mindset of educators (Noh & Abdul Karim, 2021). Secondly, design thinking is inherently challenging due to its reliance on interdisciplinary collaboration and the need to foster cooperation among teachers from diverse disciplines and perspectives (Maciver et al., 2016). Thirdly, incorporating design thinking processes within the course timeline presents a notable challenge.

In order to overcome these challenges, researchers (Beneroso & Robinson, 2022; Parmar, 2014) are integrating design thinking and project-based learning approaches. Project-based learning is a student-centered educational approach that fosters hands-on activities, fostering collaboration (Larmer et al., 2015) and allowing the teacher to incorporate design thinking processes in the course timeline (Boss & Larmer, 2018).

Academic discourse has notably emphasized the significance of affording students the opportunity to engage in substantive real-world projects (Lamer et al., 2015), endeavors which not only stimulate the development of students' abilities but also culminate in the creation of products or services that yield tangible societal advantages (Jacoby, 2014). Scholarly investigation has extended to encompass academic undertakings uniting university students with institutional collaborators, including entities such as non-governmental organizations (Arantes do Amaral, 2019), corporations (Badir et al., 2023), and educational institutions (Catapano & Gray, 2015; Kaldi et al., 2011). Nevertheless, scholars (Arantes do Amaral & Matsusaki, 2017) have also underscored the intricate challenges inherent in establishing and nurturing such collaborative alliances. Notably, scholars (Markham et al., 2003) have articulated the imperative of harmonizing institutional agendas with the academic calendar, thereby ensuring seamless integration. Furthermore, researchers (Arantes do Amaral, 2020) have underscored the potential challenges encountered within these collaborative ventures, spanning from limited partner commitment and delayed responsiveness to student inquiries to the discernment of inadequately aligned project themes.

In the present exposition, we embark on a comprehensive examination of a specific pedagogical endeavor involving university students and a public school institution. Within this context, our discourse delves into the convergence of project-based learning paradigms with the design thinking approach. Scholarly literature has extolled the virtues of amalgamating these two pedagogical frameworks, manifesting benefits such as the nurturing of students' innovation skills (Collins & Chiaramonte, 2017), fostering creativity (Cummings & Yur-Austin, 2022) and promoting empathy (Hashim et al., 2019).

However, there is still a notable gap in understanding the synergistic effects that arise from combining these approaches. This article seeks to address this gap by providing a comprehensive examination of the combined use of design thinking and project-based learning in a course, thereby offering insights into the potential benefits and outcomes of this integration.

Method

This study employed a sequential mixed methods approach (Creswell, 2014). Initially, qualitative data were collected by scrutinizing the projects' websites and conducting three focus group sessions with all student teams (one at the commencement of the course, one in the middle, and one at the conclusion). This allowed for the identification of recurring themes. Subsequently, an analysis was conducted to explore the interconnections among these themes using a causal loop diagram. This analysis revealed the need for further inquiry to gain a deeper understanding of the dynamics within the course. As a result, a questionnaire was administered to all students, gathering both quantitative and qualitative data. The collected data were analyzed, and based on the findings, refinements were made to the causal loop diagram. Through this iterative process, a comprehensive understanding of the synergistic effects arising from the combination of design thinking and project-based learning within the course was achieved.

Participants

Twelve students, consisting of eight males and four females, ranging in age from 24 to 45 years, participated in the course.

Data collection procedures

As mentioned earlier, qualitative data were collected from two sources: the students' project websites and focus group activities. The project websites served as a platform for the students to share their reflections on the completed activities, including their overall project experience and learning outcomes. Furthermore, the project websites provided insights into how the students applied the knowledge gained from other courses to their specific projects, illuminating their understanding of integrating and utilizing course content in their individual projects.

Quantitative and qualitative data were collected through the administration of a questionnaire (Appendix 1) at the end of the course. The questionnaire consisted of five distinct parts. Each part had closed-ended questions to be answered following a five-point Likert scale and one open-ended question.

Part 1 comprised four closed-ended questions that assessed the sources of stress experienced by students during the project, along with one closed-ended question related to the causal relationship between stress and willingness to perform project tasks. Part 2 included four closed-ended questions designed to gauge the students' perceptions of the application of design thinking methods in their projects, along with one closed-ended question related to the causal relationship between design thinking and motivation to learn. Part 3 consisted of an additional set of four closed-ended questions aimed at eliciting the students' perceptions and reflections on their utilization of project management tools, along with one closed-ended question related to the causal relationship between the development of project management skills and willingness to learn. Finally, Part 4 encompassed five closed-ended questions regarding the students' evaluation of the overall course structure. Part 5 consisted of five questions that explored the causal relationships between different factors. These included the relationship between stress and motivation to perform project tasks, the connection between the use of design thinking and motivation to learn, the correlation between the use of design thinking and the development of problem-solving skills, as well as the association between the development of project management skills and motivation to learn. Additionally, each part included an open-ended question prompting the students to provide any additional feedback or thoughts that were not addressed by the closed-ended questions.

Data analysis procedure

The qualitative data were analyzed following Yin's (2015) qualitative analysis method. First, we collected sentences from the project websites and notes from the focus group, creating a dataset. Then, we separated sentences with similar meanings into groups. After that, we created recurrent themes that summarized the main ideas.

Next, we created a causal loop diagram connecting variables representing the recurrent themes and performed a systemic analysis (Arantes do Amaral et al., 2023). Through this process, we realized that there were some causal relationships that needed further understanding and confirmation. Therefore, we administered the aforementioned questionnaire to the students to gather additional data. As mentioned previously, this questionnaire included both closed-ended and open-ended questions. The closed-ended questions provided us with quantitative data, while the open-ended questions provided qualitative data. To analyze the quantitative data, we developed an R program. For the qualitative data, we followed Yin's (2015) method again. Subsequently, we improved our causal loop diagram and analyzed the synergistic effects of combining design thinking and project-based learning in our course.

Results

Results from qualitative data

Recurrent Theme 1 (RT1): Delayed responses from school representatives

The students conveyed their challenges in establishing effective communication channels with the designated representatives of the educational institution. Furthermore, they expressed difficulties in obtaining timely feedback essential for the progression of their projects. Notably, one particular group of students highlighted the occurrence of delays in receiving feedback. These delays were attributed to the absence of their primary contact within the school, who had taken a holiday leave. As a result, a substitute teacher had to assume the responsibilities, impeding the timely dissemination of valuable feedback.

Recurrent Theme 2 (RT2): Lack of computational resources and internet connection at the school

During an on-site visit to the school, the students discovered a concerning inadequacy in computer availability, with only three units dispersed throughout the entire school. Additionally, the internet connection was characterized by frequent disruptions and sluggish performance. These circumstances posed a significant obstacle, considering that the majority of their projects focused on developing web-based products intended for use by students within the school.

Recurrent Theme 3 (RT3): Changes in project scope during the projects

The majority of project groups revealed that their initial project scopes underwent various modifications during the course of their work. These alterations ranged from minor adjustments to more substantial transformations. For instance, one group reported a radical shift in scope: their original project aimed to create a web-radio platform, but after two weeks, they significantly revised the scope to develop a virtual learning environment. This revised project aimed to facilitate content sharing among the school's students, encompassing textual, visual, and audio materials. The students expressed the perception that school representatives occasionally lacked a clear understanding of their own requirements and how the proposed projects could effectively address them.

Recurrent Theme 4 (RT4): The iterative process of prototype testing with the users helped to clarify the project scope

According to feedback from our students, the process of testing the prototypes has proven beneficial for both the students themselves and the teachers of the school. It has contributed to a more comprehensive understanding of the projects' scope. During the testing phase, students were able to identify limitations inherent in the prototypes, discern challenges experienced by the intended users (i.e., the children attending the school), and identify ambiguities in the product specifications. This testing phase has not only deepened our students' understanding of the final product to be created but has also empowered our school's teachers by illuminating numerous possibilities for integrating the product into their classroom activities.

Recurrent Theme 5 (RT5): The projects integrated the content of other courses, making them interdisciplinary

On their project websites, the students consistently documented their reflections on the interrelationships between the current course and their concurrent academic endeavors. They specifically articulated the application of various conceptual frameworks taught in other courses (such as universal design, project management, and innovation) This integration of diverse knowledge domains resulted in the interdisciplinary nature of their projects.

Recurrent Theme 6 (RT6): The combination of the design thinking method and the project-based learning approach enabled students to develop creative problem-solving skills, project management skills, collaboration, and empathy

As the course followed a project-based learning approach, the students were tasked with delivering specific artifacts, such as project charters, requirement analyses, and design documents, following the course's master schedule. The examination of the project websites revealed that the students diligently adhered to the six-step framework of design thinking. Notably, the students exhibited the application of empathy during their interactions with the school's teachers by adopting a perspective congruent with the educators' standpoint, engaging in active listening, extending support to develop information technology tools favorable for the enhancement of classroom activities, and sharing their experiential insights. In addition to that, the students employed project management tools to meticulously plan their projects and generated multiple innovative solutions during the imaginative phase of the design thinking process.

Results from quantitative data

In this section, we provide a summary of the findings regarding the distribution of students' responses to the closed-ended questions. Each question offered students the opportunity to select one of five options, according to the five-point Likert scale: "Totally disagree," "Disagree," "Neither agree nor disagree," "Agree," and "Totally agree." We analyze the data using four stacked charts, where each chart represents the percentage distribution of students' responses. The color scheme employed in the charts is as follows: dark brown denotes "Totally disagree," light brown corresponds to "Disagree," grey signifies "Neither agree nor disagree," light green represents "Agree," and dark green indicates "Totally agree."

Figure 2 illustrates the students' responses to four questions regarding the sources of stress experienced by students during the project (see Appendix 1). The stacked chart (Figure 2) shows that 64% of the students agreed or totally agreed that the delays in receiving feedback from Paulistinha were stressful. Only 18% of the students agreed that the changes in the project's schedule were stressful, and similarly, only 18% of the students agreed that the lack of Paulistinha's computational resources was stressful. Additionally, 18% of the students agreed or totally agreed that changes in the project's scope were stressful. Therefore, this leads to quantitative finding number 1 (QF1):

> The primary source of stress for students was found to be the delay in receiving responses to their inquiries.

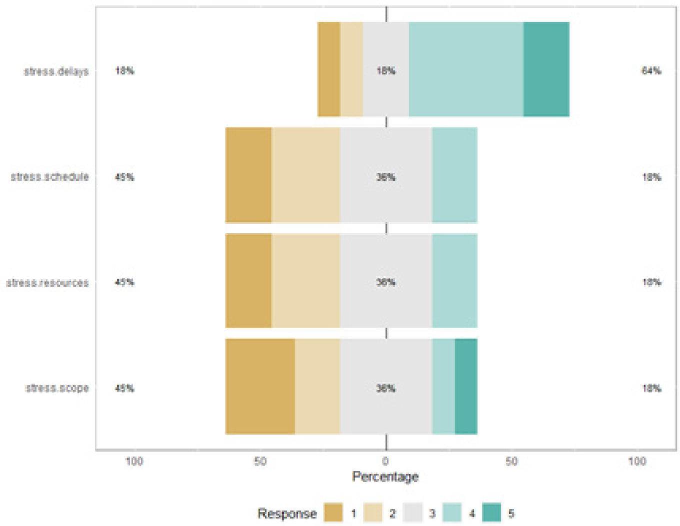


Figure 2. The students' answers related to stress.

Figure 3 presents the students' responses to four questions concerning their perceptions of the activities conducted during the design thinking stages (see Appendix 1). The stacked chart in Figure 3 reveals that a significant majority of 91% of the students agreed or completely agreed that they were able to comprehend the context of the community partner and define the requirements of the product during the 'discover' and 'focus' stages. Similarly, the same percentage of students agreed that they generated multiple solutions to the problem during the 'imagine' stage. Furthermore, all students unanimously agreed that they successfully created prototypes and conducted testing during the 'prototype' and 'try' stages. Additionally, 91% of the students agreed that they were able to engage in reflection on the processes undertaken and share insights during the reflect and share stages. Hence, these findings support quantitative finding number 2 (QF2):

> The students effectively implemented the activities advocated by design thinking methodology.

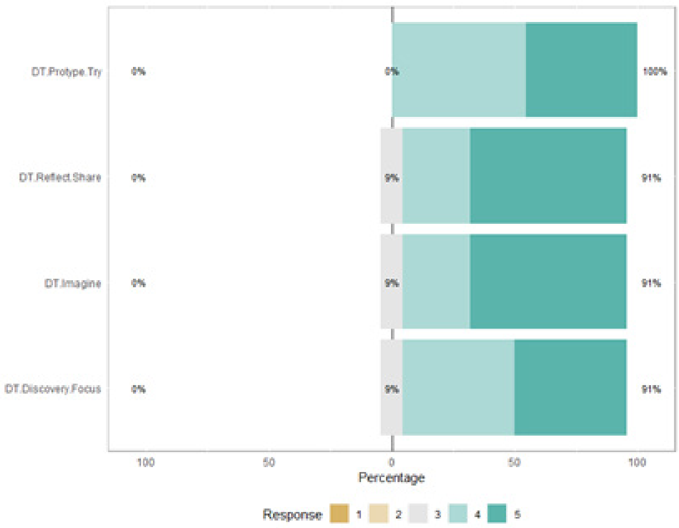


Figure 3. The students' answers related to the design thinking methodology.

Figure 4 presents the students' responses to four questions regarding their perceptions of the development of their project management skills (see Appendix 1). The stacked chart in Figure 4 reveals that all students agreed or totally agreed that they have developed project planning, project controlling, and project communication skills. Furthermore, a significant majority of students (91%) agreed or completely agreed that they were able to enhance their collaboration skills. These findings substantiate quantitative finding number 3 (QF3):

> The projects have facilitated the development of project management skills among the students.

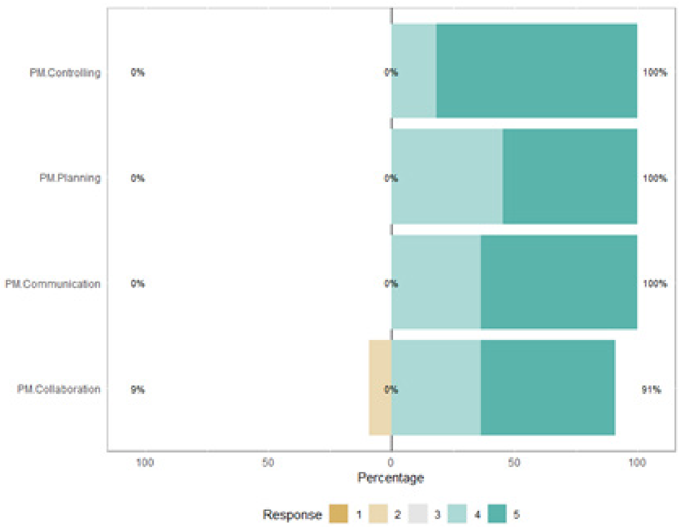


Figure 4. The students' answers related to the development of project management skills.

Figure 5 presents the students' responses to five questions pertaining to their perceptions of the course, including the teaching and learning strategy (the combination of project-based learning and design thinking), course management, and available resources (see Appendix 1). The stacked chart depicted in Figure 5 illuminates that all students agreed or totally agreed with the helpfulness of the professors' feedback, the well-organized Moodle learning environment, and the adequacy of the teaching and learning approach. Moreover, a significant majority of students (91%) agreed or completely agreed that the course met their expectations. Furthermore, a considerable 73% of the students agreed or completely agreed that working with a real-world community partner was a motivating experience. These findings substantiate quantitative finding number 4 (QF4):

> The teaching and learning strategies (the combination of project-based learning and design thinking), as well as the course management, have demonstrated effectiveness.

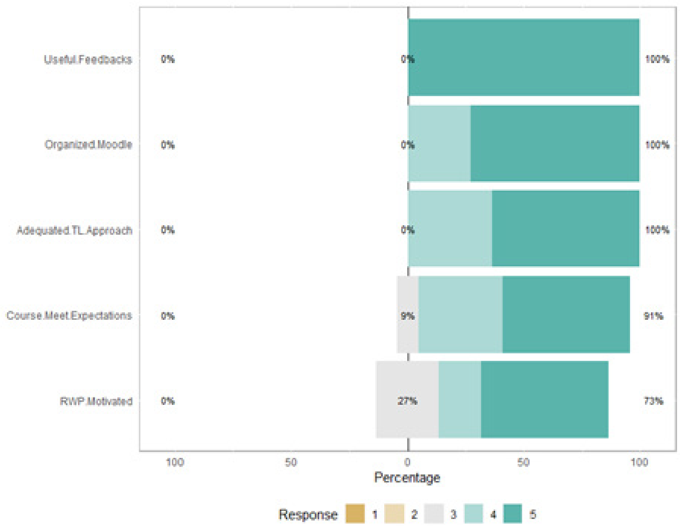


Figure 5. The students' answers related to the course.

Figure 6 illustrates the students' responses to five questions pertaining to their perception of causal relationships (refer to Appendix 1). The stacked chart depicted in Figure 6 reveals that a notable majority of 91% of the students agreed or totally agreed that there exists a positive causal relationship between the development of their project management skills and their motivation to learn. Similarly, the same percentage of students agreed or totally agreed that there is a positive causal relationship between the use of design thinking and their problem-solving skills. Additionally, 82% of the students agreed or totally agreed that a positive relationship exists between the use of design thinking and their motivation to learn. Moreover, 73% of the students concurred that there is a positive relationship between the accomplishment of project tasks and the development of project management skills. Furthermore, 45% of the students agreed or totally agreed that stress had a negative impact on their motivation to work on project tasks. These findings substantiate QF5 to QF8:

QF5: The development of project management skills and the use of design thinking positively impacted their motivation to learn.

QF6: The combination of design thinking and project-based learning positively influenced the development of problem-solving skills.

QF7: Stress had a negative impact on the motivation to work on projects for nearly half of the students.

QF8: The accomplishment of project tasks positively affected the development of project management skills.

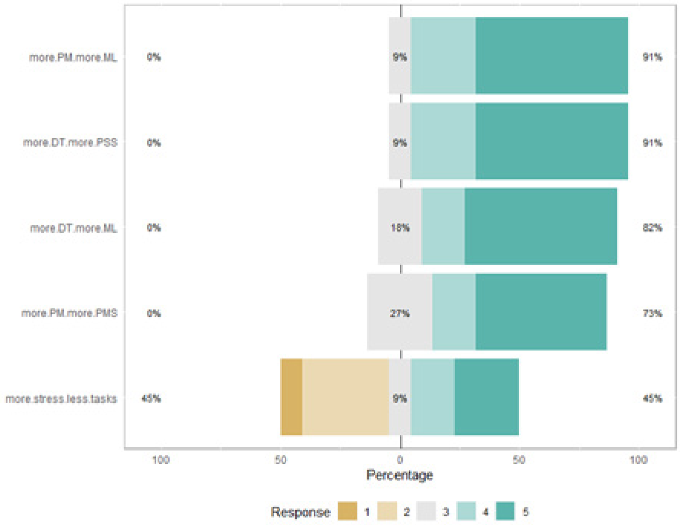


Figure 6. The students' perceptions about the causal relationships.

Discussion

The data (RT4, QF2, RT6, QF3) revealed that through the iterative processes of design thinking, the students improved their problem-solving skills. The data also revealed (RT6 and QF6) that as the students continued to refine their problem-solving skills, their motivation to learn intensified, as illustrated by the positive feedback loop labeled "Developing Problem-Solving Skills" in Figure 7. Hence, this leads us to our first finding:

> The implementation of the design thinking methodology in addressing challenges faced by community partners not only motivated the students to learn but also facilitated the development of their problem-solving skills.

This finding aligns with the research conducted by Guaman-Quintanilla et al. (2023), in which they observed that integrating design thinking in higher education settings enhances students' problem-solving abilities. This finding is also in line with the research conducted by Hashim et al. (2019), wherein they observed that the utilization of design thinking cultivates empathic relationships, thereby promoting students' motivation to learn.

The data (QF2, QF4, QF6, and QF8) revealed that the combination of project-based learning and design thinking approaches facilitated the engagement of students in authentic project-based tasks that required the sharing of ideas and collaborative efforts for project planning and execution (RT6 and QF3). Moreover, the students applied what they had learned in other courses they were taking at the same time that they were developing the project, notably the Project Management course (RT5). Based on the quantitative results (QF5), we may affirm that the development of project management skills had a positive influence on their motivation to learn, as indicated by the positive feedback loop labeled "Developing Project Management Skills" in Figure 7. This led us to our second finding:

> The integration of project-based learning and design thinking methodologies engendered the development of students' project management skills and facilitated the application of acquired knowledge across various academic disciplines, thereby promoting interdisciplinary learning.

This finding is aligned with the findings of other researchers (Ewin et al., 2017; Dijksterhuis & Silvius, 2017), who pointed out the connection between the development of project management skills and the use of design thinking methodology. It also aligns with the findings of Ge and Wang (2021), who pointed out that the combination of both approaches can promote interdisciplinary learning.

However, engaging with the school introduced various challenges for the students. Notably, they encountered issues concerning the responsiveness of school representatives in addressing their inquiries promptly (RT1). Furthermore, a significant number of projects experienced alterations to the project scope initiated by the school representatives after project initiation (RT3). Additionally, resource limitations, such as insufficient access to computers for prototype testing, were encountered by the students (RT2). Moreover, the students faced time constraints imposed by the course schedule. Collectively, these challenges heightened student stress levels, thereby exerting a detrimental impact on their determination to work on project tasks (QF7), creating a negative feedback loop (see Figure 7, loop labeled "Impacts of Real-life Project Constraints"). On the other hand, this negative feedback loop was overcome by the dynamics that led to an increase in motivation to learn (see Figure 7, the positive feedback loop "Development of Skills Motivates to Learn"). This leads us to our third finding:

> The students' determination to work on real-life project tasks was positively influenced by their motivation to learn and negatively by the stress due to real-life project constraints.

This finding is consistent with the research conducted by Lake and Whipps (2016), which highlights that involvement with community partners can be characterized by complexities and unpredictability. This finding is also in alignment with the research conducted by Arantes do Amaral (2019), which suggests that community-based learning creates learning benefits but may also induce stress among students.

Conclusion

Our empirical investigation has provided valuable insights into the synergistic effects of integrating design thinking and project-based learning in the course, resulting in a mutually reinforcing and enriched learning experience. By

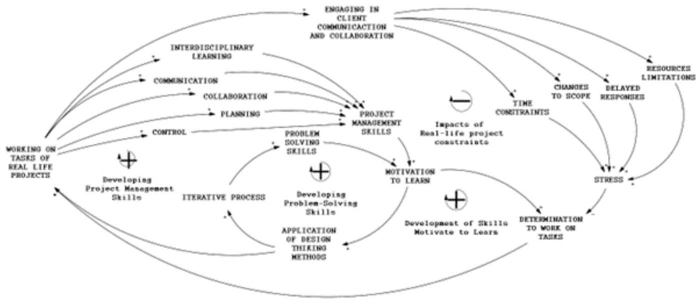


Figure 7. The causal loop diagram that represents the course's dynamics.

immersing students in real-life challenges and guiding them through the iterative design process, design thinking fosters their ability to comprehend complex problems and generate innovative solutions. This approach encourages students to think creatively, consider multiple perspectives, and embrace a human-centered approach to problem-solving.

Furthermore, the incorporation of project-based learning as a scaffold in the course has facilitated collaborative knowledge sharing among students and the development of their project management skills. Working on authentic projects with community partners has enabled students to apply their knowledge and skills in a practical context, leading to a deeper understanding of the subject matter and fostering interdisciplinary connections. The combination of design thinking and project-based learning has not only expanded students' academic horizons but also nurtured their ability to work collaboratively, communicate effectively, and manage projects successfully. Moreover, we can confidently state that this educational approach has significantly contributed to the competences and skills of our students, who are prospective educational designers.

Our findings emphasize the importance of effective communication channels with community partners, proactive management of project scope, and adequate allocation of resources to create a supportive learning environment. We understand that projects involving university students and schools can be challenging. It's a learning process; it takes time for school administrators to understand the importance of providing timely feedback to the students. Hopefully, in future projects, as the school administrators become accustomed to the communication processes and project schedules, the issues we faced in these projects will be minimized.

Additionally, strategies to manage and alleviate student stress should be considered, such as providing additional support and flexibility within the course schedule. By addressing these challenges, educators and institutions can further enhance the effectiveness of the synergistic integration of design thinking and project-based learning.

In conclusion, our study provides compelling evidence for the synergistic effects of integrating design thinking and project-based learning in promoting students' problem-solving skills, interdisciplinary learning, and motivation to learn. By embracing these approaches, educators can empower students to become innovative thinkers, adaptable problem solvers, and collaborative contributors in various professional settings. This research contributes to the growing body of knowledge on innovative teaching and learning methodologies and offers valuable insights for educators and institutions striving to enhance student learning experiences.

Limitations

One might argue that the case study involved only 12 students, potentially resulting in a small sample size for drawing robust conclusions. However, it is worth highlighting that this case study facilitated an in-depth exploration of the course dynamics, enabling a detailed examination of interactions among university students, administrators, and teachers. Speculatively, these dynamics could conceivably extend to analogous academic contexts featuring larger student cohorts. Moreover, one could contemplate the inclusion of multiple schools for comparative purposes. Nevertheless, introducing more schools might not inherently lead to a more comprehensive study. The act of comparing and contrasting across multiple institutions could introduce confounding variables, including varying school cultures, demographics, and administrative structures. These factors could potentially complicate the analysis and interpretation of the findings.

References

Acharya, S., Bhatt, A. N., Chakrabarti, A., Delhi, V. S., Diehl, J. C., van Andel, E., Jurelionis, A., Stasiuliene, L., Jussilainen Costa, L. D., & Subra, R. (2021). Problem-Based Learning (PBL) in undergraduate education: Design thinking to redesign courses. In *Design for tomorrow—volume 2: Proceedings of ICoRD 2021* (pp. 349-360). Springer Singapore.

Ananda, L. R., Rahmawati, Y., & Khairi, F. (2023). Critical thinking skills of Chemistry students by integrating design thinking with STEAM-PjBL. *JOTSE: Journal of Technology and Science Education, 13*(1), 352-367.

Arantes do Amaral, J. A. (2019). Combining community-based learning and project-based learning: A qualitative systemic analysis of the experiences and perceptions of students and community partners. *Partnerships: A Journal of Service-Learning and Civic Engagement, 10*(1), 129-145.

Arantes do Amaral, J. A., & Matsusaki, C. T. M. (2017). The dynamics of connecting universities, non-governmental organizations and community members by means of academic projects directed at people in need. *Educational Action Research, 25*(2), 280-299.

Arantes do Amaral, J. A. (2020). The problems that impact the quality of project management courses developed following a project-based learning approach with the support of community partners. *Journal of Problem Based Learning in Higher Education, 8*(2), 106-114.

Arantes do Amaral, J. A., Meister, I. P., Lima, V. S. (2022). Using design thinking tools to inform the design process of a massive open online course on using Scratch for teachers.

Journal of Design Thinking, 3(1),123-134.

Arantes do Amaral, J. A.., Meister, I. P., & Gamez, L. (2023). A systemic analysis of the impacts of the covid-19 pandemic on the studies of Brazilian graduate students: An exploratory study. *Anatolian Journal of Education, 8*(1), 173-190.

Badir, A., O'Neill, R., Kinzli, K. D., Komisar, S., & Kim, J. Y. (2023). Fostering Project-Based Learning through industry engagement in Capstone design projects. *Education sciences, 13*(4), 361.

Ballenger, S., & Sinclair, N. (2020). Inclusive design thinking-model for inclusive course development. *Online Journal of Distance Learning Administration, 23*(4), 1-7.

Beneroso, D., & Robinson, J. (2022). Online project-based learning in engineering design: Supporting the acquisition of design skills. *Education for Chemical Engineers, 38*, 38-47.

Boss, S., & Larmer, J. (2018). *Project based teaching: How to create rigorous and engaging learning experiences.* ASCD.

Brown, T. (2008). Design thinking. *Harvard Business Review, 86*(6), 84–92.

Catapano, S., & Gray, J. (2015). Saturday school: Implementing Project-Based Learning in an urban school. *Penn GSE Perspectives on Urban Education, 12*(1), 1-12.

Collins, J. A., & Chiaramonte, M. W. (2017). Project-based learning and design thinking: Fomenting agility and innovation in government. In *2017 IEEE International Professional Communication conference (ProComm)* (pp. 1-8). IEEE.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.

Cummings, C., & Yur-Austin, J. (2022). Design thinking and community impact: A case study of project-based learning in an MBA capstone course. *Journal of Education for Business, 97*(2), 126-132.

Dijksterhuis, E., & Silvius, G. (2017). The design thinking approach to projects. *The Journal of Modern Project Management, 4*(3), 33-41.

Dotson, M. E., Alvarez, V., Tackett, M., Asturias, G., Leon, I., & Ramanujam, N. (2020, February). Design thinking-based STEM learning: Preliminary results on achieving scale and sustainability through the IGNITE model. In *Frontiers in education* (Vol. 5, p. 14). Frontiers Media SA.

Earle, A. G., & Leyva-de la Hiz, D. I. (2021). The wicked problem of teaching about wicked problems: Design thinking and emerging technologies in sustainability education. *Management Learning, 52*(5), 581-603.

Ewin, N., Luck, J., Chugh, R., & Jarvis, J. (2017). Rethinking project management education: a humanistic approach based on design thinking. *Procedia Computer Science, 121*, 503-510.

Ge, X., & Wang, Q. (2021). Cultivating design thinking in an interdisciplinary collaborative project-based learning environment. *Intersections Across Disciplines: Interdisciplinarity and Learning*, 187-196.

Glen, R., Suciu, C., Baughn, C. C., & Anson, R. (2015). Teaching design thinking in business schools. *The International Journal of Management Education, 13*(2), 182-192.

Guaman-Quintanilla, S., Everaert, P., Chiluiza, K., & Valcke, M. (2023). Impact of design thinking in higher education: A multi-actor perspective on problem solving and creativity. *International Journal of Technology and Design Education, 33*(1), 217-240.

Hashim, A. M., Aris, S. R. S., & Chan, Y. F. (2019). Promoting empathy using design thinking in project-based learning and as a classroom culture. *Asian Journal of University Education, 15*(3), 14-23.

Jacoby, B. (2014). *Service-learning essentials: Questions, answers, and lessons learned.* John Wiley & Sons.

Kaldi, S., Filippatou, D., & Govaris, C. (2011). Project-based learning in primary schools: Effects on pupils' learning and attitudes. *Education, 9*(1), 35-47.

Kurokawa, T. (2013). Design thinking education at universities and graduate schools. *Science & Technology Trends Quarterly Review, 46*, 50-62.

Lake, D., Ricco, M. E., & Whipps, J. (2016). Design thinking accelerated leadership: Transforming self, transforming community. *The Journal of General Education, 65*(3-4), 159-177.

Larmer, J., Mergendoller, J., & Boss, S. (2015). *Setting the standard for project based learning*. ASCD.

Lewrick, M., Link, P., & Leifer, L. (2018). *The design thinking playbook: Mindful digital transformation of teams, products, services, businesses and ecosystems.* John Wiley & Sons.

Li, T., & Zhan, Z. (2022). A systematic review on design thinking Integrated Learning in K-12 education. *Applied Sciences, 12*(16), 8077.

Luchs, M. G. (2015). A brief introduction to design thinking. *Design thinking: New product development essentials from the PDMA*, 1-12.

Maciver, F., Malins, J., Kantorovich, J., & Liapis, A. (2016). United we stand: A critique of the design thinking approach in interdisciplinary innovation. In P. Lloyd & E. Bohemia (Eds.), *Future focused thinking - DRS international conference 2016* (pp. 37-46). Brighton, United Kingdom.

Markham, T., Larmer, J., & Ravitz, J. (2003). *Project based learning handbook: A guide to standards-focused project based learning for middle and high school teachers.* Buck Institute for Education (BIE).

MITx. (2019). *Design thinking for leading and learning: Facilitator's guide.* https://prod-edxapp.edx-cdn.org/assets/courseware/v1/efe4aa7441824dca8770b11170b32b24/asset-v1:MITx+11.155x+1T2019+type@asset+block/DTLL_FacGuide2019_v8.pdf

Mueller-Roterberg, C. (2018). *Handbook of design thinking.* Hochschule Ruhr West.

Noh, S. C., & Abdul Karim, A. M. (2021). Design thinking mindset to enhance education 4.0 competitiveness in Malaysia. *International Journal of Evaluation and Research in Education, 10*(2), 494-501.

Parmar, A. J. (2014, October). Bridging gaps in engineering education: Design thinking a critical factor for project based learning. In *2014 IEEE Frontiers in education conference (FIE) proceedings* (pp. 1-8). IEEE.

Plattner, H., Meinel, C., & Leifer, L. (Eds.). (2010). *Design thinking: Understand–improve–apply.* Springer Science & Business Media.

Ramos, A. K., Trinidad, N., Correa, A., & Rivera, R. (2016). Partnering for health with Nebraska's latina immigrant community using design thinking process. *Progress in Community Health Partnerships: Rsearch, Education, and Action, 10*(2), 311-318.

Sanger, P. A., & Ziyatdinova, J. (2014). Project based learning: Real world experiential projects creating the 21st century engineer. In *2014 International conference on Interactive Collaborative Learning (ICL)* (pp. 541-544). IEEE.

Wolniak, R. (2017). The design thinking method and its stages. *Systemy Wspomagania w Inżynierii Produkcji, 6*(6), 247-255.

Appendix

Appendix A: The questionnaire.

Questionnaire - Closed-ended Questions
For each statement, students had five choices (Totally Disagree, Disagree, Neither Agree nor Disagree, Agree, Totally Agree).

Part 1: Questions about stress:

1. The changes in the project scope caused me stress.
2. The delayed response from the institutional partner to our questions/requests regarding the project caused me stress.
3. The limited computational resources in the school for conducting tests stressed me out.
4. The short deadline for project tasks stressed me out.

Part 2: Questions about design thinking:

1. The DISCOVER and FOCUS processes helped me understand the context of our community partner and define the product requirements.
2. The IMAGINE process made me think of various solutions to the proposed problem.
3. The PROTOTYPING and TRY processes helped me develop a prototype to solve the proposed problem.
4. The REFLECT AND SHARE process made me reflect on the prototype development process and the project itself.

Part 3: Questions about project management:

1. The need to communicate with my colleagues and institutional partners helped me develop communication skills.
2. The need to work in a team helped me develop collaboration skills.
3. The need for planning helped me develop planning skills.
4. The need to record project tasks on the website helped me develop my task management skills.

Part 4: Questions about course structure:

1. The pedagogical proposal and methodology of the integrative project were adequate.
2. Working on a real project for a real institutional partner (Paulistinha) has been motivating.
3. The Moodle classroom is organized clearly, allowing easy access to materials.
4. The weekly feedback (in videos) has been useful.
5. Overall, the course met my expectations.

Part 5: Questions about causal relationships:

1. The more stressed I become, the less motivated I am to perform project tasks.
2. The use of Design Thinking has motivated me to learn more.
3. The integrative project has helped me develop managerial skills.
4. The application of DESIGN THINKING processes has improved my ability to solve real problems.
5. The development of my managerial skills has increased my motivation to learn.

Questionnaire: Open-ended Questions

1. Please tell me if you have experienced any stressful situations during the project.
2. Do you have anything else to say about your experience with the use of design thinking in the project?
3. Is there anything you would like to add regarding your experience in managing your project?
4. Do you have anything else to add about your experience in the integrative project?

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : **2591-801X**

Content Available at : **http://journals.sfu.ca/jalt/index.php/jalt/index**

The impact of NVivo in qualitative research: Perspectives from graduate students

Pongsakorn Limna[A]

[A] *Rangsit University, Pathum Thani, Thailand*

Keywords

Analysis software;
graduate students;
NVivo;
perspectives;
qualitative data;
Thailand.

Correspondence

pongsakorn.l65@rsu.ac.th [A]

Article Info

Received 14 June 2023
Received in revised form 29 July 2023
Accepted 31 July 2023
Available online 3 August 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.17

Abstract

This study aims to explore the impact of NVivo, a qualitative data analysis software, from the perspective of graduate students. The study employed a qualitative method, utilizing in-depth interviews with ten graduate students selected through purposive sampling. The data were analyzed using content analysis and NVivo. NVivo has become a valuable asset in the toolkit of graduate students engaged in qualitative research. The software's ability to organize and analyze vast amounts of qualitative data enhances research efficiency, facilitates collaboration, and supports the production of high-quality research outcomes. However, it is essential to recognize the challenges associated with implementing NVivo and to provide the necessary resources and support to students, ensuring they can fully leverage its potential. As graduate students continue to embrace NVivo in their research endeavors, its impact on qualitative research is expected to grow significantly, advancing the field and contributing to valuable insights and discoveries. This study contributes to the broader field of qualitative research by providing valuable insights into the impact of NVivo from the perspective of graduate students. It enhances our understanding of how NVivo enhances research efficiency, facilitates collaboration, and addresses challenges in implementing the software. These contributions contribute to the advancement of qualitative research methodologies and inform best practices for utilizing NVivo effectively.

Introduction

Every study must include an explicit, disciplined, and systematic approach in order to produce the most appropriate results. Qualitative research is a common component of graduate students' research projects as it offers several benefits for graduate students, particularly in disciplines such as social sciences, humanities, education, and health sciences (Hunt et al., 2009; Mohajan, 2018; Tilley, 2019; Tomaszewski et al., 2020). Qualitative research is a research method used to gather in-depth understanding and insights into human behavior, attitudes, experiences, and perceptions. It aims to explore and interpret the meanings, beliefs, and motivations underlying individuals' actions and interactions within a specific context. Unlike quantitative research that focuses on numerical data and statistical analysis, qualitative research is primarily concerned with capturing rich, descriptive data. It relies on methods such as interviews, focus groups, observations, case studies, and textual analysis to gather information. These methods allow researchers to delve into the participants' perspectives, emotions, and subjective experiences, providing a detailed understanding of the research topic (Austin & Sutton, 2014; Taherdoost, 2021; Tenny et al., 2022). In short, qualitative research is primarily concerned with understanding the "how" and "why" aspects of a phenomenon, as opposed to the quantitative approach that focuses on "how many" or "how often" something occurs (Mulisa, 2022).

Qualitative research often employs a flexible and iterative approach, allowing researchers to adapt their data collection and analysis techniques as they gain new insights. It emphasizes the researcher's involvement in the research process, recognizing their influence on data collection, interpretation, and representation. This involvement is known as reflexivity and involves acknowledging the potential biases and subjectivity that researchers bring to the study (Busetto et al., 2020; Köhler et al., 2022). When conducting qualitative research, researchers usually analyze data by coding and categorizing information. They identify patterns, themes, and relationships within the data, extracting key ideas and concepts and interpreting their significance within the broader research context. This analysis not only aids researchers in exploring complex phenomena, generating theories, informing policy decisions, and improving practices but also provides a deeper understanding of human experiences and social processes (Bengtsson, 2016; Collins & Stockton, 2018; Mohajan, 2018; Tomaszewski et al., 2020).

In recent years, the use of qualitative data analysis software such as NVivo has gained popularity among researchers in various disciplines due to its potential to streamline and enhance qualitative research processes. One of the notable advantages of NVivo is its ability to efficiently manage and organize large volumes of qualitative data, ranging from textual documents to multimedia sources, enabling researchers to delve deeper into complex datasets. By providing a user-friendly interface and a comprehensive set of tools for coding, categorization, and exploration of data, NVivo facilitates the extraction of meaningful insights and patterns from diverse sources, ultimately contributing to the rigor and depth of research findings. Moreover, NVivo supports collaborative work, allowing multiple researchers to collaborate and share data, fostering a sense of teamwork and enhancing overall research productivity (Allsop et al., 2022; Fearnley, 2022; Niedbalski & Ślęzak, 2023).

Despite these advantages, it is essential to acknowledge some potential drawbacks associated with NVivo. First and foremost, the cost of acquiring NVivo licenses and the need for additional training might pose financial and time constraints, particularly for individual researchers or smaller research projects with limited resources. Furthermore, the learning curve for mastering NVivo's advanced functionalities could be steep, potentially requiring considerable effort and practice to fully exploit the software's capabilities. Researchers must also consider the potential risk of technical issues and data compatibility challenges that may arise when dealing with large or complex datasets, which could impede the smooth progress of research activities. Notwithstanding these limitations, the benefits offered by NVivo in terms of data management, analysis, and collaboration make it a compelling choice for qualitative researchers seeking to maximize the potential of their research endeavors (Kaefer et al., 2015; Dollah et al., 2017; Lund et al., 2023).

While other articles may have reached similar conclusions regarding NVivo's benefits and challenges, they have used different participant groups, which can lead to variations in experiences and perspectives. This study aims to explain the impact of NVivo, a qualitative data analysis software, from the perspectives of graduate students. By specifically examining the perspectives of graduate students, this research provides unique insights into how NVivo influences the qualitative research process from the viewpoint of individuals who are at a critical stage of their academic and research careers. This research's originality lies in its unique focus on graduate students as the participant group, providing valuable and distinct insights into the impact of NVivo on their qualitative research experiences. The choice of participants, in this case, is crucial as graduate students represent a distinct demographic with specific research needs and challenges. Their academic pursuits often involve extensive qualitative research, making NVivo an essential tool in their research toolkit. As a result, their experiences with NVivo might differ from those of other researchers or professionals. The study highlights how NVivo contributes to the efficiency and quality of graduate students' research outcomes, fostering collaboration and facilitating the production of valuable insights and discoveries. Additionally, it acknowledges the challenges they face when implementing the software, which may differ from those encountered by other participant groups. Furthermore, this research contributes to the broader field of qualitative research by providing a targeted examination of NVivo's impact on graduate students. While other studies may have covered a wide range of participants or focused on researchers from different academic levels, this study specifically homes in on graduate students, filling a potential gap in the existing literature. The findings shed light on how NVivo benefits this particular group and provides recommendations for optimizing its utilization to meet their specific needs and circumstances.

The paper is structured into six sections, each contributing to a comprehensive and informative analysis. The first section introduces the study, emphasizing its significance and objectives. In the second section, a thorough review of existing literature is conducted. The third section outlines the research methodology employed to collect relevant data. The fourth section presents the findings and results obtained from the study. In the fifth section, a detailed discussion takes place, analyzing and interpreting the results in relation to the research objectives. The final section presents the conclusions drawn from the study, addresses any limitations encountered during the research process, and provides recommendations for future studies.

Research objective

In the realm of qualitative research, NVivo holds significant potential for enhancing the research process and outcomes. Therefore, it is a critical topic to study. This study aims to explain the impact of NVivo, a qualitative data analysis software, from the perspectives of graduate students. By providing insights into the benefits and challenges faced by graduate students when employing NVivo in their research projects, this study seeks to shed light on the value of incorporating NVivo into qualitative research methodologies.

Research question

What are the perspectives of graduate students on the impact of NVivo software on the conduct of their qualitative research?

Related literature review

NVivo is a powerful qualitative data analysis software that provides researchers and analysts with a comprehensive set of tools not only to manage and organize but also to analyze qualitative data. NVivo offers a user-friendly interface and a wide range of features designed to facilitate in-depth exploration and understanding of qualitative research data (Allsop et al., 2022; Dhakal, 2022). With NVivo, researchers can import various types of data, including text documents, audio and video recordings, images, surveys, social media data, and more. The software supports multiple file formats, allowing users to seamlessly integrate and work with diverse data sources. In addition, one of the key strengths of NVivo is its data organization capabilities. Users can create nodes or codes to label and tag specific sections of their data, making it easier to categorize and retrieve information based on themes or topics of interest. This flexible coding system enables researchers to analyze data at a granular level, identify patterns, and uncover meaningful insights (Jackson & Bazeley, 2019; Kent State University, 2023).

NVivo also offers support for those employing mixed methods in their research. NVivo enables researchers to analyze open-ended survey questions and conduct comparisons based on demographic data. Moreover, the software allows for seamless data exchange with other applications such as Microsoft Excel and IBM SPSS Statistics. This integration enhances the versatility of NVivo, empowering researchers to incorporate quantitative elements into their analysis while benefiting from the robust qualitative analysis capabilities the software provides. NVivo offers a range of analysis tools tailored to different data types. For text-based data, users can perform text searches, conduct word frequency analyses, code and annotate documents, and use various techniques for text analysis. The software also supports the analysis of multimedia content, allowing users to transcribe audio and video recordings, code specific segments, and analyze visual data to gain a deeper understanding of their research material (NVivo, n.d.; O'Donoghue, 2022; University of Illinois, 2023).

The significance of software in supporting data analysis for researchers is increasingly recognized. In the realm of social sciences, the Statistical Package for the Social Sciences (SPSS) has long been the dominant tool for quantitative data analysis. However, there exists a broader array of software tools available to facilitate qualitative data analysis. Moreover, there has been a noticeable rise in the mention of qualitative software tools in published works (Sotiriadou et al., 2014). The utilization of qualitative data analysis software has witnessed significant growth in the past twenty years, benefiting from advancements in technology that enable researchers to analyze larger volumes of data at a faster pace and employ more intricate analytical approaches (Robins & Eisen, 2017). In the field of qualitative management and business studies, NVivo has been the prevailing software package of choice (Sotiriadou et al., 2014).

Johnston (2006) provided insights gained from over 11 years of experience in teaching and utilizing qualitative data analysis software, specifically QSR NUD*IST and NVivo, within doctoral research. The author highlights the challenges that arise from the existing separation between qualitative methods training and qualitative data analysis software training, which often leads to disjointed learning curves in terms of technical skills and methodological understanding. This poses difficulties for doctoral students, as they may not receive the necessary support and exposure to navigate the ongoing 'methods revolution' through methods literature, their postgraduate training programs, or their supervisors. The paper discussed three significant factors influencing the use of qualitative data analysis software, specifically QSR NVivo, in the context of teaching and learning. Firstly, there is an increasing popularity of qualitative data analysis software among individuals from traditionally positivistic backgrounds, indicating a shift towards embracing computer-assisted approaches in qualitative research. Secondly, the software offers the promise of enhanced transparency in research processes, which can be appealing to researchers seeking to improve rigor and credibility. Lastly, many individuals rely on the free tutorials provided with the software to acquire both qualitative research methods knowledge and proficiency in software operations.

Woods et al. (2016) discussed the use of qualitative data analysis software programs, including ATLAS.ti™ and NVivo™, in qualitative research. The analysis revealed several key findings. Firstly, the number of articles reporting the use of qualitative data analysis software programs is increasing annually, indicating the growing adoption of these tools in qualitative research. The majority of studies utilizing ATLAS.ti™ and NVivo™ were published in health sciences journals, and the authors primarily originated from the United Kingdom, United States, Netherlands, Canada, and Australia. Regarding research design, qualitative data analysis software programs were employed to support various types of studies. The most commonly analyzed data sources included interviews, focus groups, documents, field notes, and open-ended survey questions. Researchers predominantly used qualitative data analysis software for data management and analysis, with fewer utilizing it for data collection or creation, as well as for visually representing their methods and findings. The study has raised important considerations regarding the extent to which qualitative data analysis software users have fully leveraged the potential of these programs to support new research approaches. While the usage of qualitative data analysis software programs has increased, the analysis suggests that there may still be untapped opportunities for utilizing these tools in innovative ways.

Dalkin et al. (2021) addressed the need for enhanced transparency in realist methods, specifically in the process of developing and refining program theories. Their study showcases the utilization of NUD*IST Vivo (NVivo), a computer-assisted qualitative data analysis software, for the purpose of building and refining program theories through the analysis of literature and interview data within a realist evaluation. The authors emphasize the intricacies involved in coding various data sources into nodes and child nodes, alongside the creation of 'attached memos' to document the progression of theory generation. NVivo plays a pivotal role in establishing a transparent and well-documented audit trail of programme theory refinement, aligning with the growing demand for increased transparency in realist analysis. The RAMESES I and II initiatives have made significant contributions by introducing consensus-based and evidence-based reporting guidelines to enhance transparency in reporting realistic research. Additionally, incorporating NVivo into realist approaches can offer a structured framework for the iterative and complex process of generating, refining, and testing intricate program theories using multiple data sources concurrently. This utilization of NVivo effectively establishes a well-organized record of the analytical process, reinforcing its rigor and transparency.

Elliott (2022) discussed the preferred methods for analyzing large volumes of text responses, particularly in the context of research involving open-ended questions (OEQs). When dealing with thousands of responses and focusing primarily on the content of the text rather than its form, automated analysis is often favored over using specialized software like NVivo (a CAQDAS package). Large sample sizes can help mitigate imprecision and potential misinterpretations caused by the ambiguity of language. An example is given of a study using Natural Language Processing (NLP) and topic modeling in R to analyze over 6,500 responses to an open-ended question (OEQ) about imagining being 60. The focus is on exploring gender and social class differences in the responses. However, this automated analysis only provides a broad overview of frequently mentioned topics by different groups. In contrast, NVivo12 is highlighted as a more comprehensive tool for analyzing smaller samples (a few hundred responses). It allows researchers to identify and count specific codes like the use of the word "not" but also provides the flexibility to delve deeper into the data through qualitative or hermeneutic analysis. This approach enables the identification of strategies individuals use when responding to questions, and it allows for closer attention to the form and style of responses. One advantage of using OEQs is that they provide access to respondents' thoughts and ideas in their own words, allowing for a more subtle analysis of the specific vocabulary and phraseology used by participants. Such nuanced examination can reveal patterns and deeper insights.

In a recent study, Tang (2023) established that NVivo software is widely favored as a valuable tool for facilitating qualitative analyses. Specifically designed for exploring and categorizing text-based data, NVivo offers a range of features, including code-and-retrieve capabilities for conducting thematic analyses. Additionally, the software provides functions that enable researchers to establish connections between codes or categories of information, facilitating the construction of conceptual frameworks and even theories based on the data.

The summary of the literature overview in this study highlights the significance of NVivo software in qualitative research. NVivo offers a comprehensive set of tools to manage, organize, and analyze qualitative data, making it a popular choice among researchers. The software's user-friendly interface and various features enable in-depth exploration and understanding of text-based data. It supports multiple file formats, provides flexible coding systems, and facilitates the analysis of multimedia content. The literature also emphasizes the growing adoption of qualitative data analysis software and the potential for increased transparency and rigor in research processes. The focus of the research being reported is on exploring the impact of NVivo from the perspective of graduate students, offering unique insights tailored to this specific group of researchers.

Methodology

In this study, a qualitative research approach was utilized as a research strategy, incorporating in-depth interviews to gather comprehensive and precise responses in line with the research objectives. The interview protocol focused specifically on how graduate students perceive NVivo as a support tool and consisted of open-ended questions that allowed for detailed responses. The interviews were conducted in English, offering both in-person and remote options, and audio recordings were utilized to facilitate analysis. Furthermore, the study employed the documentary method to examine relevant survey questions from secondary data in order to obtain primary data results. To guide the selection of participants and shape the interview questions, the researchers established clear research objectives, questions, and topics. Purposive sampling was employed to select participants based on their relevant characteristics and experiences. In addition to the structured questions, the authors also incorporated open-ended questions into the study design. This deliberate choice allowed the participants to express their opinions more freely and in-depth, contributing to a richer and more comprehensive understanding of the research topic.

Prior to the main study, a pilot test was conducted with a small group of participants to ensure the clarity of the questions. The participants were presented with the interview questions and encouraged to provide feedback on their clarity, relevance, and appropriateness. This valuable feedback helped identify any potential ambiguities or misunderstandings in the questions. Based on the input received from the participants, adjustments and improvements were made to the wording and structure of the interview questions. Furthermore, the pilot test served to assess the participants' comfort level with the questions and the overall interview process, ensuring that they would feel at ease during the actual interviews. Informed consent was obtained from the participants, and the interviews were conducted in comfortable settings, respecting their preferences. Detailed notes or recordings were taken during the interviews, which were subsequently transcribed for analysis. Qualitative data analysis techniques, such as content analysis, were applied to identify patterns, themes, and relationships within the data. The interpretation of the findings took into account the research objectives, revealing significant insights and implications.

The interview questions were as follows:

- How has NVivo helped you in organizing and managing your qualitative data effectively?
- Can you share an example of how NVivo has streamlined your analysis processes and made them more efficient?
- In what ways has NVivo facilitated collaboration among team members and enhanced teamwork in your research projects?
- How have the visual tools in NVivo enhanced your data exploration and visualization process? Could you provide an example of how NVivo has enhanced your data exploration and visualization, leading to deeper insights?
- How has the use of NVivo ensured rigorous analysis in your qualitative research?
- Can you discuss how the integration of literature within NVivo has impacted the depth and quality of your research findings?
- How has NVivo increased your productivity in qualitative research tasks? Can you provide specific examples?
- What challenges have you faced in learning and utilizing NVivo effectively, and how have you overcome them?
- How have you addressed the challenge of maintaining consistency and accuracy in data coding and analysis using NVivo?
- How have you managed the cost considerations associated with acquiring an NVivo license for your research projects?

According to Francis et al. (2010) and Namey (2017), a minimum of six interviews appears to be the optimal number for achieving data saturation in qualitative research. Therefore, this study utilized purposive sampling, a recognized qualitative research technique, to select ten key informants based on the researchers' expertise, aligning with the objectives of the study. This sampling method aimed to gather comprehensive knowledge about a specific population or phenomenon of interest. The inclusion of graduate students in the study was based on their experience, expertise, and familiarity with the subject being investigated, ensuring valuable insights. The participants needed to fulfill three inclusion criteria: being at least 18 years old, being Thai graduate students studying in Bangkok, Thailand, and possessing recent knowledge and experience in using NVivo. The data from the interviews were collected in May 2023.

Ten graduate students were interviewed to gather their perceptions of the use of NVivo. Table 1 presents the information of the respondents, including their gender, age, and major, as well as the date and time of the interviews. The sample comprised ten graduate students, with an equal distribution of five males and five females. Their ages ranged from 28 to 48. There were two graduate students pursuing Ph.D. degrees in Communication Studies, one in Entrepreneurship, two in Education, two in Business Administration, one in Political Science, one in Finance, and one in International Business.

Table 1. Demographic information on the respondents and interview dates and times.

No.	Gender	Age	Major	Date and time of interview
R1	Male	29	Ph.D. in Communication Studies	May 02, 2023 at 09:30 am
R2	Male	32	Ph.D. in Entrepreneurship	May 02, 2023 at 10:00 am
R3	Male	47	Ph.D. in Education	May 02, 2023 at 10:30 am
R4	Male	31	Ph.D. in Business Administration	May 02, 2023 at 11:00 am
R5	Male	28	Ph.D. in Communication Studies	May 02, 2023 at 11:30 am
R6	Female	43	Ph.D. in Political Science	May 03, 2023 at 09:00 am
R7	Female	31	Ph.D. in Education	May 03, 2023 at 09:30 am
R8	Female	44	Ph.D. in Finance	May 03, 2023 at 10:00 am
R9	Female	36	Ph.D. in International Business	May 03, 2023 at 10:30 am
R10	Female	48	Ph.D. in Business Administration	May 04, 2023 at 11:00 am

The researchers employed content analysis for data analysis, a widely used method in qualitative research for identifying patterns, themes, and relationships within textual data. The data collected from the in-depth interviews were transcribed and carefully reviewed to extract meaningful insights. To facilitate the data analysis process and manage the extensive qualitative data, the researchers utilized NVivo, a powerful qualitative data analysis tool. NVivo enables efficient organization, categorization, and analysis of data, supporting a systematic and rigorous exploration of themes and patterns identified during content analysis. The combination of content analysis and NVivo ensured a robust and thorough examination of the data, leading to a deeper understanding of the participants' experiences, perceptions, and the impact of NVivo, a qualitative data analysis software, from the perspective of graduate students.

Results

In the study, participants' responses were analyzed to identify various themes. The findings revealed several benefits of using NVivo, such as efficient data organization and management, streamlined analysis, improved data exploration and visualization, collaboration and teamwork, integration of literature, and time-saving and increased productivity. Additionally, the study highlighted certain challenges associated with using NVivo, including a learning curve, difficulties in data management and organization, and considerations regarding cost. The interviews and analysis were carried out using content analysis and the NVivo software. To enhance the comprehensibility of the results, a word frequency query was employed, and the commonly used words by participants during the interviews were visually represented in a word cloud (Figure 1). A word cloud is a graphical representation of text data, where words are displayed in different sizes based on their frequency of occurrence in the text. Word clouds are often used to provide a quick visual summary of the most commonly used words in a given dataset.

Figure 1. Word cloud.

Benefits of using NVivo

NVivo, as perceived by graduate students, has had a significant impact on graduate students' research and overall academic experience in several ways, including efficient organization and management of data, streamlined data analysis, enhanced data exploration and visualization, collaboration and teamwork, literature integration, and time-saving and productivity. NVivo has had a transformative impact on graduate students' qualitative research endeavors. It has improved data organization, streamlined analysis processes, facilitated collaboration, enhanced data exploration and visualization, ensured rigorous analysis, integrated literature, and increased productivity. By leveraging the capabilities of NVivo, graduate students can conduct more rigorous and insightful qualitative research, leading to high-quality scholarly outputs.

Efficient organization and management of data

NVivo allows graduate students to store, organize, and manage their qualitative data in a systematic and efficient manner. It provides a centralized platform where they can import and organize various data sources such as interviews, surveys, documents, audio/video recordings, and literature. This makes it easier to access and retrieve specific data during analysis, saving valuable time and effort.

> NVivo allows us to store and organize our qualitative data in a really systematic way. We can import all our interviews, surveys, documents, audio/video recordings, and even literature into one centralized platform. It's so much easier to access and retrieve specific data during analysis (R1).
>
> With NVivo, you can create folders and categories to sort your data. You can tag and label specific sections of interviews or documents, and the software automatically links related information together. It saves so much time because you don't have to search through endless files and documents anymore (R3).
>
> NVivo has a range of tools for data analysis. It supports various coding techniques, so you can identify patterns and themes within your data. It also has visualizations like charts and graphs, which make it easier to understand complex relationships. And the best part is that it's all in one place, so you don't have to switch between different software or applications (R4).

Streamlined data analysis

NVivo is perceived by graduate students as a valuable tool that offers a range of features to facilitate data analysis. The software provides support for various coding techniques, particularly thematic coding, which empowers graduate students to identify patterns, themes, and relationships within their data. With easy-to-use annotation, highlighting, and data segment categorization functionalities, NVivo streamlines the analysis process, providing a more structured and systematic approach to research.

> NVivo can enhance the quality and rigor of qualitative analysis. It provides a structured framework for analyzing data, making it easier to draw meaningful conclusions and support your research findings. It's a valuable tool for graduate

> students like us who rely on qualitative research methods (R8).

> NVivo is great for qualitative data analysis. It has a bunch of tools and features that make the analysis process much easier. For example, it supports different coding techniques like thematic coding. It helps you identify patterns, themes, and relationships within your data (R9).

> NVivo also allows you to annotate and categorize your data segments. You can add notes, comments, or even your own interpretations right alongside the relevant sections. This helps in keeping track of your thought process and provides a comprehensive analysis (R10).

Enhanced data exploration and visualization

NVivo offers visual tools that help graduate students explore their data in meaningful ways. It enables the creation of charts, graphs, and diagrams to visualize patterns and trends within the data. These visual representations can aid in data interpretation and provide a clearer understanding of complex relationships and concepts.

> NVivo's visual tools allow you to see connections and relationships between different themes or concepts. You can also create diagrams to illustrate complex relationships or create visual summaries of your findings. It enhances your data interpretation process and helps communicate your research more effectively (R2).

> NVivo has some great visual tools that allow you to explore your data in meaningful ways. You can create charts, graphs, and diagrams to visualize patterns and trends within your qualitative data (R6).

Collaboration and teamwork

NVivo supports collaborative work, allowing graduate students to work together on a research project. It enables sharing and synchronization of project files, coding schemes, and annotations, facilitating teamwork and promoting a coordinated approach to qualitative analysis. This is particularly beneficial for interdisciplinary research projects or when working in research teams.

> NVivo allows multiple graduate students to work on the same research project simultaneously. It enables sharing and synchronization of project files, coding schemes, and annotations. It's great for promoting teamwork and maintaining a coordinated approach to qualitative analysis (R2).

> NVivo is really convenient! It makes collaboration much smoother. It's especially useful when working in research teams or interdisciplinary projects. We can share our coding schemes and annotations, allowing everyone to see and contribute to the analysis process. It fosters a coordinated approach, ensuring consistency and collaboration among team members (R5).

> Let's say we're working on a research project together. We can both access the project files and work on them simultaneously. Any changes or updates made by one person are automatically synced for everyone else. This eliminates the need to manually merge different versions of files or worry about conflicting changes (R9).

Literature integration

NVivo enables the integration of literature and external references within the software. Graduate students can link their qualitative data with relevant literature, enabling a comprehensive analysis that combines empirical evidence with existing theoretical frameworks. This integration fosters a deeper understanding of the research context and facilitates theoretical interpretations.

> NVivo allows you to integrate literature and external references directly into the software. You can link your qualitative data with relevant literature, which helps in combining empirical evidence with existing theoretical frameworks. It's really helpful for a comprehensive analysis (R1).

> NVivo enables you to make connections between your qualitative data and the literature you've reviewed. You can reference specific sections or concepts from the literature within your coding and analysis process. This integration helps in developing theoretical interpretations and supporting your research findings with existing scholarly work (R6).

> Let's say you have a collection of journal articles or book chapters that are relevant to your research. In NVivo, you can import and link those references to your qualitative data. This integration allows you to connect your empirical findings with existing theories and concepts, fostering a deeper understanding of the research context (R9).

Time-saving and productivity

NVivo automates many manual tasks involved in qualitative analysis, such as data organization, coding, and retrieval. This saves graduate students considerable time and effort, allowing them to focus on higher-order analysis and interpretation. The software's efficiency and productivity-enhancing features enable students to work more effectively, especially when dealing with large volumes of qualitative data.

> NVivo's automation allows us to focus on higher-order analysis and interpretation rather than getting caught up in the mundane tasks of organizing and coding data. It enhances our productivity and enables us to work more effectively, even with large volumes of qualitative data. By automating manual tasks, NVivo helps us work more efficiently

and effectively. It allows us to devote our energy to the critical thinking and analysis required in qualitative research (R6).

> NVivo's automation features extend to coding as well. It offers tools that allow you to code your data, which involves assigning labels or categories to different segments. The software can even suggest potential codes based on patterns it identifies in the data. This automation speeds up the coding process and makes it more efficient (R7).

> NVivo can automate many manual tasks involved in qualitative analysis. For example, it helps with data organization, coding, and retrieval. It saves us a lot of time and effort by streamlining these processes. Let's say you have a large volume of qualitative data, such as interview transcripts, documents, and audio recordings. NVivo provides a centralized platform where you can import and organize all of this data. It automatically indexes and tags the data, making it easier to retrieve specific information when you need it (R9).

Challenges associated with using NVivo

While NVivo offers numerous benefits to graduate students in qualitative research, there are also challenges associated with using the software.

Learning curve

NVivo can be complex and may have a steep learning curve for graduate students who are new to the software. It requires time and effort to become proficient in using its various features and functionalities. Students may need to invest time in training or self-learning to maximize the potential of NVivo.

> Learning NVivo does require an investment of time and effort. However, once you get the hang of it, it can really streamline your qualitative research process and save you time in the long run. One thing that helped me was taking advantage of the available resources. NVivo provides tutorials and online documentation that can guide you through the software's features and functionalities. You can also check if your university offers any workshops or training sessions on using NVivo (R1).

> Being patient and investing time to learn the software properly is crucial. It may seem overwhelming at first, but as I dedicate myself to training and self-learning, my proficiency in using NVivo will improve gradually. It's important to keep in mind that NVivo is a research tool designed to support us. Once we become familiar with its features, we can harness its full potential to enhance our qualitative analysis. The effort put into mastering NVivo will certainly be rewarding in the end (R7).

Challenges in data management and organization

While NVivo provides tools for efficient data organization, graduate students may still face challenges in managing and structuring their data effectively. Deciding on the appropriate coding structure, developing a coding scheme, and maintaining consistency across the project can be demanding tasks. Without careful planning and organization, students may encounter difficulties in retrieving and analyzing data later on.

> It's important to have a clear idea of how I want to organize and structure my data before diving into coding and analysis. Consistency is also something I worry about as well. My advisor has suggested that having a coding manual would certainly help me stay consistent in my coding decisions. I'll make sure to develop one for my project. Also, careful planning and organization are crucial when using NVivo. It seems like the initial investment of time and effort in designing the coding structure and maintaining consistency will pay off in the analysis phase (R8).

> I've been using NVivo for my qualitative research, and I must admit, data management and organization have been quite challenging. I find myself constantly questioning whether I've organized my data in the most efficient way. Sometimes I worry that I might have to redo the entire coding structure if I realize later on that it's not working well (R10).

Cost considerations

NVivo is commercial software that requires a license, and the cost can be a factor for graduate students with limited budgets. The availability of funding or institutional support for acquiring NVivo licenses may vary, and students may need to consider alternative options or explore open-source software alternatives for qualitative analysis.

> I've been considering NVivo for my qualitative research, but the cost is also a factor that worries me. The good news is that NVivo provides a trial version, allowing you to test the software before making a purchase. This is a valuable opportunity to explore its capabilities and determine if it aligns with your research requirements. Typically, the trial period for NVivo lasts between 14 to 30 days, depending on the version and any ongoing promotions (R3).

> NVivo is commercial software, and acquiring a license can be a significant expense, especially for students with limited funding. It's definitely an important factor to consider. There are a few alternatives to NVivo that you can explore. One option is to check if your university provides institutional licenses for NVivo. Some universities, like mine, have site licenses that allow students to access the software at no additional cost (R4).

Discussion

The study explored the impact of NVivo from the perspective of graduate students engaged in qualitative research. The findings indicated that NVivo has become a valuable asset in the toolkit of graduate students engaged in qualitative research. The software's ability to organize and analyze vast amounts of qualitative data enhances research efficiency, facilitates collaboration, and supports the production of high-quality research outcomes. However, it is essential to recognize the challenges associated with implementing NVivo and to provide the necessary resources and support to students, ensuring they can fully leverage its potential. As graduate students continue to embrace NVivo in their research endeavors, the impact of this software in qualitative research is expected to grow significantly.

The perception of NVivo by graduate students indicates that it has had a profound impact on their research and academic experience. Some of the key ways in which NVivo has influenced graduate students include efficient data organization and management, streamlined data analysis, improved data exploration and visualization, facilitated collaboration and teamwork, seamless integration of literature, and notable time-saving and increased productivity. The findings were consistent with several studies. For instance, Zamawe (2015) highlighted the benefits of using NVivo, emphasizing that the software provides a centralized location for all sources. Unlike manual coding, where researchers may struggle to locate specific papers or files, NVivo streamlines the retrieval process through links and organization within a single project. This feature enables researchers to save time and enhances efficiency. Additionally, NVivo offers flexibility in reshaping and reorganizing coding and node structures, allowing for quick adjustments as needed. Kraiwanit et al. (2023) concluded that NVivo has grown in popularity as a powerful tool for managing and analyzing large volumes of qualitative data, allowing researchers to delve deeper into complex datasets and gain valuable insights. In addition, Maher et al. (2018) conducted a study indicating that while digital analysis software packages like NVivo may not fully support the analysis process, they do offer excellent data management and retrieval capabilities that facilitate analysis and write-up. The research suggests that a combination of traditional tools such as colored pens, paper, and sticky notes for coding, along with digital software packages like NVivo for data management, can provide a valid and tested method for generating grounded theory. This approach acknowledges the benefits of both traditional and digital tools, harnessing their respective strengths to enhance the analysis process. Hoover and Koerber (2009) highlighted the practical enhancements in NVivo version 8 and the expanded capabilities in version 9, particularly regarding the increased exporting power of NVivo elements. These elements include transcripts, coding summaries, and reports of demographic data, which can now be exported to Word and HTML formats. This upgrade proved beneficial in facilitating collaboration with team members involved in infant-feeding research, even if they did not have direct access to the software. However, it is acknowledged that having access to NVivo for all team members would be ideal. The software offers several collaborative features and multiple options to accommodate different styles of collaboration, further supporting collaborative research efforts.

NVivo provides many advantages to graduate students in qualitative research, but it also comes with certain challenges. According to Davidson and Jacobs (2008), qualitative researchers are confronted with the need to adapt to the technological revolution, which entails learning and teaching qualitative data analysis software in higher education research courses. This shift presents a challenge as researchers grapple with incorporating these tools into their qualitative research methodologies. Hoover and Koerber (2009) acknowledged that despite the improvements in NVivo, the software presented initial challenges due to its steep learning curve. They candidly admitted that, like many others, they faced difficulties in understanding and effectively utilizing the software and are still in the process of overcoming these challenges. Moreover, Mitchell et al.'s (2007) research illuminated the unique challenges faced by psychology students as they engage with qualitative research methods. The study highlights the difficulties inherent in learning qualitative methods within a discipline that traditionally emphasizes quantitative approaches. Sanusi (2019) also confirmed that NVivo poses challenges for beginners. While NVivo is intended to support qualitative researchers and undergoes regular updates to align with research interests, it incorporates additional features. These features, such as nodes, classification, query, and others, may be unfamiliar to novice users. As a result, beginners may be reluctant to embrace the software and may opt for manual methods instead.

Conclusion

This study sheds light on the importance and benefits, as well as challenges, of NVivo in qualitative research and emphasizes the need for continued exploration, development, and support in order to maximize its potential and contribute to the advancement of the field.

The findings revealed that NVivo has become a valuable asset in the toolkit of graduate students, offering benefits such as enhanced research efficiency, facilitated collaboration, and support for producing high-quality research outcomes. By utilizing NVivo, graduate students were able to organize and analyze vast amounts of qualitative data, leading to streamlined research processes and deeper insights. However, the study also highlighted the importance of recognizing the challenges associated with implementing NVivo and providing the necessary resources and support to students to fully leverage its potential. As graduate students increasingly adopt NVivo in their research endeavors, its impact on qualitative research is expected to grow significantly. This software has the potential to contribute to valuable insights and discoveries, further advancing the field. Therefore, it is crucial to continue supporting and promoting the use of NVivo while addressing any associated challenges.

This study holds important implications for policy and practice in the field of graduate research. It highlights the significance of NVivo in qualitative research for graduate students, emphasizing the need for support and resources to maximize its potential. Policymakers and institutions can promote NVivo adoption by providing training programs and technical assistance to graduate students. Integrating NVivo training into research methodology courses can better equip students with essential data analysis skills. Moreover, fostering a collaborative research culture and supporting innovative approaches in qualitative data analysis using NVivo can advance qualitative research methodologies. Ethical considerations related to NVivo usage should also be emphasized. Overall, by implementing these implications, policymakers and institutions can create a conducive environment for graduate researchers to leverage NVivo's capabilities effectively, ultimately enhancing the quality of qualitative research.

Furthermore, this study also contributes to the broader literature on qualitative research by providing valuable insights into the impact of NVivo from the perspective of graduate students. It enhances our understanding of how qualitative data analysis software enhances research efficiency, facilitates collaboration, and addresses implementation challenges. These implications inform best practices for utilizing NVivo effectively, advancing qualitative research methodologies, and promoting the use of technological tools in qualitative research practice.

While the study provides valuable insights into the impact of NVivo from the perspective of graduate students, it is important to acknowledge its limitations and make recommendations for future studies to build upon this research. One limitation of the study is the small sample size. Interviewing only ten graduate students through purposive sampling may limit the generalizability of the findings. Future studies could consider expanding the sample size and employing diverse sampling techniques to include a broader range of graduate students from different disciplines and institutions. This would provide a more comprehensive understanding of the impact of NVivo across various contexts. Additionally, the study focused primarily on the benefits and challenges associated with NVivo. While this is valuable information, future research could explore the specific research questions or methodologies that NVivo is most useful for. Investigating the ways in which NVivo supports specific qualitative research approaches or addresses specific research questions would provide deeper insights into the software’s potential applications. Lastly, the study primarily focused on the perspectives of graduate students. Future research could consider incorporating the views of other stakeholders, such as research advisors, faculty members, or research collaborators, to gain a more comprehensive understanding of the broader impact of NVivo on qualitative research practices. These recommendations would contribute to a more nuanced understanding of the impact of NVivo and its potential implications for qualitative research.

References

Allsop, D. B., Chelladurai, J. M., Kimball, E. R., Marks, L. D., & Hendricks, J. J. (2022). Qualitative methods with Nvivo software: A practical guide for analyzing qualitative data. *Psych, 4*(2), 142-159. https://doi.org/10.3390/psych4020013

Austin, Z., & Sutton, J. (2014). Qualitative research: Getting started. *The Canadian Journal of Hospital Pharmacy, 67*(6), 436-440. https://doi.org/10.4212/cjhp.v67i6.1406

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open, 2*, 8-14. https://doi.org/10.1016/j.npls.2016.01.001

Busetto, L., Wick, W., & Gumbinger, C. (2020). How to use and assess qualitative research methods. *Neurological Research and Practice, 2*, 1-10. https://doi.org/10.1186/s42466-020-00059-z

Collins, C. S., & Stockton, C. M. (2018). The central role of theory in qualitative research. *International Journal of Qualitative Methods, 17*(1). https://doi.org/10.1177/1609406918797475

Dalkin, S., Forster, N., Hodgson, P., Lhussier, M., & Carr, S. M. (2021). Using computer assisted qualitative data analysis software (CAQDAS; NVivo) to assist in the complex process of realist theory generation, refinement and testing. *International Journal of Social Research Methodology, 24*(1), 123-134. https://doi.org/10.1080/13645579.2020.1803528

Davidson, J., & Jacobs, C. (2008). The implications of qualitative-research software for doctoral work considering the individual and institutional context. *Qualitative Research Journal, 8*(2), 73-80. https://doi.org/10.3316/QRJ0802072

Dhakal, K. (2022). NVivo. *Journal of the Medical Library Association, 110*(2), 270-272. https://doi.org/10.5195/jmla.2022.1271

Dollah, S., Abduh, A., & Rosmaladewi, M. (2017, September). Benefits and drawbacks of NVivo QSR application. In: *2nd International conference on education, science, and technology* (pp. 61-63). Atlantis Press. https://doi.org/10.2991/icest-17.2017.21

Elliott, J. (2022). The craft of using NVivo12 to analyze open-ended questions: An approach to mixed methods analysis. *The Qualitative Report, 27*(6), 1673-1687. https://doi.org/10.46743/2160-3715/2022.5460

Fearnley, C. J. (2022). Mind mapping in qualitative data analysis: Managing interview data in interdisciplinary and multi-sited research projects. *Geo: Geography and Environment, 9*(1), e00109. https://doi.org/10.1002/geo2.109

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology and Health, 25*(10), 1229-1245. https://doi.org/10.1080/08870440903194015

Hoover, R. S., & Koerber, A. L. (2009). Using NVivo to answer the challenges of qualitative research in professional communication: Benefits and best practices tutorial. *IEEE Transactions on Professional Communication, 54*(1), 68-82. https://doi.org/10.1109/TPC.2009.2036896

Hunt, M. R., Mehta, A., & Chan, L. S. (2009). Learning to think qualitatively: Experiences of graduate students conducting qualitative health research. *International Journal of Qualitative Methods, 8*(2), 129-135. https://doi.org/10.1177/160940690900800204

Jackson, K., & Bazeley, P. (2019). *Qualitative data analysis with NVivo* (3rd ed.). Sage Publications. https://hdl.handle.net/1959.7/uws:55501

Johnston, L. (2006). Software and method: Reflections on teaching and using QSR NVivo in doctoral research. *International Journal of Social Research Methodology, 9*(5), 379-391. https://doi.org/10.1080/13645570600659433

Kaefer, F., Roper, J., & Sinha, P. N. (2015). A software-assisted qualitative content analysis of news articles: Examples and reflections. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 16*(2), Art. 8, http://nbn-resolving.de/urn:nbn:de:0114-fqs150283

Kent State University. (2023, April 25). *Statistical & qualitative data analysis software: About NVivo.* University Library. https://libguides.library.kent.edu/statconsulting/NVivo

Köhler, T., Smith, A., & Bhakoo, V. (2022). Templates in qualitative research methods: Origins, limitations, and new directions. *Organizational Research Methods, 25*(2), 183-210. https://doi.org/10.1177/10944281211060710

Kraiwanit, T., Limna, P., & Siripipatthanakul, S. (2023). NVivo for Social Sciences and Management Studies: A systematic review. *Advance Knowledge for Executives, 2*(3), 1-11. https://ssrn.com/abstract=4523829

Lund, B. D., Wang, T., Shamsi, A., Abdullahi, J., Awojobi, E. A., Borgohain, D. J., Bueno de la Fuente, G., Huerta, G. P., Isfandyari-Moghaddam, A., Islam, M. A., & Khasseh, A. A. (2023). Barriers to scholarly publishing among library and information science researchers. *International perspectives. Information Development, 39*(2), 376-389. https://doi.org/10.1177/02666669211052522

Maher, C., Hadfield, M., Hutchings, M., & De Eyto, A. (2018). Ensuring rigor in qualitative data analysis: A design research approach to coding combining NVivo with traditional material methods. *International Journal of Qualitative Methods, 17*(1), 1609406918786362. https://doi.org/10.1177/1609406918786362

Mitchell, T., Friesen, M., Friesen, D., & Rose, R. (2007). Learning against the grain: Reflections on the challenges and revelations of studying qualitative research methods in an undergraduate psychology course. *Qualitative Research in Psychology, 4*(3), 227-240. https://doi.org/10.1080/14780880701473441

Mohajan, H. K. (2018). Qualitative research methodology in social sciences and related subjects. *Journal of Economic Development, Environment and People, 7*(1), 23-48. https://doi.org/10.26458/jedep.v7i1.571

Mulisa, F. (2022). When does a researcher choose a quantitative, qualitative, or mixed research approach?. *Interchange, 53*(1), 113-131. https://doi.org/10.1007/s10780-021-09447-z

Namey, E. (2017, April 25). Riddle me this: How many interviews (or focus groups) are enough?. *R&E Research for Evidence.* https://researchforevidence.fhi360.org/riddle-me-this-how-many-interviews-or-focus-groups-are-enough

Niedbalski, J., & Ślęzak, I. (2023, January). NVivo as a tool for supporting teamwork in the context of qualitative research conducted remotely-opportunities, limitations, and practical tips. In: *World conference on qualitative research* (pp. 38-59). Springer International Publishing. https://doi.org/10.1007/978-3-031-31346-2_3

NVivo. (n.d.) NVivo10 for Windows: Getting started. *QSR International.* https://download.qsrinternational.com/Document/NVivo10/NVivo10-Getting-Started-Guide.pdf

O'Donoghue, T. (2022). Is conducting interpretive studies within mixed methods research projects justified? Methinks not. *Journal of Applied Learning and Teaching, 5*(2), 14-19. https://doi.org/10.37074/jalt.2022.5.2.14

Robins, C. S., & Eisen, K. (2017). Strategies for the effective use of NVivo in a large-scale study: Qualitative analysis and the repeal of don't ask, don't tell. *Qualitative Inquiry, 23*(10), 768-778. https://doi.org/10.1177/1077800417731089

Sanusi, I. (2019). *Strengths and weaknesses of using NVivo software in analyzing qualitative data* [Doctoral dissertation, Universitas Islam Negeri (UIN) Ar-Raniry]. https://repository.ar-raniry.ac.id/id/eprint/15044/

Sotiriadou, P., Brouwers, J., & Le, T. A. (2014). Choosing a qualitative data analysis tool: A comparison of NVivo and Leximancer. *Annals of Leisure Research, 17(*2), 218-234. https://doi.org/10.1080/11745398.2014.902292

Taherdoost, H. (2021). Data collection methods and tools for research: A step-by-step guide to choose data collection technique for academic and business research projects. *International Journal of Academic Research in Management, 10*(1), 10-38. https://hal.science/hal-03741847/

Tang, R. (2023). Harnessing insights with NVivo. In J. M. Okoko., S. Tunison,. & K. D. Walker. (Eds.), *Varieties of qualitative research methods.* Springer Texts in Education. Springer, Cham. https://doi.org/10.1007/978-3-031-04394-9_34

Tenny, S., Brannan, J. M., & Brannan, G. D. (2022). *Qualitative study.* StatPearls Publishing. https://pubmed.ncbi.nlm.nih.gov/29262162/

Tilley, S. (2019). The role of critical qualitative research in educational contexts: A Canadian perspective. *Educar em Revista, 35*, 155-180. https://doi.org/10.1590/0104-4060.66806

Tomaszewski, L. E., Zarestky, J., & Gonzalez, E. (2020). Planning qualitative research: Design and decision making for new researchers. *International Journal of Qualitative Methods, 19*. https://doi.org/10.1177/1609406920967174

University of Illinois. (2023, March 6). *Text mining tools and methods: About NVivo.* University of Illinois. https://guides.library.illinois.edu/textmining/nvivo

Woods, M., Paulus, T., Atkins, D. P., & Macklin, R. (2016). Advancing qualitative research using qualitative data analysis software (QDAS)? Reviewing potential versus practice in published studies using ATLAS. ti and NVivo, 1994–2013. *Social Science Computer Review, 34*(5), 597-617. https://doi.org/10.1177/0894439315596311

Zamawe, F. C. (2015). The implication of using NVivo software in qualitative data analysis: Evidence-based reflections. *Malawi Medical Journal, 27*(1), 13-15. https://doi.org/10.4314/mmj.v27i1.4

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Examining Unified Theory of Acceptance and Use of Technology 2 through Meta-analytic Structural Equation Modelling

Caleb Or[A]

[A] *Academic Director, Total English Learning Global*

Keywords

MASEM;
meta-analytic structural equation modelling;
one-stage meta-analytic structural equation modelling;
OSMASEM;
Unified Theory of Acceptance and Use of Technology 2;
UTAUT2.

Correspondence

caleb.chin.poh.or@gmail.com [A]

Article Info

Received 17 March 2023
Received in revised form 18 June 2023
Accepted 12 July 2023
Available online 14 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.7

Abstract

Like its predecessor, the Unified Theory of Technology and Use of Technology (UTAUT), UTAUT2 has been adopted, adapted and used in extended forms due to its simplicity, parsimony and robustness. This study synthesised 39 empirical studies based on the UTAUT2 model in educational contexts, using the One-stage Meta-Analysis and Structural Equation Modelling (OSMASEM). Although the findings in this study aligned with the initial findings by Ventakesh et al. (2012), the model did not perform well compared to those in the initial UTAUT2 study in the explained variance in both behavioural intention and use behaviour. When new relationships were introduced into the UTAUT2 model in this study, constructs like performance expectancy, hedonic motivation, social influence, and price value were new predictors of use behaviour. The meta-analytic structural equation modelling approach used in this study, OSMASEM, allows researchers to use past empirical study data to examine the UTAUT2 framework without replicating similar studies. Using OSMASEM, researchers could easily add past empirical data to train the UTAUT2 model to study the trends in technology acceptance and use in educational contexts.

Introduction

Recent research has attempted to examine technology acceptance through meta-analytic approaches (Feng et al., 2021; Leong et al., 2022; Jeyaraj & Dwivedi, 2020; Mishra et al., 2023; Than et al., 2021; Zaremohzzabieh et al., 2022). Meta-analytic structural equation modelling (MASEM) is a powerful mechanism for synthesising prior research findings, reconciling inconsistent conclusions, and resolving variable relationships (Cheung, 2014; Jeyaraj & Dwivedi, 2020; Viswesvaran & One, 1995). The advantage of using the MASEM is that it can test models that involve variables not included in the primary studies (Bergh et al., 2016; Steinmetz & Block, 2022). This approach combines the strengths of meta-analysis, which quantitatively summarises the results of individual studies and structural equation modelling. MASEM is a widely used statistical technique in educational research for synthesising and integrating data from multiple studies because of its ability to synthesise data from multiple studies and estimate a weighted average effect size, which measures the strength of the relationship between two variables (Cheung, 2019; Furlow & Beretvas, 2010; Herhausen et al., 2021; Raeisi-Vanani et al., 2022). It allows researchers to overcome the limitations of individual studies and arrive at a more comprehensive and robust understanding of the relationship between educational variables and outcomes.

MASEM can be used in studies that examine the adoption and usage of technology in organisations, such as those based on popular models like the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) (Ventakesh et al., 2012). In UTAUT2 studies, MASEM can be used to synthesise data from multiple studies to understand the relationships between the factors proposed in the UTAUT2 model and the adoption and usage of technology. For example, MASEM can estimate each factor's weighted average effect size on the adoption and usage of technology, allowing researchers to determine which factors impact technology adoption and usage.

In recent years, structural equation modelling is gaining popularity as one of the meta-analyses methods (Jak & Cheung, 2020; Steel et al., 2021; Wilson et al., 2016). Tang and Cheung (2016) demonstrated that researchers could benefit from MASEM by introducing a two-stage meta-analytic structural equation modelling (TSSEM) using R packages such as metaSEM, while Jak et al. (2021) developed a one-stage MASEM (OSMASEM) for random-effects models. OSMASEM is a specific approach to MASEM where all of the data from multiple studies is combined in a single analysis rather than conducting separate meta-analyses for each moderator variable or each dependent variable, providing an advantage over traditional meta-analyses. While TSSEM (Tang & Cheung, 2016) and OSMASEM (Jak et al., 2021) gathered traction, such approaches were not commonly used in UTAUT2 studies. UTAUT2 studies could benefit significantly from the OSMASEM approach as it allows researchers to synthesise and cumulate research findings into a single effect size (Bergh et al., 2016). The effect size reflects the magnitude and directionality of the association between the two or more UTAUT2 variables. OSMASEM can also provide information on the degree of fit of the entire UTAUT model and can handle samples with missing correlations (Cheung & Cheung, 2016). As such, this study aims to utilise OSMASEM to synthesise past UTAUT2 research data and examine their findings from 2013 to 2022.

Literature review

UTAUT2

UTAUT2 was developed later to tailor to the consumer acceptance and use of technology. There were three critical features in UTAUT2: (1) the introduction of hedonic motivation (HM), price value (PV) and habit (H) as additional factors in consumer products and technology use; (2) some existing relationships were changed in the original UTAUT model; and (3) introduction of new relationships (Venkatesh et al., 2012) (Figure 1). According to Venkatesh et al. (2012), the effect of HM on BI is moderated by age, gender, and experience. The effect of PV on BI is moderated by age and gender. H has direct and mediated effects on UB, and individual differences moderate these effects.

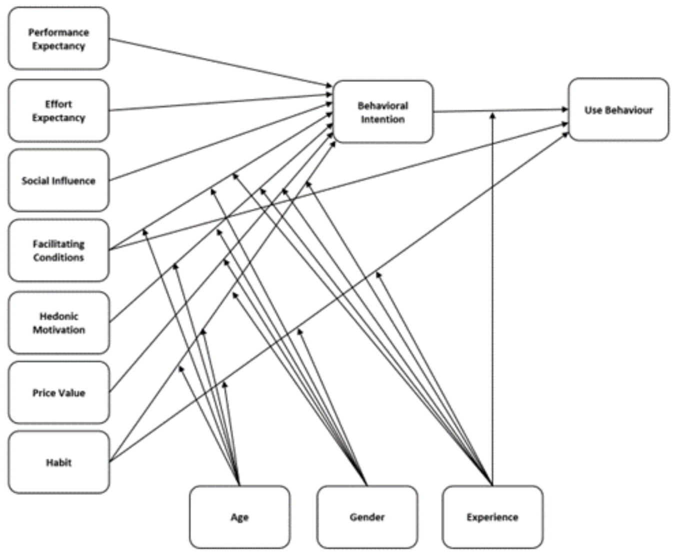


Figure 1: UTAUT2. Note: Adapted from Venkatesh et al. (2012).

UTAUT2 is considered the most comprehensive model in information system and technology adoption research (Tamilmani et al., 2017; Tamilmani et al., 2021). The model has been used in many past studies to examine factors influencing technology acceptance. For instance, Goto and Munyai (2022) utilised UTAUT2 to examine factors affecting law students' acceptance and use of online learning, while Avci and Avci (2022) examined the factors affecting teachers' use of digital learning resources.

As in UTAUT, Venkatesh et al. (2012) posited that PE was a predictor of BI, and the proposition remains constant in later empirical studies utilising UTAUT2. For instance, Hu et al. (2020), in their UTAUT2 study with 638 academic staff that explored factors affecting the adoption of emerging mobile technologies, revealed that PE remained a predictor of BI. Similarly, Jung & Lee (2020) found that PE was a predictor of BI in their cross-cultural study examining the adoption of open educational resources with 152 Korean and Japanese educators.

Similar to the UTAUT findings, the empirical results from UTAUT2 studies with EE as a predictor of BI have been inconsistent. Some studies showed that EE did not significantly affect BI. For instance, in the study with 206 undergraduates on the acceptance of Google Classroom, Kumar and Bervell (2019) found that EE was not a predictor of BI. In a similar research on the acceptance of Google Classroom with 163 students, Bervell et al. (2021) found that EE had a significant effect on SI instead of BI. De Moraes and Cabello (2017), in their study on the use of educational applications by 133 Brazilian students, revealed that EE has no significant effect on BI.

Based on the literature, SI was posited to be a predictor of BI. In many later UTAUT2 studies, it was found that SI continued to have a significant effect on BI (Ashraf et al., 2023; Aziz et al., 2020; Fathima Sanjeetha & Sabraz Nawaz, 2020; Goto & Munyai, 2022; Moorthy et al., 2019a; Raman & Don, 2013; Raman & Thabbimalai, 2021; Tseng et al., 2019).

One of the critical features of UTAUT2 is the change of some existing relationships in the original UTAUT model (Venkatesh et al., 2012). In the original UTAUT model, FC is posited to be a predictor of UB (Venkatesh et al., 2003). However, in the UTAUT2 model, FC is posited to predict both BI and UB (Venkatesh et al., 2012). FC remained a predictor of BI in many later UTAUT2 studies (Arain et al., 2018; Azizi et al., 2020; Bhimasta & Suprapto, 2016; El-Masri & Tarhini, 2017; Faqih & Jaradat, 2021; Farooq et al., 2017; Fathima Sanjeetha & Sabraz Nawaz, 2020; Gengfu & Chotiyaputta, 2019; Gunawan et al., 2019; Hu et al., 2020; Kalinkara & Talan, 2022; Meet et al., 2022; Mishra et al., 2022; Raman & Don, 2013; Rudhumbu, 2022; Tseng et al., 2019; Widjaja et al., 2020; Zacharis & Nikolopoulou, 2022). The discussion on FC as a predictor of UB is sometimes not straightforward as in many studies. UB was often omitted in many UTAUT2 empirical studies (Abdul Rabu et al., 2019; Al-Azawei & Alowayr, 2020; Almahri et al., 2020; Arain et al., 2018; Arain et al., 2019; Bhimasta & Suprapto; 2016; de Moraes & Cabello, 2017; El-Masri & Tarhini, 2017; Faqih & Jaradat, 2021; Gengfu & Chotiyaputta, 2019; Gunawan et al., 2019; Jung & Lee; 2020; Kaur et al., 2021; Le et al., 2022; Meet et al., 2022; Mehta et al., 2019; Moorthy et al., 2019a; Moorthy et al., 2019b; Rudhumbu, 2022; Sharif et al., 2019; Xu et al., 2022). For studies that included UB as a construct, in most cases, findings revealed that FC was a predictor of UB (Ain et al., 2016; Ashraf et al., 2023; Cao & Nguyen, 2022; Goto & Munyai, 2022; Hu et al., 2020; Kalinkara & Talan, 2022; Musa et al., 2022; Nikolopoulou et al., 2020; Raman & Don, 2013; Tseng et al., 2019; b et al., 2020; Zhou et al., 2020; Zawain, 2019; Zawin & Haboobi, 2019).

HM is the fun or pleasure of using a system or technology (Brown & Venkatesh, 2005). HM has been included as a critical predictor in past consumer behaviour research and prior information system research in the consumer technology use context (Brown & Venkatesh, 2005; Holbrook & Hirschman, 1982). In information system research, HM has been found to influence technology acceptance and use (Childers et al., 2001; Thong et al., 2006; Van der Heijden, 2004). From the literature, HM is generally a predictor of BI, a finding that is aligned with what was proposed by Venkatesh et al. (2012) (Ashraf et al., 2023; Avci & Acvi, 2022, Azizi et al., 2020; Bervell et al., 2021; de Moraes & Cabello, 2017; Fathima Sanjeetha & Sabraz Nawaz, 2020; Hu et al., 2020, Kalinkara & Talan, 2022; Kumar & Bervell, 2019; Moorthy et al., 2019b; Nikolopoulou et al., 2020; Raman & Don, 2013; Zhou et al., 2022). However, when Tamilmani et al. (2019) conducted a meta-analysis of 79 UTAUT2 studies, the researchers found that only 46 (58%) of the studies utilised HM as a construct, while 33 studies (42%) omitted the construct.

Venkatesh et al. (2012) extended the original UTAUT to examine the use of information technology in consumer contexts. Hence, in UTAUT2, PV is crucial as consumers have to bear the costs associated with purchasing devices and services. Consumer behaviour research has included cost-related constructs to explain consumers' actions (Dodds et al., 1991). In marketing research, PV is conceptualised with the quality of products and services to determine their perceived value (Zeithaml, 1988).

While adding PV as a construct may set UTAUT2 apart from the original UTAUT2, many later studies did not include it as part of the latter model. Tamilmani et al. (2018a) conducted a meta-analysis on 79 UTAUT2 empirical studies and found that only 32 studies (41%) utilised PV, while 47 studies (59%) omitted the construct from their research models. The main argument for excluding PV as a construct in their UTAUT2 models was that the technology involved in the studies was free of cost, like mobile applications and social networking sites. Among the 47 studies examined, only 4 were in the educational contexts examining LMS, informal learning and podcasting (Lai et al., 2016; Lin et al., 2013; Raman & Don, 2013). The researchers recommended that PV be a key predictor of individual technology adoption with UTAUT2. In other words, for utilising the UTAUT2 model for studies, PV should be one of the essential constructs in future research. Or (2023a) argued that since past studies had shown that PV had no significant effect on BI when examining technologies that were free of charge, it was recommended that the original UTAUT model be adopted or extended with added constructs instead of citing it as UTAUT2 research. For studies that included PV as a construct, it has been found that PV was a predictor of BI (Azizi et al., 2020; Farooq et al., 2017; Gengfu & Chotiyaputta, 2019; Meet et al., 2022; Mehta et al., 2019; Moorthy et al., 2019b; Tseng et al., 2019; Xu et al., 2022).

H is critical in predicting technology use (Kim & Malhotra, 2005; Kim et al., 2005; Limayem et al., 2007). It is defined as the degree to which individuals tend to perform behaviours automatically because of learning (Limayem et al., 2007), while Kim et al. (2005) equate H with automaticity. In other words, H is viewed as prior behaviour measured as the extent to which an individual believes the behaviour to be automatic (Kim & Malhotra 2005; Limayem et al. 2007). Tamilmani et al. (2018b) discovered in their systematic review that out of 66 empirical studies that utilised UTAUt2, only 23 (35%) included H as a construct in the studies. They recommended that researchers studying the initial stages of technology adoption in mandatory user settings should refrain from using H as a construct. On the other hand, using H as a construct is encouraged in research to examine established technologies driven by intrinsic consumer motivation. From the literature, H was generally found to have a significant effect on BI (Almahri et al., 2020; Ashraf et al., 2023; Avci & Avci, 2022; Azizi et al., 2020; de Moraes et al., 2017; Fathima Senjeetha & Sabraz Nawaz, 2020; Hu et al., 2020; Jung & Lee, 2020; Malešević et al., 2021; Mishra et al., 2021; Moorthy et al., 2019; Nikolopoulou et al., 2020; Raman & Thannimalai, 2021, Zhou et al., 2022) and UB (Avci & Avci; 2022; Azizi et al., 2020, Hu et al., 2020; Malešević et al., 2021; Nikolopoulou et al., 2020).

The current UTAUT2 study using OSMASEM

The current study synthesised 39 empirical research on UTAUT2 in educational contexts and capitalised on the advantage of synthesising correlation matrices through correlation-based OSMASEM (Jak et al., 2021). The current UTAUT2 study addresses these research questions:

1. To what degree do pooled correlation matrix relationships among the constructs show significant variations in UTAUT2 empirical studies from 2013 to 2022 using the OSMASEM approach?

2. To what degree does the UTAUT2 model fit the data from a pooled correlation matrix using the OSMASEM?

3. Are there new direct relationships among the UTAUT2 constructs found using the OSMASEM?

Method

Literature search and screening procedures

The Google Scholar database was searched to identify the relevant literature to the current UTAUT2 study. The following search terms and Boolean operators were used, "UTAUT2" AND "education". The other advanced search settings were included "anywhere in the articles" and "return articles dated between 2013 and 2023." After the search, an initial screening of the identified 10,900 studies was performed based on the following criteria: (1) the studies must address school or university's technology acceptance; (2) the studies must describe the relationships between the UTAUT2 constructs; and (3) the studies must analyse, report and discuss the findings in English. The initial screening resulted in 1,130 eligible empirical studies. Some studies were then excluded by applying the following criteria: (1) the studies did not target teachers, lecturers, educators or students in K-12, college or university education; (2) the studies were not based on the UTAUT2 model, but the UTAUT model. Past research using the OSMASEM approach had been conducted previously (Or, 2023a); (3) the studies had insufficient statistical reporting of the correlations between UTAUT2 constructs; (4) correlations between variables were negative where R package, metaSEM, is unable to compute; and (5) UTAUT2 was examined outside of educational contexts. Figure 2 summarises the results of the literature search and screening procedures. Table 1 lists the various research from which the data is used in this OSMASEM study.

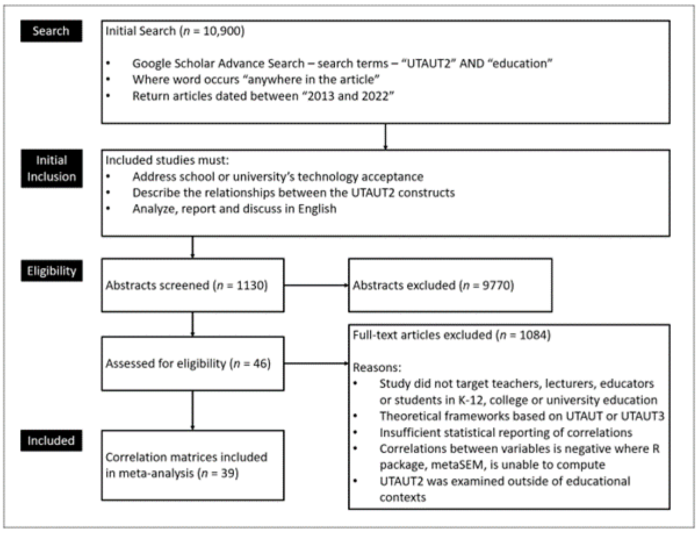


Figure 2. Diagram describing the literature search and the selection of eligible studies for meta-analysis.

Table 1. UTAUT2 studies from which data are used.

S/N	Technology / System	Sample Size	UTAUT2 Constructs	Study
1	AR	281	PE; EE; SI; FC; HM; PV; BI	Faqih & Jaradat (2021).
2	Blended Learning	203	PE; SI; HM; BI	Kaur et al. (2021).
3	Blended Learning	432	PE; EE; SI; FC; HM; PV; H; BI	Rudhumbu. (2022).
4	Chatbots	431	PE; EE; SI; FC; HM; H; BI	Almahri et al. (2020, March).
5	Digital Learning Resources	355	PE; EE; SI; FC; HM; PV; H; BI; UB	Avci & Avci. (2022).
6	Digital Library	483	PE; EE; SI; FC; HM; H; BI	Moorthy et al. (2019).
7	Distance Learning System	208	PE; EE; SI; FC; HM; H; BI; UB	Kalinkara & Talan. (2022).
8	e-Books	326	PE; EE; SI; FC; HM; PV; H; BI	Bhimasta & Suprapto. (2016, November).
9	e-Books	257	PE; EE; SI; FC; HM; PV; H; BI	Gengfu & Chotiyaputta. (2019).
10	e-Learning	418 (American) 389 (Qatari)	PE; EE; SI; FC; HM; PV; H; BI	El-Masri & Tarhini. (2017).
11	e-Learning	197	PE; EE; SI; FC; HM; PV; H; BI; UB	Goto, J & Munyai. (2022).
12	e-Learning	160	PE; EE; SI; HM; PV; H; BI	Mehta et al. (2019).
13	e-Learning	1024	PE; EE; FC; PV; H; BI; UB	Osei et al. (2022).
14	e-Learning	159	PE; EE; SI; FC; HM; PV; H; BI	Raman & Thannimalai. (2021).
15	e-Learning	314	PE; EE; SI; FC; HM; H; BI; UB	Zacharis & Nikolopoulou. (2022).
16	e-Services	173	PE; EE; SI; FC; H; BI; UB	Malešević et al. (2021).
17	Google Classroom	163	PE; EE; SI; FC; HM; PV; H; BI; UB	Bervell et al. (2022).
18	Google Classroom	163	PE; EE; SI; FC; HM; PV; H; BI; UB	Kumar & Bervell. (2019).
19	Interactive Whiteboards	171	PE; EE; SI; FC; HM; H; BI; UB	Zhou et al. (2022).
20	Learning Management System	320	PE; EE; SI; FC; HM; H; BI; UB	Raman & Don. (2013).

S/N	Technology / System	Sample Size	UTAUT2 Constructs	Study
21	Learning Managemen t System	178	PE; EE; SI; FC; HM; PV; H; BI	Sharif et al. (2019).
22	Learning Managemen t System	228	PE; EE; SI; FC; HM; H; BI; UB	Zwain. (2019).
23	Learning Managemen t System	113	PE; EE; SI; FC; HM; H; BI; UB	Zwain & Haboobi. (2019).
24	Lecture Capture System	481	PE; EE; SI; FC; HM; PV; H; BI; UB	Farooq et al. (2017).
25	Micro-lectures	161	PE; EE; SI; FC; HM; PV; BI; UB	Tseng et al. (2022).
26	Mobile Internet	262	PE; EE; SI; FC; HM; PV; H; BI; UB	Nikolopoulou et al. (2021).
27	Mobile Learning	469	PE; EE; SI; HM; PV; BI	Al-Azawei & Alowayr. (2020).
28	Mobile Learning	730	PE; EE; SI; FC; HM; H; BI	Arain et al. (2019).
29	Mobile Learning	730	PE; EE; SI; FC; HM; H; BI	Arain et al. (2018).
30	Mobile Technologi es	152	PE; EE; SI; FC; HM; PV; H; BI	Jung & Lee. (2020).
31	MOOCs	321	PE; EE; SI; FC; HM; PV; BI	Wijaya & Weinhandl. (2022).
32	Online Learning	376	PE; EE; SI; FC; HM; H; BI; UB	Musa et al. (2022).
33	Online Learning	566	PE; EE; SI; FC; HM; H; BI	Xu et al. (2022).
34	QR Code	200	SI; FC; HM; BI	Abdul Rabu et al. (2019).
35	Smartphone	831	PE; EE; SI; FC; HM; H; BI; UB	Cao & Nguyen. (2022).
36	Smartphone	540	PE; EE; SI; FC; HM; PV; H; BI: UB	Nikolopoulou et al. (2020).
37	Social Media	291	PE; EE; SI; FC; HM; H; BI; UB	Fathima Sanjeetha & Sabraz Nawaz. (2020).
38	Social Media	315	PE; EE; SI; FC; HM; H; BI; UB	Mishra et al. (2022).

Internal structure

R Studio (version 2022.12.0, Build 353) and its metaSEM package (version 1.3.0) were used to examine the fit of Model 1. The analysis examined whether the actual factor structure and loadings aligned with the theorised structure. It is done by statistically testing the fit between the proposed measurement model and the observed correlations (Albright & Park, 2009; Bollen, 1989; Hair et al., 2006; Kline, 2005). The following indices were used to assess the fit of Model 1 to the data: (a) *χ2/* Degree of Freedom *χ2/df*), (b) Root Mean Square Error of Approximation (RMSEA) (Steiger, 1990), (c) Standardised Root Mean Square Residual (SRMR), (d) Comparative Fit Index (CFI) (Bentler, 1990) and (e) Tucker-Lewis fit index (TLI; Bentler & Bonett, 1980) (Table 2). The values for the UTAUT2 model were within the recommended thresholds for acceptable model fit based on all five indices (*χ2/df* = 2.062; RMSEA = .008; SRMR = .026; CFI = 1.000, TLI = .984) (Table 2). The data reliability was analysed using IBM SPSS (version 28.0.1.1) and was highly reliable (N = 39; α = .993).

Table 2. Goodness-of-fit indices of Model 1.

Measure	Threshold	Value
χ^2	--	10.886
Df	--	5.000
χ^2/df	< 3.000	2.172
p-value	> .050	.053
RMSEA	< .050	.009
SRMR	<.080	.002
CFI	> .950	.997
TLI	> .950	.981

The correlation matrices obtained from the 39 UTAUT2 studies were analysed with the R package, metaSEM (version 1.3.0). With the R software, the metaSEM package derived originally from the openMX package provides analysis for the OSMASEM method using the SEM approach. The OSMASEM approach, most suitable for processing longitudinal relationships between variables at continuous time points (Cheung, 2014), was a good fit for this study that extracted empirical studies from the last decade, 2013 to 2022. Furthermore, the metaSEM package increased the sensitivity of significance tests by utilising the maximum likelihood estimation for analyses and used the sum rather than the average of sample sizes to compute the standard errors for the path coefficients.

Model 1 in this current meta-analysis underperformed as compared to the original model by Venkatesh et al. (2012). The original UTAUT2 model performed at an adjusted R2 of 74% for BI. The UTAUT2 model in this study only attained an R2 of 53.6%. For the explained variance of UB, Model 1 also underperformed compared to the original UTUAT2 model at R2 of 48.4% (Table 3). The original UTAUT model attained an explained variance at 52% for UB.

Table 3. Comparison of variances explained.

Measure	Threshold	Value
χ^2	--	10.886
Df	--	5.000
χ^2/df	< 3.000	2.172
p-value	> .050	.053

Like the original UTAUT2 model proposed by Ventakesh et al. (2012), H remained the best predictor of BI (β= .250; *p*<.001) compared to PE, EE, SI, FC, HM and PV in the current model: (1) PE had a significant effect on BI (β= .173; *p*< .001); (2) EE had a significant positive effect on BI (β= .068; *p*< .001); (3) SI had a significant positive effect on BI (β= .204; *p*<.001); (4) FC had a significant positive effect on BI (β= .070; *p*<.001); (5) HM had a significant positive effect on BI (β= .172; p<.001); and (6) PV had a significant positive effect on BI (β= .094; *p*<.001). Similar to the original UTAUT2 findings by Ventakesh et al. (2012), BI had a significant positive effect on UB (β= .525; *p*< .001); FC had a significant effect on UB (β= .193; *p*< .001), and H had a significant effect on UB (β= .264; *p*< .001). In Model 1, BI continued to be the best predictor of UB, consistent with Ventakesh et al.'s findings (2012). The results for the variables are summarised in Figure 3 and Table 2.

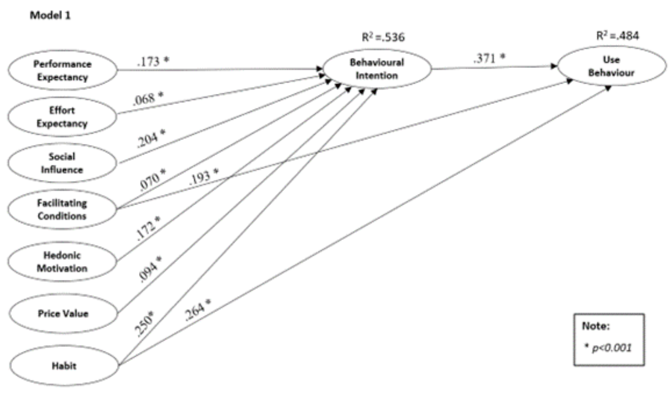


Figure 3. Path Diagram of UTAUT2 Model 1.

One additional model (Model 2) tested in this MASEM study was to include all possible exogenous variables and stimulate the various possible direct relationships between them (Figure 4). It was observed that when a direct relationship between EE and UB was added, the model fit indices fell below the desired thresholds. However, without a direct relationship between EE and UB, it was found in Model 2 that PE, EE, SI, FC, PV, HM and H were all predictors of BI and PE, SI, FC, HM, PV and PV and H were also predictors of UB. The goodness-of-fit indices for Model 2 fell within the recommended thresholds for acceptable model fit ($\chi 2/df$ = 2.226; RMSEA = .010; SRMR = .008; CFI = .999, TLI = .980) (Table 3).

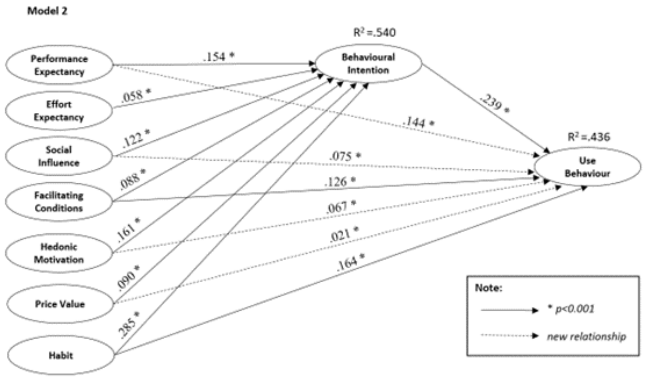


Figure 4. Path Diagram of UTAUT2 Model 2.

Table 4. Alternative UTAUT2 Model Goodness-of-fit Indices

Measure	Threshold	Value
		Model 2
χ^2	--	2.226
Df	--	1.000
χ^2/df	< 3.000	2.226
p-value	> .050	.136
RMSEA	< .050	.010
SRMR	<.080	.008
CFI	> .950	.999
TLI	> .950	.980

While there was an excellent internal data structure in Model 2, the explained variance for BI (54%) and UB (43.6%) underperformed as compared to the initial UTAUT2 model introduced by Venkatesh et al. (2012) (BI, 74%; UB, 52%) (Table 4). In Model 2, H remained the strongest predictor of BI (β= .285; p<.001), as compared to PE, EE, SI, FC, HM and PV: (1) PE had a significant effect on BI (β= .154; p<.001); (2) EE had a significant effect on BI (β= .058; p<.001); (3) SI had a significant effect on BI (β= .122; p<.001); (4) FC had a significant effect on BI (β= .088; p<.001); (5) HM had a significant effect on BI (β= .161; p<.001); and (6) PV had a significant effect on BI (β= .090; p<.001). Although BI remained to be the strongest predictor of UB (β= .239; p<.001), four other direct relationships between PE, SI, HM and PV were observed: (1) PE had a significant effect on UB (β= .144; p<.001); (2) SI had a significant effect on UB (β= .075; p<.001); HM had a significant effect on UB (β= .067; p<.001); and PV had a significant effect on UB (β= .021; p<.001). Like the initial UTAUT2 model introduced by Ventakesh et al. (2012), FC had a significant effect on UB (β= .126; p<.001), and H had a significant effect on UB (β= .164; p<.001). Compared to Model 1(53.6%), Model 2 performed slightly better, with a BI variance of 54%. However, in terms of variance explained for UB, Model 2 underperformed (43.6%) as compared to Model 1 (48.4%) (Table 5).

Table 5. Comparison of model variances explained.

	Variance Explained		
	Original Model	Model 1	Model 2
BI	.740	.536	.540
UB	.520	.484	.436

Discussion

The MASEM approach was employed to revisit the UTAUT2 model first introduced by Ventakesh et al. (2012). In Model 1, the results showed that H remained the strongest predictor of BI, with PE, EE, SI, FC, HM and PV having a significant positive effect on BI. FC, H and BI served as predictors of UB, and BI as a mediator. These results are all in line with the findings from the original UTAUT2 model. In Model 2, after adding new direct relationships into the alternative model, the findings showed that while PE, SI, FC, HM, PV and H each had a significant effect on UB, EE did not. Recent UTAUT2 studies have reported other direct relationships similar to those simulated in Model 2. For instance, Goto and Munyai (2022) reported that PV had a significant effect on UB in their study on the acceptance and use of online learning with 197 South African law students. Hu et al. (2020) found that PE and HM had a significant effect on UB when the researchers explored the factors affecting the adoption of mobile technologies with 638 Chinese academics. However, the direct relationship between SI and UB was not reported thus far in the educational context. Among the 39 studies included in this MASEM research, it was observed that two variables were commonly omitted from the UTAUT2 model: PV and UB. Of the 39 studies, 17 (43.59%) omitted PV, and 19 (48.72%) did not examine UB as an exogenous variable in the theoretical models.

While there was an attempt to examine other direct relationships between the variables in Model 2, the explained variance of both BI and UB did not perform better than the original UTAUT2 model proposed by Venkatesh et al. (2012). The possible reason would be that behavioural intentions had shifted as educational technologies changed between the period 2012 to 2023. The various technologies examined among the 39 studies covered mainly e-learning, learning management system and mobile learning. Take mobile learning, for example; in the surveys conducted by Educause Review in 2016 and 2018, students were asked to identify reasons why they did not want their teachers to use mobile apps and devices for coursework (Chen et al., 2023). For 2016 and 2018, limited internet connectivity and limited funds were among the cited reasons. In 2021, while the lack of mobile device access, limited technical support and funds were not problems for students in the 2021 survey, lack of interest was the reason. 53% of the students in the 2021 survey indicated that they would not want to use mobile apps or devices in their studies because they were not interested in mobile learning.

Conclusions

While UTAUT2 was developed for the consumer context, the findings from this MASEM study supported the model's applicability in the educational context. In Model 2, some new relationships of variables were discovered, including the direct effects of PE, SI, HM and PV on UB, which is a departure from the original findings by Ventakesh et al. (2012). Recalling that the UTAUT2 was developed for the consumer context, in the case of HM being a predictor of both BI and UB, the acceptance and use of educational technologies are driven through the extrinsic motivation of teachers and students to improve the performance of their intended tasks (Tamilmani et al., 2019). It is an important reminder to policymakers and higher education executives that extrinsic motivation plays a vital role in the successful implementation of education technologies.

PV was discovered as a predictor of both BI and UB in this study. However, only 22 of the 39 studies (56.41%) included PV as a construct in the research model. Researchers had chosen not to include PV because the users of the intended educational technologies did not need to incur any monetary cost. In contrast, some did not explain why PV was omitted in their research. Both Tamilmani et al. (2018a) and Or (2023b) suggested that PV is not an appropriate construct to be included in research models examining the adoption and use of technology made available freely to students and faculty members in higher education.

The current study synthesised empirical data from UTAUT2 studies from 2013 to 2022 in the educational context using the OSMASEM approach (Jak et al., 2021). OSMASEM synthesises correlation matrices rather than single correlations, demonstrating how the approach can be applied to examine theory-driven models. Tamilmani et al. (2019) suggested that researchers use correlation-based analysis to calculate explained variances, which this study managed to do. Many diverse findings have been discovered from past UTAUT2 studies since its inception in 2012. OSMASEM, the method introduced in this study, offers an alternative approach for researchers to use past empirical data to examine the UTAUT2 model without replicating similar studies. As more empirical data in the near future are added to train the UTAUT2 data model, researchers utilising methods like the OSMASEM can study how educational technology trends change over time, an observation established by Mishra et al. (2023) in their MASEM study on TAM research. As such, the OSMASEM approach allows researchers to focus on the critical relationships within the UTAUT2 model and advise their colleagues and executives accordingly who are implementing technologies in higher educational institutions. At the time of this writing, OSMASEM has never been utilised in the meta-analysing of the UTAUT2 model in educational contexts.

The popularity of OSMASEM in educational research is not well-established at the time of this writing, as its use is relatively recent compared to other methods in the field. One limitation of the metaSEM package used in the R software is that it cannot compute negative correlations. Future research will benefit as the software package develops in the next few years to enable it to do so. Nevertheless, OSMASEM is gaining popularity as a valuable tool for synthesising and analysing data from multiple studies, particularly in education and psychology. Its popularity may increase as researchers become more aware of its potential benefits over traditional meta-analytic methods and the availability of software packages such as metaSEM that supports the implementation of OSMASEM increases. In conclusion, OSMASEM is a recent yet valuable tool for technology acceptance studies like the UTAUT2 model. It allows researchers to synthesise data from multiple studies and evaluate measurement invariance, leading to a more comprehensive and robust understanding of the relationships between the factors proposed in the UTAUT2 model and the adoption and usage of technology in higher education.

References

Abdul Rabu, S. N., Hussin, H., & Bervell, B. (2019). QR code utilisation in a large classroom: Higher education students' initial perceptions. *Education and Information Technologies, 24*, 359-384. https://doi.org/10.1007/s10639-018-9779-2

Ain, N., Kaur, K., & Waheed, M. (2016). The influence of learning value on learning management system use: An extension of UTAUT2. *Information Development, 32*(5), 1306-1321. https://doi.org/10.1177/0266666915597546

Albright, J. J., & Park, H. M. (2009). *Confirmatory factor analysis using Amos, LISREL, Mplus, SAS/STAT CALIS*. https://pdfs.semanticscholar.org/887d/174122f76e2a232ba313866cb1b191bd15ad.pdf

Almahri, F. A. J., Bell, D., & Merhi, M. (2020, March). Understanding student acceptance and use of chatbots in the United Kingdom universities: A structural equation modelling approach. In *2020 6th International conference on Information Management (ICIM)* (pp. 284-288). IEEE. https://doi.org/10.1109/ICIM49319.2020.244712

Al-Azawei, A., & Alowayr, A. (2020). Predicting the intention to use and hedonic motivation for mobile learning: A comparative study in two Middle Eastern countries. *Technology in Society, 62*, 101325. https://doi.org/10.1016/j.techsoc.2020.101325

Arain, A. A., Hussain, Z., Rizvi, W. H., & Vighio, M. S. (2019). Extending UTAUT2 toward acceptance of mobile learning in the context of higher education. *Universal Access in the Information Society, 18*(3), 659-673. https://doi.org/10.1007/s10209-019-00685-8

Arain, A. A., Hussain, Z., Vighio, M. S., & Rizvi, W. H. (2018). Factors influencing acceptance of mobile learning by higher education students in Pakistan. *Sindh University Research Journal-SURJ (Science Series), 50*(01), 141-146. https://doi.org/10.26692/sujo/2018.1.0025

Ashraf, M. A., Shabnam, N., Tsegay, S. M., & Huang, G. (2023). Acceptance of Smart technologies in blended learning: Perspectives of chinese medical students. *International Journal of Environmental Research and Public Health, 20*(3),

2756. https://10.3390/ijerph20032756

Avcı, S., & Avcı, S. (2022). Examining the factors affecting teachers' use of digital learning resources with UTAUT2. *Malaysian Online Journal of Educational Technology, 10,* 200-214. https://doi.org/10.52380/mojet.2022.10.3.399

Azizi, S. M., Roozbahani, N., & Khatony, A. (2020). Factors affecting the acceptance of blended learning in medical education: Application of UTAUT2 model. *BMC Medical Education, 20*(1), 1-9. https://doi.org/10.1186/s12909-020-02302-2

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238. https://doi.org/10.1037//0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588. https://doi.org/10.1037/0033-2909.88.3.588

Bergh, D. D., Aguinis, H., Heavey, C., Ketchen, D. J., Boyd, B. K., Su, P., Lau, C.L.L., & Joo, H. (2016). Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership-performance relationship. *Strategic Management Journal, 37*(3), 477-497. https://doi.org/ 10.1002/smj.2338

Bervell, B. B., Kumar, J. A., Arkorful, V., Agyapong, E. M., & Osman, S. (2021). Remodelling the role of facilitating conditions for Google Classroom acceptance: A revision of UTAUT2. *Australasian Journal of Educational Technology,* 115-135. https://doi.org/10.14742/ajet.7178

Bhimasta, R. A., & Suprapto, B. (2016, November). An empirical investigation of student adoption model toward mobile e-textbook: UTAUT2 and TTF model. In *Proceedings of the 2nd international conference on communication and information processing* (pp. 167-173). https://doi.org/10.1145/3018009.3018046

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*(3), 303-316. https://doi.org/10.1177/0049124189017003004

Brown, S. A., & Venkatesh, V. (2005). A model of adoption of technology in the household: A baseline model test and extension incorporating household life cycle. *Management Information Systems Quarterly, 29*(3), 11. https://doi.org/10.2307/25148690

Cao, T. M., & Nguyen, N. P. (2022). Factors affecting students in Vietnam's intention on using smartphones for learning on the mobile learning platforms. *Journal of Educational and Social Research, 12*(2), 113-113. https://doi.org/10.36941/jesr-2022-0038

Chen, B., Denoyelles, A., Brown, T., & Seilhamer, R. (2023, January 25). The evolving landscape of students' mobile learning practices in higher education. *Educase Review.* https://er.educause.edu/articles/2023/1/the-evolving-landscape-of-students-mobile-learning-practices-in-higher-education

Cheung, M. W. L. (2014). Fixed-and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior research methods, 46*(1), 29-40. https://doi.org/10.3758/s13428-013-0361-y.

Cheung, M. W. L., & Cheung, S. F. (2016). Random-effects models for meta-analytic structural equation modeling: Review, issues, and illustrations. *Research Synthesis Methods, 7*(2), 140-155. https://10.1002/jrsm.1166

Cheung, M. W. L. (2019). Some reflections on combining meta-analysis and structural equation modeling. *Research Synthesis Methods, 10*(1), 15-22. https://doi.org/10.1002/jrsm.1321

Childers, T. L., Carr, C. L., Peck, J., & Carson, S. (2001). Hedonic and utilitarian motivations for online retail shopping behavior. *Journal of Retailing, 77*(4), 511-535. https://doi.org/10.1016/S0022-4359(01)00056-2

de Moraes, G. H. S. M., & Cabello, O. G. (2017). The information technology and the university: Use of educational applications by students. *Revista Tecnologia e Sociedade, 13*(28). https://periodicos.utfpr.edu.br/rts

Dodds, W. B., Monroe, K. B., & Grewal, D. (1991). Effects of price, brand, and store information on buyers' product evaluations. *Journal of Marketing Research, 28*(3), 307-319. https://doi.org/10.1177/002224379102800305

El-Masri, M., & Tarhini, A. (2017). Factors affecting the adoption of e-learning systems in Qatar and USA: Extending the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2). *Educational Technology Research and Development, 65*(3), 743-763. https://doi.org/10.1007/s11423-016-9508-8

Faqih, K. M., & Jaradat, M. I. R. M. (2021). Integrating TTF and UTAUT2 theories to investigate the adoption of augmented reality technology in education: Perspective from a developing country. *Technology in Society, 67*, 101787. https://doi.org/10.1016/j.techsoc.2021.101787

Farooq, M. S., Salam, M., Jaafar, N., Fayolle, A., Ayupp, K., Radovic-Markovic, M., & Sajid, A. (2017). Acceptance and use of lecture capture system (LCS) in executive business studies: Extending UTAUT2. *Interactive Technology and Smart Education.* https://doi.org/10.1108/ITSE-06-2016-0015

Fathima Sanjeetha, M. B., & Sabraz Nawaz, S. (2020). *Students' intention to use social media for academic purposes: A study of South Eastern university of Sri Lanka.* http://ir.lib.seu.ac.lk/bitstream/123456789/5428/1/JISIT-5207.pdf

Feng, G. C., Su, X., Lin, Z., He, Y., Luo, N., & Zhang, Y. (2021). Determinants of technology acceptance: Two model-based meta-analytic reviews. *Journalism & Mass Communication Quarterly, 98*(1), 83-104. https://doi.org/10.1177/1077699020952400

Furlow, C. F., & Beretvas, S. N. (2010). An evaluation of multiple imputation for meta-analytic structural equation modeling. *Journal of Modern Applied Statistical Methods, 9*(1), 14. https://doi.org/10.22237/jmasm/1272687180

Gengfu, M., & Chotiyaputta, V. (2019). Acceptance and use predictors of e-books: A case at universities in Sichuan, China. *Asian Administration & Management Review, 2*(2). https://so01.tci-thaijo.org/index.php/AAMR/article/download/242839/164807

Goto, J., & Munyai, A. (2022). The acceptance and use of online learning by law students in a South African University: An application of the UTAUT2 model. *The African Journal of Information Systems, 14*(1), 3. https://digitalcommons.kennesaw.edu/ajis/vol14/iss1/3

Gunawan, A., Muchardie, B. G., & Liawinardi, K. (2019, August). Millennial behavioral intention to adopt e-book using Utaut2 model. In *2019 International Conference on Information Management and Technology (ICIMTech)* (Vol. 1, pp. 98-102). IEEE. https://doi.org/10.1109/ICIMTech.2019.8843826

Hair Jr, J. F., Black, W. C., & Babin, B. J. (2006). *BJ, RE Anderson, and RL Tatham. Multivariate data analysis.* New Jersey.

Herhausen, D., Morgan, R. E., Brozović, D., & Volberda, H. W. (2021). Re-examining strategic flexibility: a meta-analysis of its antecedents, consequences and contingencies. *British Journal of Management, 32*(2), 435-455. 10.1111/1467-8551.12413

Holbrook, M. B., & Hirschman, E. C. (1982). The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of Consumer Research, 9*(2), 132-140. https://doi.org/10.1086/208906

Hu, S., Laxman, K., & Lee, K. (2020). Exploring factors affecting academics' adoption of emerging mobile technologies-an extended UTAUT perspective. *Education and Information Technologies, 25*(5), 4615-4635. https://doi.org/10.1007/s10639-020-10171-x

Jak, S., & Cheung, M. W. L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods, 25*(4), 430. https://psyarxiv.com/ce85j/download?format=pdf

Jak, S., Li, H., Kolbe, L., de Jonge, H., & Cheung, M. W. L. (2021). Meta-analytic structural equation modeling made easy: A tutorial and web application for one-stage MASEM. *Research Synthesis Methods, 12*(5), 590-606. https://doi.org/10.1002/jrsm.1498

Jeyaraj, A., & Dwivedi, Y. K. (2020). Meta-analysis in information systems research: Review and recommendations. *International Journal of Information Management, 55,* 102226. https://doi.org/10.1016/j.ijinfomgt.2020.102226

Jung, I., & Lee, J. (2020). A cross-cultural approach to the adoption of open educational resources in higher education. *British Journal of Educational Technology, 51*(1), 263-280. https://doi.org/10.1111/bjet.12820

Kalinkara, Y., & Talan, T. (2022). Rethinking evaluating the use of distance learning systems in the context of the Unified Theory of Acceptance and Use of Technology-2. *Journal of Learning for Development, 9*(2), 229-252. https://files.eric.ed.gov/fulltext/EJ1358467.pdf

Kaur, D. J., Saraswat, N., & Alvi, I. (2021). Exploring the effects of blended learning using WhatsApp on language learners' lexical competence. *Rupkatha Journal on Interdisciplinary Studies in Humanities, 13*(4). https://doi.org/10.21659/rupkatha.v13n4.60

Kim, S. S., & Malhotra, N. K. (2005). A longitudinal model of continued IS use: An integrative view of four mechanisms underlying postadoption phenomena. *Management Science, 51*(5), 741-755. https://doi.org/10.1287/mnsc.1040.0326

Kline, R. B. (2005). *Methodology in the social sciences. Principles and practice of structural equation modeling*, 2.

Kumar, J. A., & Bervell, B. (2019). Google Classroom for mobile learning in higher education: Modelling the initial perceptions of students. *Education and Information Technologies, 24*(2), 1793-1817. https://doi.org/10.1007/s10639-018-09858-z

Lai, C., Wang, Q., Li, X., & Hu, X. (2016). The influence of individual espoused cultural values on self-directed use of technology for language learning beyond the classroom. *Computers in Human Behavior, 62,* 676-688. http://dx.doi.org/10.1016/j.chb.2016.04.039

Le, T. L. H., Hoang, V. H., Hoang, M. D. M., Nguyen, H. P., & Bui, X. B. (2022). Impact of digital literacy on intention to use technology for online distribution of higher education in Vietnam: A study of Covid19 context. *Journal of Distribution Science, 20*(6), 75-86. https://doi.org/10.15722/jds.20.06.202206.75

Leong, L. Y., Hew, T. S., Ooi, K. B., Metri, B., & Dwivedi, Y. K. (2022). Extending the theory of planned behavior in the social commerce context: A Meta-Analytic SEM (MASEM) approach. *Information Systems Frontiers,* 1-33. https://doi.org/10.1007/s10796-022-10337-7

Limayem, M., & Hirt, S. G. (2003). Force of habit and information systems usage: Theory and initial validation. *Journal of the Association for information Systems, 4*(1), 65-97. https://doi.org/10.17705/1jais.00030

Lin, S., Zimmer, J. C., & Lee, V. (2013). Podcasting acceptance on campus: The differing perspectives of teachers and students. *Computers & Education,* (68), pp. 416-428. http://dx.doi.org/10.1016/j.compedu.2013.06.003

Malešević, A., Barać, D., Soleša, D., Aleksić, E., & Despotović-Zrakić, M. (2021). Adopting xRM in higher education: E-services outside the classroom. *Sustainability, 13*(14), 7522. https://doi.org/10.3390/su13147522

Meet, R. K., Kala, D., & Al-Adwan, A. S. (2022). Exploring factors affecting the adoption of MOOC in Generation Z using extended UTAUT2 model. *Education and Information*

Technologies, 1-23. https://doi.org/10.1007/s10639-022-11052-1

Mehta, A., Morris, N. P., Swinnerton, B., & Homer, M. (2019). The influence of values on e-learning adoption. *Computers & Education, 141*, 103617. https://doi.org/10.1016/j.compedu.2019.103617

Mishra, N., Gupta, S. L., Srivastava, P., Srivastava, S., & Kabir, M. (2022). Student acceptance of social media in higher education: An application of UTAUT2 model. *Thailand and The World Economy, 40*(1), 88-108. https://so05.tci-thaijo.org/index.php/TER/article/download/257041/173461/941274

Mishra, A., Shukla, A., Rana, N. P., Currie, W. L., & Dwivedi, Y. K. (2023). Re-examining post-acceptance model of information systems continuance: A revised theoretical model using MASEM approach. *International Journal of Information Management, 68,* 102571. https://doi.org/10.1016/j.ijinfomgt.2022.102571

Moorthy, K., Chun T'ing, L., Ming, K. S., Ping, C. C., Ping, L. Y., Joe, L. Q., & Jie, W. Y. (2019). Behavioral intention to adopt digital library by the undergraduates. *International Information & Library Review, 51*(2), 128-144. https://doi.org/10.1080/10572317.2018.1463049

Moorthy, K., Yee, T. T., T'ing, L. C., & Kumaran, V. V. (2019). Habit and hedonic motivation are the strongest influences in mobile learning behaviours among higher education students in Malaysia. *Australasian Journal of Educational Technology, 35*(4). https://doi.org/10.14742/ajet.4432

Musa, M., Ismail, M. N., Tahir, S., Fudzee, M. F. M., & Jofri, M. H. (2022). Student Acceptance Towards Online Learning Management System based on UTAUT2 model. *International Journal of Advanced Computer Science and Applications, 13*(11). https://10.14569/IJACSA.2022.0131115

Nikolopoulou, K., Gialamas, V., & Lavidas, K. (2020). Acceptance of mobile phone by university students for their studies: An investigation applying UTAUT2 model. *Education and Information Technologies, 25*(5), 4139-4155. https://doi.org/10.1007/s10639-020-10157-9

Or, C. (2023a). Revisiting Unified Theory of Technology and Use of Technology using meta-analytic structural equation modelling. *International Journal of Technology in Education and Science, 7*(1), 83-103. https://doi.org/10.46328/ijtes.420

Or, C. C. P. (2023b). Towards an integrated model: Task-Technology fit in Unified Theory of Acceptance and Use of Technology 2 in education contexts. *Journal of Applied Learning and Teaching, 6*(1), 151-163. https://doi.org/10.37074/jalt.2023.6.1.8

Osei, H. V., Kwateng, K. O., & Boateng, K. A. (2022). Integration of personality trait, motivation and UTAUT 2 to understand e-learning adoption in the era of COVID-19 pandemic. *Education and Information Technologies,* 1-26. https://doi.org/10.1007/s10639-022-11047-y

Raeisi-Vanani, A., Plonsky, L., Wang, W., Lee, K., & Peng, P. (2022). Applying meta-analytic structural equation modeling to second language research: An introduction. *Research Methods in Applied Linguistics, 1*(3), 100018. https://doi.org/10.1016/j.rmal.2022.100018

Raman, A., & Don, Y. (2013). Pre-service teachers' acceptance of learning management software: An application of the Utaut2 model. *International Education Studies, 6*(7), 157-164. http://dx.doi.org/10.5539/ies.v6n7p157

Raman, A., & Thannimalai, R. (2021). Factors impacting the behavioural intention to use e-learning at higher education amid the Covid-19 pandemic: UTAUT2 model. *Psychological Science and Education, 26*(3), 82-93. https://doi .org/10 .17759/pse .2021260305

Rudhumbu, N. (2022). Applying the UTAUT2 to predict the acceptance of blended learning by university students. *Asian Association of Open Universities Journal, 17*(1), 15-36. https://www.emerald.com/insight/2414-6994.htm

Sharif, A., Afshan, S., & Qureshi, M. A. (2019). Acceptance of learning management system in university students: An integrating framework of modified UTAUT2 and TTF theories. *International Journal of Technology Enhanced Learning, 11*(2), 201-229. https://10.1504/IJTEL.2019.098810

Steinmetz, H., & Block, J. (2022). Meta-analytic structural equation modeling (MASEM): New tricks of the trade. *Management Review Quarterly,* 1-22. https://doi.org/10.1007/s11301-022-00293-6

Steel, P., Beugelsdijk, S., & Aguinis, H. (2021). The anatomy of an award-winning meta-analysis: Recommendations for authors, reviewers, and readers of meta-analytic reviews. *Journal of International Business Studies, 52*(1), 23-44. https://doi.org/10.1057/s41267-020-00385-z

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173-180. http://www.statpower.net/Steiger%20Biblio/Steiger90b.pdf

Tamilmani, K., Rana, N. P., & Dwivedi, Y. K. (2017, November). A systematic review of citations of UTAUT2 article and its usage trends. In *Conference on e-business, e-services and e-society* (pp. 38-49). Springer, Cham. https://doi.org/10.1007/978-3-319-68557-1_5

Tamilmani, K., Rana, N., Dwivedi, Y., Sahu, G. P., & Roderick, S. (2018a). *Exploring the role of 'price value' for understanding consumer adoption of technology: A review and meta-analysis of UTAUT2 based empirical studies*. https://aisel.aisnet.org/pacis2018/64

Tamilmani, K., Rana, N. P., & Dwivedi, Y. K. (2018b, June). Use of 'habit' is not a habit in understanding individual technology adoption: a review of UTAUT2 based empirical studies. In *International working conference on transfer and diffusion of IT* (pp. 277-294). Springer, Cham. https://doi.org/10.1007/978-3-030-04315-5_19

Tamilmani, K., Rana, N. P., Prakasam, N., & Dwivedi, Y. K.

(2019). The battle of brain vs. heart: A literature review and meta-analysis of "hedonic motivation" use in UTAUT2. *International Journal of Information Management, 46*, 222-235. https://doi.org/10.1016/j.ijinfomgt.2019.01.008

Tamilmani, K., Rana, N. P., & Dwivedi, Y. K. (2021). Consumer acceptance and use of information technology: A meta-analytic evaluation of UTAUT2. *Information Systems Frontiers, 23*, 987-1005. https://doi.org/10.1007/s10796-020-10007-6

Tang, R. W., & Cheung, M. W. L. (2016). Testing IB theories with meta-analytic structural equation modeling: The TSSEM approach and the Univariate-r approach. *Review of International Business and Strategy*. https://doi.org/10.1108/RIBS-04-2016-0022

Than, W. W., Kyaw, E. M., & Htoo, H. Z. (2021) A meta-analytic structural equation modelling on the Unified Theory of Acceptance and Use of Technology in higher education. *International Journal of Educational Management and Development Studies, 2*(4), 44-71. https://doi.org/10.53378/352074

Thong, J. Y., Hong, S. J., & Tam, K. Y. (2006). The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. *International Journal of Human-Computer Studies, 64*(9), 799-810. https://doi.org/10.1016/j.ijhcs.2006.05.001

Tseng, T. H., Lin, S., Wang, Y. S., & Liu, H. X. (2019). Investigating teachers' adoption of MOOCs: The perspective of UTAUT2. *Interactive Learning Environments*, 1-16. https://doi.org/10.1080/10494820.2019.1674888

Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS quarterly,* 695-704. https://doi.org/10.2307/25148660

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly,* 425-478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS quarterly,* 157-178. https://doi.org/10.2307/41410412

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel psychology, 48*(4), 865-885. https://doi.org/10.1111/j.1744-6570.1995.tb01784.x

Widjaja, H. A. E., Santoso, S. W., Fernando, E., & Condrobimo, A. R. (2020, November). Improving the quality of Learning Management System (LMS) based on student perspectives using UTAUT2 and trust model. In *2020 4th International conference on Informatics and Computational Sciences (ICICoS)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICICoS51170.2020.9298985

Wilson, S. J., Polanin, J. R., & Lipsey, M. W. (2016). Fitting meta-analytic structural equation models with complex datasets. *Research Synthesis Methods, 7*(2), 121-139. https://doi.org/10.1002/jrsm.1199

Xu, W., Shen, Z. Y., Lin, S. J., & Chen, J. C. (2022). Improving the behavioral intention of continuous online learning among learners in higher education during COVID-19. *Frontiers in Psychology, 13*, 857709-857709. https://doi.org/10.3389/fpsyg.2022.857709

Zacharis, G., & Nikolopoulou, K. (2022). Factors predicting University students' behavioral intention to use eLearning platforms in the post-pandemic normal: An UTAUT2 approach with 'learning value'. *Education and Information Technologies,* 1-18. https://doi.org/10.1007/s10639-022-11116-2

Zaremohzzabieh, Z., Roslan, S., Mohamad, Z., Ismail, I. A., Ab Jalil, H., & Ahrari, S. (2022). Influencing factors in MOOCs adoption in higher education: A meta-analytic path analysis. *Sustainability, 14*(14), 8268. https://doi.org/10.3390/su14148268

Zhou, Y., Li, X., & Wijaya, T. T. (2022). Determinants of behavioral intention and use of interactive whiteboard by K-12 teachers in remote and rural areas. *Frontiers in Psychology, 13*, 934423. https://10.3389/fpsyg.2022.934423

Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *Journal of Marketing, 52*(3), 2-22. https://doi.org/10.1177/002224298805200302

Zwain, A. A. A. (2019). Technological innovativeness and information quality as neoteric predictors of users' acceptance of learning management system: An expansion of UTAUT2. *Interactive Technology and Smart Education*. https://doi.org/10.1108/ITSE-09-2018-0065

Zwain, A. A. A., & Haboobi, M. N. H. (2019). Investigating determinants of faculty and students' acceptance of e-learning management systems using UTAUT2. *Technology, 7*(8). https://www.ijicc.net/images/vol7iss8/7817_Zwain_2019_E_R.pdf

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Predicting students' multidimensional learning outcomes in public secondary schools: The roles of school facilities, administrative expenses and curriculum

Valentine J. Owan[A]
John A. Ekpenyong[B]
Usen F. Mbon[C]
Kingsley B. Abang[D]
Nse N. Ukpong[E]
Maria O. Sunday[F]
Samuel O. Ekaette[G]
Michael E. Asuquo[H]
Victor U. Agama[I]
Garieth O. Omorobi[J]
John A. Undie[K]

A Department of Educational Foundations, University of Calabar, Calabar, Nigeria; Ultimate Research Network (URN), Calabar, Cross River State, Nigeria
B Department of Educational Management, University of Calabar, Calabar, Nigeria; Ultimate Research Network (URN), Calabar, Cross River State, Nigeria
C Department of Educational Management, University of Calabar, Calabar, Nigeria
D Department of Educational Foundations, University of Calabar, Calabar, Nigeria
E Department of Educational Management, University of Calabar, Calabar, Nigeria
F Department of Educational Foundations, University of Calabar, Calabar, Nigeria
G Department of Educational Management, University of Calabar, Calabar, Nigeria
H Department of Educational Management, University of Calabar, Calabar, Nigeria; Ultimate Research Network (URN), Calabar, Cross River State, Nigeria
I Department of Educational Foundations, University of Calabar, Calabar, Nigeria; Ultimate Research Network (URN), Calabar, Cross River State, Nigeria
J Department of Educational Management, University of Calabar, Calabar, Nigeria
K Department of Educational Management, University of Calabar, Calabar, Nigeria

Keywords

Factor analysis;
hierarchical regression;
learning;
Nigeria;
school inputs;
students' achievement.

Correspondence

owanvalentine@gmail.com [A]

Article Info

Received 12 May 2023
Received in revised form 1 July 2023
Accepted 15 July 2023
Available online 19 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.9

Abstract

Previous research has assessed school facilities, administrative expenditures and curriculum and their relative contributions to students' cognitive learning outcomes. This suggested the need to investigate further how these predictors may impact students' affective and psychomotor outcomes. The current research studied the combined and relative prediction of school facilities, administrative expenses and curriculum on students' overall cognitive, affective and psychomotor learning outcomes in public secondary schools. A cross-sectional research design was employed in this study, involving 87 school administrators and a randomly selected group of 915 senior secondary class II (SS2) students. For data collection, we utilised the School Inputs Questionnaire (SIQ) and Educational Outcomes Questionnaire (EOQ), both developed by the researchers and validated through expert assessments, including content validity, Exploratory Factor Analysis (EFA) for dimensionality, Confirmatory Factor Analysis (CFA) for goodness of fit, and reliability using Cronbach's alpha. The results of these assessments demonstrated acceptable outcomes aligned with international standards. Hierarchical regression analysis was conducted to analyse the collected data. The findings indicated that enhancing the provision of quality school facilities, administrative expenses, and school curricula improved students' overall cognitive, affective, and psychomotor learning outcomes. Specifically, administrative expenses and school curriculum had significant predictive power for students' overall cognitive, affective, and psychomotor learning outcomes. However, while school facilities significantly predicted students' overall, affective, and psychomotor dimensions, they did not significantly predict the cognitive dimension. These findings offer valuable insights for policymakers and educators aiming to enhance the educational quality in public secondary schools.

Introduction

Students' learning outcomes refer to the knowledge, skills, attitudes, and values students acquire due to their education. These outcomes are important because they impact students' ability to succeed in school and their future careers and reflect the education system's effectiveness. Learning outcomes are measured by the quality of students' cognitive, affective, and psychomotor skills after exposure to lesson contents and experiences in the short- or long-term (Ekpenyong et al., 2022; Owan et al., 2022). Cognitive learning outcomes refer to students' knowledge, understanding, and intellectual development from their educational experiences. These can include understanding and recalling information, solving problems, analysing and synthesising information, applying knowledge to new situations, and thinking critically and creatively (Ali, 2013; Bassey & Owan, 2020). The measure of cognitive skills follows Bloom's taxonomy of educational objectives, including "knowledge", "comprehension", "application", "analysis", "synthesis", and "evaluation" (Bloom et al., 1956). These have, however, been revised to "remember", "understand", "apply", "analyse", "evaluate", and "create" (Anderson & Krathwohl, 2001; Krathwohl, 2002). Cognitive learning outcomes are often measured through tests and may be assessed at the individual, class, school, or district levels (Ekpenyong et al., 2022).

Affective outcomes refer to students' personality, social and emotional traits, as well as their behaviour, morals and values. It can be measured using constructs such as self-concept, motivation, study habits, respect for rules and regulations, and positive attitudes towards learning, others, and the world around them (Almlund et al., 2011; Lipnevich et al., 2013). Affective learning outcomes are often more difficult to measure than cognitive learning outcomes, as they involve changes in students' attitudes and emotions rather than their knowledge or skills (Anderson & Bourke, 2013; Levin, 2012). However, they are no less important, as students' attitudes and emotions can significantly impact their learning and overall wellbeing (Ekpenyong et al., 2022; Owan et al., 2022). These affective attributes predict students' school learning behaviour and determine their societal adjustment. Thus, a key educational outcome must be attained for the functionality of society at large.

Similarly, the psychomotor learning outcomes measure how students have acquired functional skills in schools to adapt and function in a dynamic society. The psychomotor skills of the students are commonly referred to as the physical or vocational abilities normally developed during practical activities in schools (Zhao & Hong, 2012). When students acquire basic vocational skills in school, they become functional and useful in society and can take up menial employment for self-reliance (Nathan et al., 2017). Individuals who excel academically, possess a strong aptitude for learning and have admirable personal qualities are considered excellent students and valuable assets to society (Olaitan, 2017). This implies that adequate assessment of goal attainment in secondary schools should go beyond measuring students' cognitive attainment to capturing their affective attributes and psychomotor skills (Bassey et al., 2019; Ekpenyong et al., 2022, 2023; Owan et al., 2022).

In the last two decades, studies on students' learning outcomes in Nigeria, just as with most African or developing countries, have continually lamented over the quality of students produced from the secondary education system (e.g., Ajayi & Yusuf, 2010; Arop et al., 2018; Odigwe et al., 2018). Some scholars have complained that Nigeria's laudable secondary education goals are not maximally attained due to an increased number of youths without functional skills (Suleiman, 2018; Ololube et al., 2016) and a high rate of social upheaval among Nigerian youths (Adelaja & George, 2020). Others have pointed to students' low scores on standardised and teacher-made tests (Eze, 2021; Owan & Ekpenyong, 2022; Ugwuanyi et al., 2020), high rate of indiscipline (Gcelu et al., 2020) and moral decadence (Sanga, 2022) as indicators of poor learning outcomes in African countries. Despite the laudable secondary education policies in Nigeria, many youths are still idle, do not live usefully in society, are not able to transit into institutions of higher learning, and lack any sense of self-worth or regard for the opinions and emotions of others (Pastore, 2019; Robert & Owan, 2019; Watson et al., 2016). Thus only a few per cent of the students can gain access to tertiary institutions every year (Herbaut & Geven, 2020; Ilie & Rose, 2016) due to their high rate of involvement in examination malpractices (Agwu et al., 2022; Okolie et al., 2019; Owan et al., 2023). These observed inadequacies point to the fact that secondary schools, which are supposed to prepare students for functional living through acquiring the right skills, values, and learning and boosting their dispositions for higher learning, have not attained their goals.

Many factors can influence students' learning outcomes. Some of these factors are internal to the student, such as their innate ability, health status, social capital (Owan et al., 2022), motivation (Baber, 2020), self-regulation (Shing & Rameli, 2020), and prior knowledge (Alabdulkarem et al., 2021). Other factors are external to the student and include the quality of teaching (Belsito, 2016; Robert & Owan, 2019), teachers' pedagogic service discharge (Ngware & Mutisya, 2022; Owan et al., 2022), the curriculum (Kazima et al., 2022; Peterson & Mlynarczyk, 2016), and the learning environment (Matthews & Mercer-Mapstone, 2018), among others. In this paper, the emphasis is on school inputs as predictors of students' learning outcomes. School inputs are all the factors or characteristics available in schools that can influence the entire education production process. These include infrastructural provisions, student-teacher ratio, administrative expenses, funding, classroom time utilisation rates, class size, and school curriculum (Nghambi, 2015). Although these school inputs have since been identified as crucial for students' learning outcomes, the degree of such a relationship has rarely been investigated in Cross River State, Nigeria. This creates a knowledge gap since it remains unclear the degree to which school inputs predict students' learning outcomes across the cognitive, affective and psychomotor dimensions of learning outcomes. Previous studies on school inputs have mostly assessed the contribution of individual inputs to students' cognitive outcomes, ignoring the affective and psychomotor dimensions. In the present study, we used three specific variables (school facilities, administrative expenses, and school curriculum) as proxies for school inputs. The next section reviews previous studies on each specific input about students' learning outcomes in secondary schools.

Studies on school facilities

In the last decade, a large body of research has focused on analysing the role that school facilities play in the education of students. Previous research has attempted to analyse the extent to which different types of school facilities are available (Akah et al., 2022; Nurabadi et al., 2020; Owan & Owan, 2022), adequate (Ademiluyi, 2019; Alabi, 2021), functional, accessible (Islam et al., 2020; Oluwalola, 2021), and utilised (Bervell & Arkorful, 2020) for teaching and learning. These studies have revealed different degrees of resource availability in secondary schools. For instance, some studies discovered a low extent in the availability and adequacy of school resources (Lawanson & Gede, 2011; Takwate, 2018). On the contrary, other studies have found a great extent in the availability of diverse resources for teaching and learning after the COVID-19 pandemic (Akah et al., 2022; Owan & Owan, 2022), with public schools revealed as having more facilities than private (Arshad et al., 2020). The low availability of facilities recorded in some studies and the high availability recorded in others may be due to the undersupply, optimum supply and oversupply of school materials resources.

It has been shown that over- and under-provision of resources or inequitable distribution of materials to schools result in the waste of school material resources (Mbon et al., 2020). Nevertheless, the disparity in the results of previous studies regarding the availability status of school facilities creates an evidence gap. It is a sign that further research is necessary for more clarification. Besides, most studies did not consider the role school facilities played in the educational outcomes of learners in secondary schools. Bridging this gap, however, other researchers have linked the availability of school facilities to students' motivation (Sidi, 2019) and learning outcomes (Arshad et al., 2018; Takwate, 2018). It has been proven that the availability and layout of school physical facilities can promote students' learning outcomes (Ariani, 2015; Daramola et al., 2017). Other researchers have argued that the mere availability of facilities does not promote learning outcomes, as some study suggests; instead, they proved that teachers' utilisation of available resources has a nexus with students' learning outcomes (Akah et al., 2022; Issacar & Hesbon, 2021; Owan & Ekpenyong, 2022).

Similarly, the functionality of school facilities has been linked to students' learning outcomes in secondary schools (Dube, 2019). Although different terms and phrases have been used to mean learning outcomes in most previous studies, one issue is common among them. Most studies have treated students' learning outcomes as a unidimensional construct by focusing too much on the cognitive domain of Bloom's taxonomy (Bassey et al., 2019; Ekpenyong et al., 2022; Owan et al., 2022). It has been argued that how well a child learns is reflected in the cognitive, affective and psychomotor attributes (Akhiruyanto et al., 2022; Orak et al., 2020; Robert & Owan, 2019). Therefore, any measurement of students' learning outcomes must consider the three domains of learning (cognitive, affective and psychomotor) to be considered adequate (Ekpenyong et al., 2022; Owan et al., 2022).

The few studies that have assessed students' learning outcomes from the three domains of Bloom's taxonomy did not focus on school facilities as the predictor. Their foci were on variables such as chemistry laboratory curriculum (Enneking et al., 2019), students' variables (Owan et al., 2022), instructional videos (Cooper & Higgins, 2015), teachers and administrators' inputs (Ekpenyong et al., 2022), and quality assurance practices (Bassey et al., 2019) among others. The existing gap in the literature was the driving force behind the present study.

Studies on administrative expenses

Administrative expenses are overhead expenditures that educational managers make in the day-to-day running of the school. These expenditures depend on how much income the school has available (Odigwe & Owan, 2022). Therefore, the government's national expenditure in funding the education system is important for local expenses at the institutional level (Ekaette et al., 2019; Odigwe & Owan, 2019). Other sources of funds for administrative expenses are internal revenue generation (Mbah & Onuora, 2018; Odigwe, 2020) and alternative funding (Onyeche, 2018). Despite the importance of internal funding and expenditure, the literature has been silent on school managers' administrative expenses. Previous studies on administrative expenditures over the last decade have focused on government ministries, agencies and other parastatals (e.g., Chernew & Mintz, 2021; Cunha, 2018). Other studies have assessed administrative expenses as a criterion variable responding to different predictors in the context of corporate firms, non-governmental and banking organisations (Fan & Liu, 2017; Venieris et al., 2015).

In the education sector, most studies have focused on school leaders' budgeting (Sinclair & Malen, 2021), accountability (Keddie & Holloway, 2020; Paletta et al., 2020; Wang et al., 2022), cost-sharing (Alazmi & Al-Kubaisi, 2020; Hayes & Burkett, 2021), internal revenue generation (Mbah & Onuora, 2018; Odigwe, 2020; Onyeche, 2018) fund management (Aliyu, 2018; Odigwe & Owan, 2022; Owan et al., 2021), and resource procurement practices (Buys et al., 2020; Prabhakar et al., 2022) and other related constructs. Although some of these variables are tied to administrative expenses in one or the other (for example, resource procurement), the extent to which principals' day-to-day expenditure predicts students' outcomes was not the focus of the cited studies.

Admittedly, studies have documented that school spending was associated with students' academic achievement (Hægeland et al., 2012; Nicoletti & Rabe, 2018). The importance of instructional expenses for achieving high student test results in any educational system was also emphasised (Webber, 2012). Similarly, it was shown that rural students' academic performance is positively correlated with the amount of money spent on their education and the number of years they spend in school (Munda & Odebero, 2014). A study has also shown that increased school funding was associated with increased administrative expenses, students' discipline, attendance and academic success (Huntoon, 2021). A case was also presented by the finding of Gigliotti and Sorensen (2018) that sustained financial investment in schools was crucial for districts to maintain quality public education. Even though school financing and administrative expenditures are important predictors of students' learning outcomes (Bruce et al., 2019; Strickland, 2021), the link is known for the cognitive aspect of students' learning outcomes. There seems to be no existing study connecting principals' administrative expenses to students' affective and psychomotor learning outcomes. Based on this gap, the present study assessed how principals' administrative expenses predict secondary school students' affective, cognitive and psychomotor learning outcomes.

Studies on the school curriculum

The curriculum is a structured plan of education developed by schools or other organisations to help learners gain a deeper understanding and mastery of the material. It is designed to improve their abilities and contribute to society's overall wellbeing (Megbo & Saka, 2015). In a study, Demir et al. (2012) revealed that students acquired efficient studying skills through the curriculum for increased academic achievements. It has been discovered that students exposed to a new science curriculum improved their analytic ability, processing capacity and other skills, such as reading, mathematics and communication, than those taught using the traditional curriculum (Alghamdi, 2017; Shymansky et al., 1983). Furthermore, students' achievement levels in both coordinate and synthetic geometry improved after exposure to quality curriculum contents compared to students in the control group (Senk, 2020).

Different studies have linked the different dimensions of students' school achievement to the school curriculum. For instance, it was discovered that students' academic and social skills were enhanced as they could communicate, organise their ideas, share information and express opinions due to improvements in the curriculum contents and experiences (Alismail & McGuire, 2015). Again, in China, it was found that the new curriculum used in schools changed students' attitudes and led to more positive views of the government (Cantoni et al., 2017). In the same direction, research in South Africa indicated a direct association between the curriculum implemented and students' success in the schools (Dhunpath & Subbaye, 2018). The correlation between school-based curriculum and students' academic achievement was moderated by factors such as students' ability, the quality of school resources, internal and external support for schools and the quality of the curriculum arrangement process (Wiyono, 2018). Also, when exposed to an improved school curriculum, students showed higher learning achievements and higher motivation to learn science using digital technologies (Alnajjar, 2022). This implies that the school curriculum affected students' academic performance and psychomotor and affective skills.

However, the study conducted by Ni et al. (2011) found mixed results when studying the effectiveness of the school curriculum in improving students' performance in all three areas of learning. The cited authors found that implementing a quality school curriculum improved the cognitive performance of all the students and their psychomotor attributes (such as routine problem-solving and complex problem-solving skills), and it improved the affective attributes of learners. However, another research provided contrary findings that the curriculum only improved students' achievement scores of cognitive ability but not their post-school affective (such as the behaviour of the students in society) and psychomotor (functional skills for employment) outcomes (Bouck & Joshi, 2012).

Several scholars have attempted to link the school curriculum to student achievement. Only a handful of these studies have attempted to relate the quality of the school curriculum to the three domains. In fact, in some cases, studies focus on one domain per time, with just a few addressing the three concurrently (Bouck & Joshi, 2012; Ni et al., 2011). Among the studies that capture the three learning domains relative to the school curriculum, there also seem to be disagreements on how strongly the two variables are related, the direction of the relationship, and the importance of the correlation. This situation creates an evidence gap since the findings in the literature are inconclusive and warrant further studies to clarify the ongoing debate. Also, a cursory look at the literature further shows a decline in recent studies, making the area seem neglected. Moreover, limited literature on curriculum development and its impact on students' learning outcomes in Nigeria across the cognitive, affective, and psychomotor domains has created a knowledge gap. Due to the gap, it is yet to be known whether the school curriculum has a role to play in deciding how students think (cognitive), behave (affective) and showcase their skills (psychomotor). However, conducting studies in these areas is important to improve the curriculum's effectiveness and enhance students' learning outcomes. Based on this identified gap, the present study also examined how the curriculum predicts secondary school students' affective, cognitive and psychomotor learning outcomes.

Theoretical and conceptual frameworks

This study is grounded in the Input-Process-Output (IPO) model of education production. The IPO model is widely used in educational research and provides a conceptual framework for understanding the relationships between inputs, processes, and outputs. In the IPO model, inputs refer to the resources and factors that influence the learning process (Decius et al., 2021; Huang et al., 2021). School facilities, administrative expenses, and curriculum can be considered as inputs that impact students' learning outcomes. These inputs provide the foundation for the teaching and learning processes within the educational setting (Ekpenyong et al., 2022, 2023; Owan et al., 2022).

Processes represent the instructional activities, teaching methods, and interactions between teachers and students that occur within the learning environment (Chen et al., 2022; Wong et al., 2022). The inputs influence these processes and play a critical role in shaping students' learning outcomes (Decius et al., 2021). In this case, output is students' learning outcomes, including cognitive, affective, and psychomotor dimensions. The outputs are the result of the interactions between the inputs and the teaching and learning processes (Ekpenyong et al., 2023; Owan & Ekpenyong, 2022; Robert

& Owan, 2019).

In the current study, the IPO model was useful in understanding how school facilities, administrative expenses, and curriculum can influence the teaching and learning processes and subsequently impact students' learning outcomes. It provides a theoretical framework to examine the relationships between these variables and offers insights into the mechanisms through which inputs affect outputs. Although the process (teaching and learning) was not measured nor examined in the current, the state of the output (students' learning outcomes) across the three dimensions offers insight into the process. Based on this theoretical underpinning, the conceptual model of this study was developed, as shown in Figure 1.

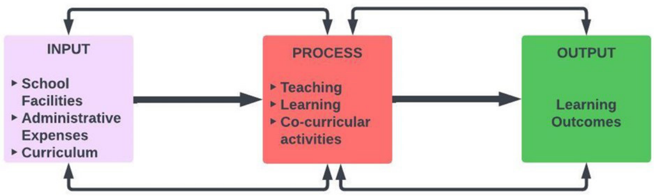


Figure 1: Conceptual model of the study.

Research question

The primary research question that underly this study is: What are the relative and composite contributions of school facilities, administrative expenses and curriculum to students' overall cognitive, affective and psychomotor learning outcomes in public secondary schools?

Hypothesis

The hypothesis tested in this study is as follows:

Ho: There are no significant relative and composite contributions of school facilities, administrative expenses and curriculum to students' overall cognitive, affective and psychomotor learning outcomes in public secondary schools.

H1: There are significant relative and composite contributions of school facilities, administrative expenses and curriculum to students' overall cognitive, affective and psychomotor learning outcomes in public secondary schools.

Methods

Research design

The study adopted the quantitative research method, drawing from the positivist research philosophy. The use of quantitative research involves the collection and analysis of numerical data to examine relationships, patterns, and statistical significance. This approach provides objective and empirical evidence to support or reject hypotheses and research questions. Positivism is a philosophical stance that uses scientific methods to understand and explain the social world (Tamminen & Poucher, 2020). Positivists believe that knowledge can be gained through systematic observation and measurement, and they seek to establish causal relationships between variables (Zyphur & Pierides, 2020). The research design adopted for this study was the cross-sectional research design. A key feature of cross-sectional studies is the observation of variables in a single moment (Zangirolami-Raimundo et al., 2018). This design was deemed appropriate for this study because the researchers concentrated on SS2 students as the unit of measurement for learning outcomes since school facilities, administrative expenses, and curriculum also affect them. Secondly, the design allowed for estimating the contributions of all the explanatory variables on the criterion variables based on data collected at a time point.

Study participants

Both secondary school administrators (N= 87) and senior secondary class II (SS2) students (N = 53,255; males = 26,206; females = 27,047) constituted the targeted population for this research. 915 SS2 students were randomly selected using a multistage selection technique, while principals were not sampled since their population was manageable (More details about the sampling process can be found in (Ekpenyong et al., 2022, 2023; Owan et al., 2022). For students' demographics, 44.1% are males, while 55.9% are females. For age, 48.9% of the students are between 10 and 20 years, while 51.1% are 21 or older. Regarding socioeconomic status (SES), 50.5% of the students are from families with a high SES, while 48.2% are from families with a low SESs. Conversely, 51.8% of students came from broken families, while the remaining 48.2% were from families with intact structures. Again, 47.2% of the students are members of small families, while 52.8% are members of large families. For principals, 50.7% were males, while 49.3% were females; 3.8% were Nigeria Certificate in Education (NCE) or Ordinary National Diploma (OND) holders, 79.9% were Higher National Diploma (HND) or First Degree holders, 13.9% were Master's degree holders, and 2.4% were doctorate holders. Regarding years of work experience, 25.7% had 0 to 10 years of experience, 22.6% had 11 to 20 years of experience, 24.8% had 21 to 30 years of experience, and 26.8% had 31 years of experience or higher.

Measures

The study has three independent variables: school facilities, administrative expenses, and curriculum. School facilities refer to the physical resources and infrastructure available in schools. It encompasses various aspects such as classrooms, libraries, laboratories, computer facilities, sports facilities, and other amenities. The study aimed to understand how the provision and quality of these resources influence students' learning outcomes across different dimensions. Administrative expenses pertain to the financial resources allocated and utilised by school administrators for the management and operation of the educational institution. It encompasses budgeting, expenditure on administrative functions, financial planning, resource allocation, and fund management. The study investigated how the prudent utilisation of administrative expenses impacts students' learning outcomes. The curriculum serves as a framework for educational instruction and encompasses the content, pedagogical approaches, learning objectives, and assessment methods employed in schools. The study examined the design, content, and implementation of the curriculum. It aimed to understand how the curriculum influences students' overall, cognitive, affective, and psychomotor learning outcomes. These independent variables were selected based on their potential influence on students' learning outcomes.

The dependent variable in the study is students' learning outcomes. Specifically, the study examined students' overall, cognitive, affective, and psychomotor learning outcomes as the dependent variables. Overall learning outcomes refer to the comprehensive assessment of students' learning achievements across various domains. It encompasses the cognitive, affective, and psychomotor dimensions of learning. Cognitive learning outcome is the dimension of learning outcomes related to developing students' intellectual abilities, knowledge acquisition, critical thinking skills, problem-solving skills, and academic achievements. Affective learning outcomes pertain to students' emotional and attitudinal development, including their motivation, engagement, attitudes towards learning, and social-emotional skills. Psychomotor learning outcomes refer to developing students' physical and motor skills, coordination, dexterity, and ability to perform practical tasks.

Instrumentation

This research included two data collection devices: the School Inputs Questionnaire (SIQ) and the Educational Outcomes Questionnaire (EOQ). The researchers created new instruments because none already existed that had suitable psychometric properties for measuring the variables of this study. The items in both instruments were based on previous studies (e.g., Bassey et al., 2019; Lili et al., 2018; Odigwe, 2020; Robert & Owan, 2019), theories/ models (such as "human capital development theory" by Schultz, 1961; "contemporaneous educational production model by Coleman et al. (1966) and ideas from consulted field experts. Based on principals' perspectives, the SIQ was designed to measure school facilities, curriculum and administrative expenses. The EOQ was designed to assess students' affective and psychomotor learning outcomes and was administered to the SS2 students. On the other hand, cognitive learning outcome was measured using the average sessional results of the students, which were expressed as percentage estimates of their scores. The SIQ and EOQ instruments utilised a six-point Likert scale, ranging from Very Strongly Agree (VSA) to Very Strongly Disagree (VSD), to capture the participants' responses. The choice of a six-point Likert scale was based on the unique nature of the measured variables. The SIQ consisted of 21 items, while the EOQ comprised 30 items.

Validity and reliability

The draft copy of the research instrument underwent a thorough review process by a panel of experts to assess its face and content validity. The panel included two experts specialising in the Economics of Education, two experts in Measurement and Evaluation, and one expert in Educational Psychology from the University of Calabar. Additionally, the quantitative validity of the instrument was evaluated by a group of ten experts, consisting of four experts in Measurement and Evaluation and six experts in Educational Management. These experts were asked to rate the relevance and clarity of the items related to the measured domains using a 1-4 scale, where higher scores indicated greater relevance or clarity. The ratings were used to calculate the instrument's Content Validity Index (CVI).

The Content Validity Index (CVI) for the SIQ was assessed at both the item and scale levels. The item-level CVI scores for relevance and clarity of the SIQ ranged from 0.8 to 1 and 0.9 to 1, respectively. The scale-level CVI scores for relevance and clarity of the SIQ were 0.98 and 0.98, respectively. Similarly, the item-level and scale-level CVIs for the EOQ were evaluated, with relevance scores ranging from 0.9 to 1 and clarity scores ranging from 0.8 to 1. The scale-level CVI scores for relevance and clarity of the EOQ were 0.99 and 0.98, respectively. It is important to note that different researchers have established standards for revising, dropping, and retaining items, and for validation by ten experts, the acceptable CVI threshold typically falls within the range of 0.78 to 0.83. Since all the item-level CVIs (I-CVIs) and scale-level CVIs (S-CVIs) in this study were within this range, the instrument can be considered valid.

Additionally, a pilot test was conducted involving 110 school leaders, 50 principals and 60 vice principals, and 412 SS2 students from non-participating schools. This pilot test aimed to assess the dimensionality and factorial validity of the research instruments. The Cronbach alpha approach was employed to evaluate the instruments' reliability. The results of this test indicated that the internal consistency reliability coefficients of the items in the SIQ ranged from .77 to .90. In contrast, those of the EOQ ranged from .71 to .90. Further details regarding these findings can be found in the results section of the study.

Procedure for data collection and analysis

The researchers, along with the assistance of four trained research assistants, physically administered the instruments to the participants. Prior permission was obtained from the school leaders, including principals and vice principals, who were provided with a clear explanation of the research purpose. A letter requesting their consent for participation in the study was given to all respondents. Participants who were willing to participate were encouraged to respond sincerely to the items in the instruments. The instruments were then distributed to the selected school leaders and students involved in the study. The respondents were given three days to complete and return the questionnaires. The researchers and research assistants visited the schools again to collect the completed copies of the instruments. Only the questionnaire copies that were properly filled out and retrieved were utilised for data analysis in the study.

The items in each questionnaire were categorised based on the specific research variables they were designed to measure. A scoring system was established for both instruments, ranging from 6 to 1 for positively worded items, while negatively worded items were reverse scored. A coding schedule was developed to guide the scoring and coding of responses, which were then entered into a spreadsheet using a spreadsheet package. The scores for each respondent on the respective sub-scales were summed and recorded in the research project's prepared spreadsheet. As for the assessment of cognitive learning outcomes, the average sessional results of each student per school were used, and these average scores were entered in the appropriate column of the spreadsheet. Descriptive statistics were employed to analyse the demographic data of the respondents. However, to address the research question and test the previously stated hypothesis, multiple hierarchical linear regression analysis was conducted.

Results

Exploratory Factor Analysis (EFA)

The dimensionality and structure of the instruments were analysed using Principal Axis Factoring (PAF). During the extraction process, the Promax rotation method was applied. Factors were selected based on Eigenvalues greater than 1, and items with loadings below .40 were eliminated. The correlation matrix determinant value for the School Input Questionnaire (SIQ) exceeded the criterion value of .00001, indicating the absence of multicollinearity among the items in the matrix. However, one problematic item (SI18) was identified, which loaded exclusively onto factor 4 and did not correlate with any other item. After removing this item, the PAF was rerun using the same parameters. The results revealed that three factors accounted for 51.19% of the total variance in the data. The sample size of 110 school leaders was considered adequate for factor analysis, as indicated by a Kaiser-Meyer-Olkin (KMO) measure of .83. The Bartlett's test of sphericity, which assesses the correlation between variables, was significant at the .05 level with a Chi-Square value of 834.69 and 153 degrees of freedom, indicating that the variables were not redundant (Owan et al., 2021). The three factors were retained as they aligned with the study's theoretical framework. The pattern matrix was also examined to illustrate the relationship between each item and the latent factors (see Table 1).

The dimensionality test of the Educational Outcomes Questionnaire (EOQ) was also based on principal axis factoring. The complete EFA procedure for this questionnaire and its results can be found in two already published works from this project (see Ekpenyong et al., 2022; Owan et al., 2022).

Table 1: Factor Analysis of the School Input Questionnaire (SIQ) Structure.

Factor	Label	Items	$\bar{X}$	SD	Factor loadings	
					EFA	CFA
1	School facilities (Variance explained: 20.48%; Cronbach Alpha reliability: .90)	SI6	3.69	1.76	.82	.81
		SI9	3.75	1.68	.78	.77
		SI5	3.83	1.65	.78	.76
		SI7	3.60	1.64	.76	.77
		SI8	3.58	1.68	.76	.77
		SI4	3.42	1.68	.76	.76
2	School Curriculum (Variance explained: 18.25%; Cronbach Alpha reliability: .87)	SI21	3.24	1.71	.79	.79
		SI22	3.40	1.75	.78	.79
		SI19	3.35	1.66	.76	.75
		SI16	3.28	1.75	.69	.70
		SI20	3.44	1.73	.69	.69
		SI17	3.38	1.71	.69	.68
3	Administrative Expenses (Variance explained: 12.46%; Cronbach Alpha reliability: .77)	SI13	3.79	1.74	.68	.67
		SI12	3.60	1.70	.64	.63
		SI10	3.54	1.66	.64	.65
		SI15	3.52	1.81	.61	.60
		SI11	3.52	1.79	.53	.55
		SI14	3.58	1.77	.52	.53
	Instrument Total	Kaiser-Meyer-Olkin (KMO) = .83 Bartlett's Test of Sphericity at 153 df = 834.69, p < .05 Cronbach Alpha = .77 Corr. Det. Matrix = .000				

Confirmatory Factor Analysis

The confirmatory factor analysis (CFA) was performed using the Maximum Likelihood (ML) estimation technique to assess the measurement capability of the items in capturing their respective latent constructs. This CFA confirmed the findings from the earlier exploratory factor analysis (EFA) conducted and provided additional validation. Table 1 presents the CFA and EFA results, while Figure 2 depicts the CFA model for the School Input Questionnaire (SIQ). The CFA model for the EOQ can be referenced from two previous works by Ekpenyong et al. (2022) and Owan et al. (2022).

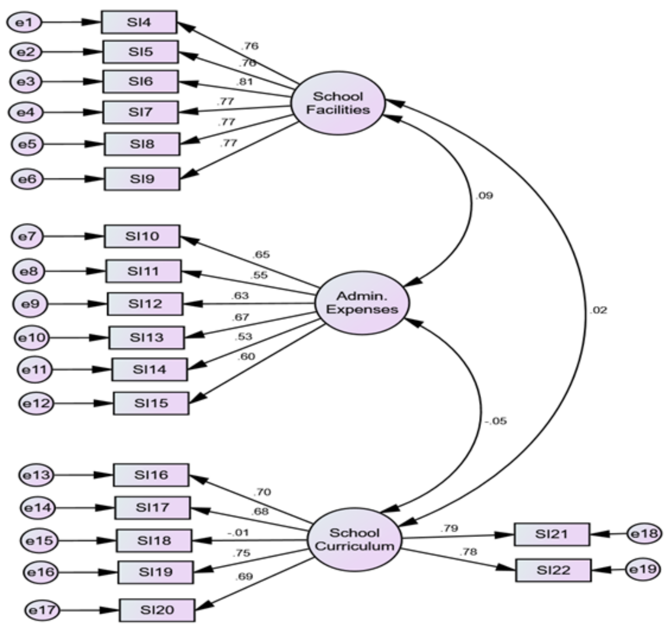


Figure 2: Standardised Latent-Trait CFA Model of the School Input Questionnaire (SIQ).

Eight fit indices were utilised to establish the adequacy of the CFA model and determine its acceptance. These include "Chi-Square", "Comparative Fit Index" (CFI), "Goodness of Fit Index" (GFI), "HOELTER's Critical N", "Incremental Fit Index" (IFI), "Normed Fit Index" (NFI), "Relative Fit Index" (RFI), "Root Mean Square Error of Approximation" (RMSEA), and "Tucker-Lewis Index" (TLI). The specific details about each of these indices are already documented in the literature (see Brown, 2015; Hooper et al., 2008; Hu & Bentler, 1999; Owan et al., 2021). Multiple fit indices were employed to compensate for each index's complementary strengths and weaknesses, in line with instrument validation research recommendations. For example, Kline (2005) suggested utilising a minimum of four fit indices ($\chi2$, RMSEA, CFI, and SRMR) to determine the acceptance of a CFA model.

Assessing the goodness of fit information reveals that the SIQ model met the criteria for Chi-Square ($\chi2 = 144.69$, df = 149, p = .59 > .05), IFI (1.00 > .95), TLI (1.00 > .95), CFI (1.00 > .95) and RMSEA (.00 < .08) but did not meet the GFI (.88 < .95), NFI (.84 < .95) and RFI criteria (.82 < .05). However, the model was retained as it met the criteria of most of the fit indices. Besides, the GF1, NFI and RFI values were all approaching 1.00, and values closer to 1.00 have been suggested to indicate a good fit (Hooper et al., 2008). The EOQ met almost all the criteria except the Chi-Square criteria. More information about the EOQ model can be found in the two published reports (Ekpenyong et al., 2022; Owan et al., 2022). All models generally met the requirements of at least four assessment indices. The TLI and CFI supported earlier research findings as less vulnerable to sample size concerns. These justifications and the fact that all retained items had acceptable factor loadings led to the acceptance of both models. The instruments were deemed suitable for data collection—not only for this study but also for future researchers.

Relative and composite contributions to students' learning outcomes

The findings in Table 2 demonstrate the impact of different factors on students' overall learning outcomes. In the initial model, school facilities accounted for 21% of the variance in students' learning outcomes. However, when administrative expenses were introduced in model 2, the contribution of school inputs increased to 30%, resulting in a significant R2 change of 9%. In model 3, the contribution of the school curriculum was added to that of model 2, further raising the contribution of school inputs to 37%, with an additional R2 change of 7%. This indicates that school facilities contributed 21%, administrative expenses contributed 9%, and the school curriculum contributed 7% to the overall variance in students' learning outcomes. Collectively, these three predictors explain 37% of the variance in students' overall learning outcomes in secondary schools. However, it is important to note that 63% of the variance remains unexplained and may be attributed to other variables not considered in this study. Among the predictors, school facilities had the most significant impact on students' overall learning outcomes (ΔF [1, 868] = 230.15, p < .05), followed by administrative expenses (ΔF [1, 867] = 105.62, p < .05), and the school curriculum (ΔF [1, 866] = 95.39, p < .05).

Table 2 reveals the contributions of different factors to students' cognitive learning outcomes. In the first model, school facilities accounted for 21% of the variance in cognitive outcomes. When administrative expenses were introduced in model 2, the contribution of school inputs increased by 9%, resulting in a total contribution of 29%. Model 3 included the school curriculum, which brought about a 6% change, raising the composite contribution of school inputs to 36% from the 29% recorded in model 2. This indicates that school facilities, administrative expenses, and the curriculum contributed 21%, 9%, and 6%, respectively, to students' cognitive learning outcomes. Furthermore, Table 2 demonstrates that the composite contribution of the three predictors (school inputs) to the total variance in cognitive learning outcomes was 36%, leaving 64% of the variance unaccounted for and attributable to other variables not considered in model 3. Among the predictors, school facilities had the most significant impact on students' cognitive learning outcomes (ΔF [1, 868] = 226.13, p < .05), followed by administrative expenses (ΔF [1, 867] = 104.43, p < .05), and the school curriculum (ΔF [1, 866] = 94.19, p < .05).

Regarding affective learning outcomes, according to Table 2, school facilities accounted for 27% of the variance in students' outcomes in model 1. With the introduction of administrative expenses in model 2, there was a 9% change, leading to a total contribution of 36% from the initial 27%. In model 3, including the school curriculum resulted in an 8% change, bringing the composite contribution of the three predictors to 44% from the 36% recorded in model 2. Therefore, school facilities, administrative expenses, and the curriculum contributed 27%, 9%, and 8%, respectively, to students' affective learning outcomes. Furthermore, Table 2 indicates that the composite contribution of the three predictors (school inputs) to the total variance in affective learning outcomes was 44%. In contrast, the remaining 56% of the variance was unaccounted for and attributable to other predictors not included in model 3. Among the predictors, school facilities had the highest predictive power for students' affective learning outcomes (ΔF [1, 868] = 321.63, p < .05), followed by administrative expenses (ΔF [1, 867] = 127.50, p < .05), and the school curriculum (ΔF [1, 866] = 120.33, p < .05).

In terms of psychomotor learning outcomes, as presented in Table 2, school facilities accounted for 27% of the variance in model 1. With the inclusion of administrative expenses in model 2, there was a 10% increase, resulting in a total contribution of 37% from the initial 27%. In model 3, adding the school curriculum led to a 7% shift, raising the composite contribution of the three predictors to 44% from the 37% recorded in model 2. Therefore, school facilities, administrative expenses, and the curriculum contributed 27%, 10%, and 7%, respectively, to students' psychomotor learning outcomes. Furthermore, Table 2 indicates that the composite contribution of the three predictors (school inputs) to the total variance in psychomotor learning outcomes was 44%, while the remaining 56% was unaccounted for and attributed to factors not included in model 3. Among the predictors, school facilities had the highest predictive power for students' psychomotor learning outcomes (ΔF [1, 868] = 318.95, p < .05), followed by administrative expenses (ΔF [1, 867] = 134.54, p < .05), and the school curriculum (ΔF [1, 866] = 114.84, p < .05).

Table 2: Hierarchical regression analysis of the relative prediction of school facilities, administrative expenses and curriculum on students’ learning outcomes.

Criterion variables	Model	R	R^2	$Adj.\ R^2$	SE	ΔR^2	ΔF	df_1	df_2	$p\Delta F$
Overall	1	.46[a]	.21	.21	182.65	.21	230.15	1	868	.000
Learning	2	.54[b]	.30	.29	172.55	.09	105.62	1	867	.000
Outcomes	3	.60[c]	.37	.36	163.86	.07	95.39	1	866	.000
Cognitive	1	.46[a]	.21	.21	175.47	.21	226.13	1	868	.000
Learning	2	.54[b]	.29	.29	165.87	.09	104.43	1	867	.000
Outcomes	3	.60[c]	.36	.36	157.61	.07	94.19	1	866	.000
Affective	1	.52[a]	.27	.27	4.65	.27	321.63	1	868	.000
Learning	2	.60[b]	.36	.36	4.34	.09	127.50	1	867	.000
Outcomes	3	.66[c]	.44	.44	4.07	.08	120.33	1	866	.000
Psychomotor	1	.52[a]	.27	.27	2.74	.27	318.95	1	868	.000
Learning	2	.61[b]	.37	.37	2.55	.10	134.54	1	867	.000
Outcomes	3	.66[c]	.44	.44	2.40	.07	114.84	1	866	.000

a Predictors: (Constant), School facilities
b Predictors (Constant): School facilities, administrative expenses
c Predictors (Constant): School facilities, administrative expenses, School curriculum

Hypothesis testing: Composite contributions

The ANOVA results, as presented in Table 3, were utilised to test the hypothesis of this study at a significance level of .05. It is revealed that school facilities made a significant contribution to students’ overall learning outcomes in model 1, with $F(1, 868) = 230.15$, $p < .05$. In model 2, there was a significant composite contribution of school facilities and administrative expenses to students’ overall learning outcomes, with $F(2, 867) = 181.76$, $p < .05$. Similarly, in model 3, a significant composite contribution of school facilities, administrative expenses, and school curriculum to students’ overall learning outcomes was observed, with $F(3, 866) = 166.16$, $p < .05$. These results provide partial support for the alternative hypothesis, indicating that the predictors have a significant impact on students’ overall learning outcomes. Conversely, the null hypothesis is rejected concerning the composite contribution of the three predictors to students’ overall learning outcomes.

According to Table 3, significant contributions were observed concerning students’ cognitive learning outcomes. In model 1, school facilities made a significant contribution, with $F(1, 868) = 226.13$, $p < .05$. Model 2 indicated a significant composite contribution of school facilities and administrative expenses to students’ cognitive learning outcomes, with $F(2, 867) = 178.75$, $p < .05$. Furthermore, in model 3, a significant composite contribution of school facilities, administrative expenses, and school curriculum to students' cognitive learning outcomes was observed, with $F(3, 866) = 163.37$, $p < .05$. As a result, the null hypothesis, which pertained to the composite contribution of the three predictors on students' cognitive learning outcomes, was rejected. Conversely, the alternative hypothesis was supported, suggesting that these predictors significantly influence students’ cognitive learning outcomes.

According to Table 3, the contribution of school facilities to students’ affective learning outcomes was found to be statistically significant in model 1, with $F(1, 868) = 321.63$, $p < .05$. In model 2, the composite contribution of school facilities and administrative expenses to students' affective learning outcomes was also significant, with $F(2, 867) = 248.00$, $p < .05$. Similarly, in model 3, the composite contribution of school facilities, administrative expenses, and school curriculum to students’ affective learning outcomes was significant, with $F(3, 866) = 228.20$, $p < .05$. Based on the evidence presented in Table 3, the null hypothesis was rejected, which presumably suggested no significant contribution of the three predictors to students' affective learning outcomes. Instead, the alternative hypothesis was supported, indicating a significant contribution of these three predictors to students’ affective learning outcomes.

According to Table 3, the contribution of school facilities to students’ psychomotor learning outcomes in model 1 was found to be statistically significant, with $F(1, 868) = 318.95$, $p < .05$. In model 2, the composite contribution of school facilities and administrative expenses to students’ psychomotor learning outcomes was also statistically significant, with $F(2, 867) = 251.00$, $p < .05$. Furthermore, in model 3, the composite contribution of school facilities, administrative expenses, and school curriculum to students’ psychomotor learning outcomes was found to be statistically significant, with $F(2, 867) = 227.00$, $p < .05$. Based on the results presented in Table 3, the null hypothesis, suggesting no significant contribution of the three predictors to students’ psychomotor learning outcomes, was rejected. Conversely, the alternative hypothesis was supported, which proposed a significant composite contribution of school facilities, administrative expenses, and school curriculum to students’ psychomotor learning outcomes.

Table 3: ANOVA results of hierarchical regression analysis on the composite prediction of school facilities, administrative expenses and curriculum on students’ learning outcomes.

Criterion variable	Model	Source	SS	Df	MS	F	P
Overall learning outcomes	1	Regression	7678024	1	7678024	230.15	.00[b]
		Residual	28956775	868	33360.34		
		Total	36634799	869			
	2	Regression	10822445	2	5411223	181.76	.00[c]
		Residual	25812353	867	29772.03		
		Total	36634799	869			
	3	Regression	13383631	3	4461210	166.16	.00[d]
		Residual	23251167	866	26848.92		
		Total	36634799	869			
Cognitive skills	1	Regression	6962332	1	6962332	226.13	.00[b]
		Residual	26725221	868	30789.43		
		Total	33687552	869			
	2	Regression	9835343	2	4917672	178.75	.00[c]
		Residual	23852209	867	27511.2		
		Total	33687552	869			
	3	Regression	12175183	3	4058394	163.37	.00[d]
		Residual	21512370	866	24841.07		
		Total	33687552	869			
Non-Cognitive skills	1	Regression	6951.83	1	6951.83	321.63	.00[b]
		Residual	18761.08	868	21.614		
		Total	25712.91	869			
	2	Regression	9357.102	2	4678.551	248.00	.00[c]
		Residual	16355.81	867	18.865		
		Total	25712.91	869			
	3	Regression	11352.48	3	3784.161	228.20	.00[d]
		Residual	14360.42	866	16.582		
		Total	25712.91	869			
Practical skills	1	Regression	2389.589	1	2389.589	318.947	.00[b]
		Residual	6503.153	868	7.492		
		Total	8892.741	869			
	2	Regression	3263.181	2	1631.59	251.279	.00[c]
		Residual	5629.561	867	6.493		
		Total	8892.741	869			
	3	Regression	3922.329	3	1307.443	227.797	.00[d]
		Residual	4970.413	866	5.74		
		Total	8892.741	869			

b Predictors (Constant): School facilities
c Predictors (Constant): School facilities, administrative expenses
d Predictors (Constant): School facilities, administrative expenses, School curriculum

Hypothesis testing: Relative contributions

According to Table 4, school facilities, administrative expenses, and curriculum individually contributed significantly to students’ overall cognitive, affective, and psychomotor learning outcomes. However, in model 3, the relative contribution of school facilities to students’ cognitive learning outcomes was insignificant. As a result, the alternative hypothesis was supported, indicating that school facilities have a significant relative contribution to students’ overall affective and psychomotor learning outcomes. On the other hand, the null hypothesis regarding the relative contribution of administrative expenses and school curriculum to students' overall cognitive, affective, and psychomotor learning outcomes was rejected. This suggests that both administrative expenses and school curriculum have significant relative contributions to students' overall, affective, and psychomotor learning outcomes. However, it's important to note that the null hypothesis was not rejected for the relative contribution of school facilities to students' cognitive learning outcomes, indicating that the impact of school facilities on cognitive learning outcomes may not be statistically significant.

Table 4: Specific prediction of school facilities, administrative expenses and curriculum on students' learning outcomes.

Criterion variables	Models	Predictors	β	*t*	*SE*	*p*
Students' Overall Learning Outcomes	1	School facilities	.46	15.17	2.65	.00
	2	School facilities	.19	4.83	3.41	.00
		Administrative expenses	.40	10.28	3.24	.00
	3	School facilities	.08	1.97	3.39	.05
		Administrative expenses	.20	4.84	3.50	.00
		School curriculum	.39	9.77	2.98	.00
Cognitive Learning Outcomes	1	School facilities	.46	15.04	2.55	.00
	2	School facilities	.19	4.76	3.28	.00
		Administrative expenses	.40	10.22	3.12	.00
	3	School facilities	.07	1.91	3.26	.06
		Administrative expenses	.20	4.81	3.37	.00
		School curriculum	.39	9.71	2.86	.00
Affective Learning Outcomes	1	School facilities	.52	17.93	0.07	.00
	2	School facilities	.24	6.45	0.09	.00
		Administrative expenses	.42	11.29	0.08	.00
	3	School facilities	.12	3.32	0.08	.00
		Administrative expenses	.21	5.34	0.09	.00
		School curriculum	.41	10.97	0.07	.00
Psychomotor Learning Outcomes	1	School facilities	.52	17.86	0.04	.00
	2	School facilities	.23	6.23	0.05	.00
		Administrative expenses	.43	11.60	0.05	.00
	3	School facilities	.11	3.16	0.05	.00
		Administrative expenses	.23	5.72	0.05	.00
		School curriculum	.40	10.72	0.04	.00

Discussion

This study quantified the degree to which school facilities, administrative expenses, and curriculum cumulatively and relatively predict students' overall cognitive, affective and psychomotor learning outcomes in public secondary schools. The results showed that improving the combined provision of quality school facilities, administrative expenses, and curriculum predicts students' overall learning outcomes. The result aligns with earlier studies that the availability and layout of school physical facilities can promote students' learning outcomes (Ariani, 2015; Daramola et al., 2017). Furthermore, the study corroborates earlier evidence that school spending was associated with students' academic achievement (Hægeland et al., 2012; Nicoletti & Rabe, 2018). The result also strengthens the findings of Demir et al. (2012) that the school curriculum is important for increasing students' learning efficiency and achievement. This result is crucial for school administrators looking to promote secondary education goal attainment through the overall development of students for progression towards higher education, good living, and economic advancement.

The study also found that factors such as the quality of school facilities, administrative expenses, and the school curriculum content significantly affect students' cognitive learning outcomes in public secondary schools. This result suggests that the quality of school facilities, the cost of administrative expenses, and the content of the school curriculum all play important roles in deciding how secondary school students can meaningfully think, comprehend, apply, analyse, synthesise and evaluate information. This result agrees with other studies (e.g., Munda & Odebero, 2014; Webber, 2012) that documented a significant positive correlation between educational expenditure and students' academic achievement. Furthermore, another study found significant improvement in students' cognitive outcomes due to exposure to quality curriculum content compared to control group students (Senk, 2020). This may be of interest to educators, policymakers, and others who are concerned with improving the academic performance and outcomes of students in secondary schools. It is also useful for finding areas where improvements can be made to better support students' learning and development.

The results showed a strong, combined positive prediction of school facilities, administrative expenses, and curriculum on students' affective learning outcomes in public secondary schools. This result suggests that students in schools with good facilities, prudent administrative expenditures, and a well-designed curriculum are likely to have better affective learning outcomes (such as increased motivation, engagement, and enjoyment of learning) compared to those in schools with poorer facilities and wasteful administrative expenditures, and a poorly designed curriculum. This result is not surprising since the effective provision of school facilities, proper management of school funds, and curriculum development can be useful in shaping students' values, characters, attitudes and behaviours. The finding agrees with some earlier studies that the availability of school facilities is related to students' motivation (Sidi, 2019) and learning outcomes (Arshad et al., 2018; Takwate, 2018). The findings of this study also align with an earlier study which documents an important link between administrative expenditures and students' learning outcomes (Strickland, 2021). Moreover, an earlier study also documented that school curricula changed students' attitudes (Cantoni et al., 2017). The implications of these findings are relevant for educational stakeholders to invest more in supplying school facilities and curriculum development. The results can also be useful for school principals to minimise the misuse and wastage of school funds.

This study also revealed that school facilities, administrative expenses and curriculum cumulatively predict students' psychomotor learning outcomes in public secondary schools. This finding implies that secondary students in schools with adequate facilities, reasonable administrative expenses and quality curricula tend to buy more skills and develop competencies for perception, adaptation, origination, creation and innovation than their school counterparts without such provisions. The result is explainable since schools with better facilities, lower administrative expenses, and more effective curricula may successfully promote practical teaching and learning. This result supports an earlier study (Ni et al., 2011) that discovered that implementing the school curriculum improved students' cognitive and psychomotor attributes, such as routine and complex problem-solving skills. The result also strengthens the findings of earlier studies in China (Cantoni et al., 2017) and South Africa (Dhunpath & Subbaye, 2018).

In terms of individual contributions, this study revealed that school facilities significantly predicted students' overall, affective and psychomotor learning outcomes. However, school facilities did not significantly predict the cognitive dimension of learning outcomes in public secondary schools. This result implies that students who learn in better facilities are more engaged and motivated, leading to better affective and psychomotor learning outcomes. On the other hand, the quality of school facilities may not be as important for cognitive learning outcomes, which may depend more on other factors such as teaching effectiveness, curriculum quality, and students' cognitive abilities and prior knowledge. This result agrees with other studies that how well a child has learnt is reflected in their cognitive, affective and psychomotor abilities (Akhiruyanto et al., 2022; Orak et al., 2020; Robert & Owan, 2019). Similarly, other researchers found an important link between the availability of school facilities and students' motivation (Sidi, 2019). Nevertheless, further research is needed to understand why school facilities do not significantly predict students' cognitive learning outcomes.

It was also proved that administrative expenses significantly predicted students' learning outcomes holistically and across all dimensions. Thus, higher administrative expenses are related to better learning outcomes for students in public secondary schools. This relationship extends to multiple dimensions of learning outcomes, including cognitive (related to knowledge and understanding), affective (related to emotions and attitudes), and psychomotor (related to physical skills and movements). This evidence supports a long list of studies (e.g., Hægeland et al., 2012; Munda & Odebero, 2014; Nicoletti & Rabe, 2018; Webber, 2012) reporting a substantial correlation between administrative expenditures and students' learning outcomes. It is not uncommon for research to find that certain factors, such as resources and funding, can significantly impact student learning outcomes. In this case, administrative expenses may play a particularly important role. It is worth noting that the nature and extent of this relationship may vary depending on the specific context in which the research was conducted.

Further research may be needed to understand the mechanisms behind this relationship fully. One potential implication of these findings is that schools and educational institutions may want to consider increasing prudent expenditure of school finances and minimising waste to improve students' learning outcomes. This agrees with the findings of Mbon et al. (2020) that the wastage of school resources was associated with poor school effectiveness in promoting teaching and learning.

Lastly, this study documented a significant positive prediction of school curriculum on students' overall cognitive, affective and psychomotor learning outcomes in public secondary schools. This suggests that the curriculum used in these schools is effective in helping students learn and achieve positive outcomes in various areas. The result aligns with an earlier study that reported improvements in students' academic and social skills due to improvements in the curriculum contents and experiences (Alismail & McGuire, 2015). Other studies also reported that students showed higher learning achievements and motivation to learn when exposed to an improved school curriculum (Alnajjar, 2022; Bouck & Joshi, 2012; Ni et al., 2011). Overall, the findings of this study highlight the importance of having a well-designed and effective curriculum in promoting positive learning outcomes for students. This is especially important in public secondary schools, as these schools serve a diverse population and play a critical role in preparing students for higher education and the workforce. It would be interesting to explore further the specific aspects of the curriculum that contributed to the positive learning outcomes observed in this study, as well as to examine the potential long-term effects of this curriculum on students' academic and professional success.

Constraints and direction for future research

There are several limitations to this study. First, the study was conducted in public secondary schools, so the results may not be generalisable to other types of schools, such as private or primary schools. Therefore, conducting a similar study in different types of schools, such as private or primary schools, will be important to determine if the findings can be generalised to these settings. Second, the study only looked at the prediction of these factors on learning outcomes and did not assess their actual causal relationship. Future studies may consider using experimental or quasi-experimental designs to assess the causal relationship between the predictor variables and students' learning outcomes. Third, the study was conducted in a specific geographic location in Nigeria, so the results may not apply to other regions with different education systems or socio-cultural contexts. Future researchers should consider expanding their scope to other geographic locations to address this limitation and see if the results hold up in different education systems and socio-cultural contexts. Finally, the study was cross-sectional, so whether the observed relationships between the predictor variables and learning outcomes are consistent over time is unclear. Therefore, future longitudinal studies need to be conducted to assess the stability of the relationships between predictor variables and learning outcomes over time.

Conclusion

This study aimed to examine the combined and individual impact of school facilities, administrative expenses, and school curriculum on students' overall cognitive, affective, and psychomotor learning outcomes in public secondary schools. The study's findings provide compelling evidence that these factors play a significant role in shaping students' educational achievements. The results indicate that school facilities significantly contribute to students' overall affective and psychomotor learning outcomes. Additionally, administrative expenses significantly influence students' overall cognitive, affective, and psychomotor learning outcomes. Furthermore, the school curriculum significantly predicts students' overall cognitive, affective, and psychomotor learning outcomes. These findings hold considerable implications for policymakers, educators, and researchers in the field of education. Policymakers can utilise these findings to prioritise investments in school facilities and promote efficient allocation of administrative expenses, leading to improved student learning outcomes. Educators are encouraged to recognise the impact of these factors and optimise school facilities and resources to maximise student achievement. The study emphasises the importance of investing in high-quality school facilities, managing administrative expenses effectively, and implementing a comprehensive curriculum to enhance learning outcomes in public secondary schools. Overall, this study provides valuable insights for educators and policymakers striving to enhance the quality of education in public secondary schools. It also highlights the need for further research to deepen our understanding of the relationship between these factors and student learning outcomes, facilitating the identification of best practices and evidence-based approaches for enhancing student achievement.

Based on the conclusion of this study, it is recommended that school administrators and education authorities should allocate resources and oversee the enhancement of school facilities, including classrooms, libraries, laboratories, and recreational spaces. They should create a conducive learning environment that supports students' overall cognitive, affective, and psychomotor learning outcomes. School administrators and financial managers should thoroughly review administrative expenses to identify inefficiency and reduce unnecessary costs. Resources should be allocated to educational initiatives, student support services, and teacher professional development opportunities. Curriculum development committees and Education Authorities should be responsible for reviewing and updating the school curriculum to ensure it is comprehensive, well-rounded, and aligned with educational standards. The curriculum should incorporate practical and experiential learning activities, critical thinking exercises, and relevant teaching materials to enhance students' overall cognitive, affective, and psychomotor learning outcomes. Teacher training institutions and departments should provide regular professional development opportunities for teachers to enhance their teaching skills and subject knowledge. The focus should be on strategies that promote effective instructional practices, student engagement, and differentiation to improve students' overall cognitive, affective, and psychomotor learning outcomes. Schools, community organisations, and stakeholders should foster collaborative partnerships to support students' learning outcomes. Schools should collaborate with policymakers, parents, and community organisations to provide additional resources, mentorship programs, and extracurricular activities that promote students' holistic development.

References

Adelaja, A., & George, J. (2020). Is youth unemployment related to domestic terrorism? *Perspectives on terrorism, 14*(5), 41-62. https://www.jstor.org/stable/26940038

Ademiluyi, L. F. (2019). Adequacy and utilisation of ICT resources for teaching business subjects in senior secondary schools in Osun State, Nigeria. *African Journal of Teacher Education, 8*, 139–158. https://doi.org/10.21083/ajote.v8i0.4899

Agwu, P., Orjiakor, C. T., Odii, A., Onalu, C., Nzeadibe, C., Roy, P., Onwujekwe, O., & Okoye, U. (2022). "Miracle examination centres" as hubs for malpractices in senior secondary school certificate examination in Nigeria: A systematic review. *International Journal of Educational Development, 88,* 102538. https://doi.org/10.1016/j.ijedudev.2021.102538

Ajayi, I. A., & Yusuf, M. A. (2010). School plants planning and students' learning outcomes in South West Nigerian secondary schools. *International Journal of Educational Sciences, 2*(1), 47–53. https://doi.org/10.1080/09751122.2010.11890000

Akah, L. U., Owan, V. J., Alawa, D. A., Ojie, F. C., Usoro, A. A., Dada, O. A., Olofu, M. A., Ebuara, V. O., Ajigo, I., Essien, E. E., Essien, C. K., Unimna, F. A., Ukpong, J., Adeleke, O. P., & Neji, H. A. (2022). ICT deployment for teaching in the COVID-19 era: A quantitative assessment of resource availability and challenges in public universities. *Frontiers in Education, 7,* Article 920932. https://doi.org/10.3389/feduc.2022.920932

Akhiruyanto, A., Hidayah, T., Yudhistira, D., & Fahmi, H. (2022). Assessment of the learning outcomes of physical education in children with intellectual disabilities. *Journal of Innovation in Educational and Cultural Research, 3*(3), 471-477. https://doi.org/10.46843/jiecr.v3i3.207

Alabdulkarem, A., Alhojailan, M., & Alabdulkarim, S. (2021). Comprehensive investigation of factors influencing university students' academic performance in Saudi Arabia. *Education Sciences, 11*(8), Article 375. https://doi.org/10.3390/educsci11080375

Alabi, O. (2021). Assessing the adequacy of aesthetic nature and maintenance of school grounds of the Lagos state public schools: A step toward sustainable development. *Interdisciplinary Journal of Education, 4*(2), 167-175. https://doi.org/10.53449/ije.v4i2.65

Alazmi, M. S., & Al-Kubaisi, H. S. (2020). School principals' perceptions on the diversification of school financing sources: A study using Delphi method. *Management in Education, 34*(3), 94-106. https://doi.org/10.1177/0892020619897372

Ali, S. (2013). Factors affecting academic achievement of students. *American Journal of Educational Research, 1*(8), 283-289. https://doi.org/10.12691/education-1-8-3

Aliyu, B. B. (2018). Assessment of financial management practices among secondary school principals in Kaduna State. *International Journal of Hospitality & Tourism Management, 2*(2), 22–27. https://doi.org/10.11648/j.ijhtm.20180202.11

Almlund, M., Duckworth, A., Heckman, J. & Kautz, T. (2011). *Personality psychology and economics*. National Bureau of Economic Research. https://doi.org/10.3386/w16822

Anderson, L. W., & Bourke, S. F. (2013). *Assessing affective characteristics in the schools*. Routledge. https://doi.org/10.4324/9781410605443

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's* *taxonomy of educational objectives*. Longman. https://eduq.info/xmlui/handle/11515/18345

Ariani, M. G. (2015). The impact of the physical learning spaces on the learning process. In S. Tang, & L. Logonnathan (Eds), *Taylor's 7th teaching and learning conference 2014 proceedings.* Springer. https://doi.org/10.1007/978-981-287-399-6_16

Arop, F. O., Ekpang, M. A., Nwannunu, B. I., & Owan, V. J. (2018). Personnel management and corrupt academic practices in universities in Cross River State, Nigeria. *International Journal of Economics, Commerce and Management, 6*(9), 405–419. https://doi.org/10.2139/ssrn.3250019

Arshad, M., Haq, M. N. U., & Khan, M. G. (2020). Status of physical facilities and students' achievement at public and PEF partner schools in Punjab, Pakistan. *Global Political Review, 5*(1), 163-171. https://doi.org/10.31703/gpr.2020(V-I).19

Arshad, M., Qamar, Z. A., & Gulzar, F. H. (2018). Effects of physical facilities at public schools on students' achievement in Punjab, Pakistan. *Global Social Sciences Review, 3*(4), 102-113. https://doi.org/10.31703/gssr.2018(III-IV).07

Baber, H. (2020). Determinants of students' perceived learning outcome and satisfaction in online learning during the pandemic of COVID-19. *Journal of Education and e-learning Research, 7*(3), 285-292. https://doi.org/10.20448/journal.509.2020.73.285.292

Bassey, B. A., & Owan, V. J. (2020). Higher-ordered test items as assessment practice in higher education during Pandemics: Implications for effective e-learning and safety. In V. C. Emeribe, L. U. Akah, O. A. Dada, D. A. Alawa, & B. A. Akuegwu (Eds.), *Multidisciplinary issues in health, human kinetics and general education practices* (pp. 395–409). University of Calabar Press. https://bit.ly/31J4ZiI

Bassey, B. A., Owan, V. J., & Agunwa, J. N. (2019). Quality assurance practices and students' performance evaluation in universities of South-South Nigeria: A structural equation modelling approach. *British Journal of Psychology Research, 7*(3), 1–13. https://doi.org/10.5281/zenodo.4458641

Belsito, C. (2016). The importance of 'teacher quality' and 'quality teaching' on academic performance. *Journal of Student Engagement: Education Matters, 6*(1), 28-38. https://ro.uow.edu.au/jseem/vol6/iss1/5

Bervell, B., & Arkorful, V. (2020). LMS-enabled blended learning utilisation in distance tertiary education: Establishing the relationships among facilitating conditions, voluntariness of use and use behaviour. *International Journal of Educational Technology in Higher Education, 17,* Article 6. https://doi.org/10.1186/s41239-020-0183-9

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals - Handbook 1, Cognitive domain.* David McKay Publishing. https://bit.ly/3GY9SpQ

Bruce, M. D., Ermasova, N., & Mattox, L. (2019). The fiscal disparity and achievement gap between extremely wealthy and poor school districts in Illinois. *Public Organization Review, 19*(4), 541–565. https://doi.org/10.1007/s11115-018-0417-7

Buys, M., du Plessis, P., & Mestry, R. (2020). The resourcefulness of school governing bodies in fundraising: Implications for the provision of quality education. *South African Journal of Education, 40*(4), 1–9. https://doi.org/10.15700/saje.v40n4a2042

Chen, Z., Bao, Y., & Zhu, T. (2022). An empirical study on IPO model construction of undergraduate education quality evaluation in China from the statistical pattern recognition approach in NLP. *ACM Transactions on Asian and Low-Resource Language Information Processing, 21*(6), 1-18. https://doi.org/10.1145/3543851

Chernew, M., & Mintz, H. (2021). Administrative expenses in the US health care system: Why so high? *JAMA, 326*(17), 1679-1680. https://doi.org/10.1001/jama.2021.17318

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, F., Mood, A. M., & Weinfeld, F.D. (1966). *Equality of educational opportunity.* US Government Printing Office. https://files.eric.ed.gov/fulltext/ED012275.pdf

Cooper, D., & Higgins, S. (2015). The effectiveness of online instructional videos in the acquisition and demonstration of cognitive, affective and psychomotor rehabilitation skills. *British Journal of Educational Technology, 46*(4), 768-779. https://doi.org/10.1111/bjet.12166

Cunha, C. M. P. D. (2018). Sponsor bias in pension fund administrative expenses: The Brazilian experience. *Brazilian Administration Review, 15*(1), Article e170072. https://doi.org/10.1590/1807-7692bar2018170072

Daramola, D., Olutola, A., & Ogunjimi, M. (2017). Assessing the impact of school environment on academic performance of senior secondary school students in economics. *International Journal of Educational Benchmark, 8*(2), 41-49. https://bit.ly/3k0W7xv

Decius, J., Schaper, N., & Seifert, A. (2021). Work characteristics or workers' characteristics? An input-process-output perspective on informal workplace learning of blue-collar workers. *Vocations and Learning, 14*(2), 285-326. https://doi.org/10.1007/s12186-021-09265-5

Demir, S., Kilinc, M. & Dogan, A. (2012). The effect of curriculum for developing efficient studying skills on academic achievements and studying skills of learners. *International Electronic Journal of Elementary Education, 4*(3), 427-440. https://files.eric.ed.gov/fulltext/EJ1068605.pdf

Dube, P. (2019). Utilisation of instructional media in teaching and learning by secondary school teachers: A case of MKOBA 1 high school. *International Journal of Innovative Science and Research Technology, 4*(9), 409-418. https://bit.ly/3iscW3U

Ekaette, S. O., Owan, V. J., & Agbo, D. I. (2019). External debts and the financing of education in Nigeria from 1988 – 2018:

Implication for effective educational management. *Journal of Educational Realities, 9*(1), 1–14. https://doi.org/10.5281/zenodo.4320606

Ekpenyong, J. A., Owan, V. J., Ogar, J. O., & Undie, J. A. (2022). Hierarchical linear modelling of educational outcomes in secondary schools: What matters – teachers' or administrators' input? *Cogent Education, 9*(1), Article 2133491. https://doi.org/10.1080/2331186X.2022.2133491

Ekpenyong, J. A., Owan, V. J., Mbon, U. F., & Undie, S. B. (2023). Family and community inputs as predictors of students' overall, cognitive, affective and psychomotor learning outcomes in secondary schools. *Journal of Pedagogical Research, 7*(1), 103–127. https://doi.org/10.33902/JPR.202319099

Enneking, K. M., Breitenstein, G. R., Coleman, A. F., Reeves, J. H., Wang, Y., & Grove, N. P. (2019). The evaluation of a hybrid, general chemistry laboratory curriculum: Impact on students' cognitive, affective, and psychomotor learning. *Journal of Chemical Education, 96*(6), 1058-1067. https://doi.org/10.1021/acs.jchemed.8b00637

Eze, E. (2021). Why secondary school geography students perform poorly in external examinations. *Journal of Geography, 120*(2), 51-60. https://doi.org/10.1080/00221341.2020.1860114

Fan, Y., & Liu, X. K. (2017). Misclassifying core expenses as special items: Cost of goods sold or selling, general, and administrative expenses? *Contemporary Accounting Research, 34*(1), 400–426. https://doi.org/10.1111/1911-3846.12234

Gcelu, N., Padayachee, A. S., & Sekitla, D. M. (2020). Management of indiscipline among secondary school students in Ilembe District, KwaZulu-Natal, South Africa. *African Journal of Gender, Society & Development, 9*(4), Article 139. https://hdl.handle.net/10520/ejc-aa_jgida1-v9-n4-a7

Gigliotti, P., & Sorensen, L. C. (2018). Educational resources and student achievement: Evidence from the save harmless provision in New York State. *Economics of Education Review, 66*, 167–182. https://doi.org/10.1016/j.econedurev.2018.08.004

Hægeland, T., Raaum, O., & Salvane, K. G. (2012). Pennies from heaven? Using exogenous tax variation to identify effects of school resources on pupil achievement. *Economics of Education Review, 31*(5), 601-614. https://doi.org/10.1016/j.econedurev.2012.03.004

Hayes, S. D., & Burkett, J. R. (2021). Almost a principal: Coaching and training assistant principals for the next level of leadership. *Journal of School Leadership, 31*(6), 502-525. https://doi.org/10.1177/1052684620912673

Herbaut, E., & Geven, K. (2020). What works to reduce inequalities in higher education? A systematic review of the (quasi-) experimental literature on outreach and financial aid. *Research in Social Stratification and Mobility, 65*, Article 100442. https://doi.org/10.1016/j.rssm.2019.100442

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*(1), 53-60. https://doi.org/10.21427/D7CF7R

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, X., Sun, S., & Law, R. (2021). A reflection of core marketing subjects in e-hospitality programmes: The IPO Model. *Journal of Quality Assurance in Hospitality & Tourism, 22*(3), 336-344. https://doi.org/10.1080/1528008X.2020.1774033

Huntoon, M. (2021). *A study to examine the effect of School Improvement Grant (SIG) funds on student performance at seven California K-8 schools.* (Doctoral Dissertation), Concordia University Irvine. http://hdl.handle.net/11414/3485

Ilie, S., & Rose, P. (2016). Is equal access to higher education in South Asia and sub-Saharan Africa achievable by 2030? *Higher Education, 72*(4), 435-455. https://doi.org/10.1007/s10734-016-0039-3

Islam, A. A., Gu, X., Crook, C., & Spector, J. M. (2020). Assessment of ICT in tertiary education applying structural equation modelling and Rasch model. *SAGE Open, 10,* Article 2158244020975409. https://doi.org/10.1177/2158244020975409

Issacar, N., & Hesbon, A. O. (2021). Instructional learning materials' use and students' academic outcomes in private secondary schools in Rwanda: A case study of Nyarugenge district. *Journal of Education, 4*(7), 76-93. https://doi.org/10.53819/810181025026

Kazima, M., Longwe, J., & Gobede, F. (2022). Improving teaching and learning mathematics in Malawi primary schools: A review of reforms, interventions, successes and challenges. *Journal of Applied Learning & Teaching, 5*(Sp. Iss. 2), 81–89. https://doi.org/10.37074/jalt.2022.5.S2.3

Keddie, A., & Holloway, J. (2020). School autonomy, school accountability and social justice: Stories from two Australian school principals. *School Leadership & Management, 40*(4), 288-302. https://doi.org/10.1080/13632434.2019.1643309

Kline, R. B. (2016). *Principle and practice of structural equation modelling* (4th ed). The Guilford Press. https://psycnet.apa.org/record/2015-56948-000

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice, 41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

Lawanson, O. A., & Gede, N. T. (2011). Provision and management of school facilities for the implementation of the UBE programme. *Journal of Educational and Social Research, 1*(4), 47-47. https://bit.ly/3CCHWVM

Levin, H. M. (2012). More than just test scores. *Prospects, 42*(3), 269-284. https://doi.org/10.1007/s11125-012-9240-z

Lili, L., Hongyu, G., & Rozelle, S. (2018). The correlation of expenditure on school level and students' academic performance: Based on the empirical study in western poor rural China. *Journal of East China Normal University (Educational Sciences), 36*(6), 100-106. https://doi.org/10.16382/j.cnki.1000-5560.2018.06.011

Lipnevich, A. A., MacCann, A., & Roberts, R. D. (2013). *Assessing non-cognitive constructs in education: A review of traditional and innovative approaches.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199796304.013.0033

Matthews, K. E., & Mercer-Mapstone, L. D. (2018). Toward curriculum convergence for graduate learning outcomes: Academic intentions and student experiences. *Studies in Higher Education, 43*(4), 644-659. https://doi.org/10.1080/03075079.2016.1190704

Mbah, A. N., & Onuora, J. K. J. (2018). Effect of internally generated revenue on infrastructural development of South-East States of Nigeria. *IIARD International Journal of Economics and Business Management, 4*(7), 1–10. https://bit.ly/3k86n7d

Mbon, U. F., Ukpabio, G. E., Ekanem, E. E., Okon, J. E., Uko, E. S., Ngaji, M. N., & Okon, E. E. (2020). Wastage of school material resources and secondary school system effectiveness: Evidence from a survey in Nigeria. *Humanities and Social Sciences Letters, 8*(3), 264–279. https://doi.org/10.18488/journal.73.2020.83.264.279

Megbo, B. C., & Saka, A. (2015). Assessment of secondary school curriculum in Nigeria. *International Journal of Multidisciplinary Research and Development, 2*(9), 141-145. http://www.allsubjectjournal.com/download/1241/2-9-14.pdf

Munda, S. W., & Odebero, S. (2014). The influence of education costs on students' academic performance in Kenya: An empirical study of Bungoma County secondary schools. *Asian Journal of Educational Research, 2*(1), 1-11. https://bit.ly/3GTXk2A

Nathan, S. S., Berahim, M., & Ramle, R. (2017). Rubric for measuring psychomotor and affective learning domain. *Pertanika Journal of Social Science & Humanities, 25*, 101-108. https://bit.ly/3ivCFbQ

Nghambi, G. H. (2015). *Factors contributing to poor academic performance in certificate of secondary education examination for community secondary schools in Urambo District, Tabora, Tanzania.* (Unpublished M.Ed Thesis), Open University of Tanzania. https://core.ac.uk/download/pdf/44684818.pdf

Ngware, M. W., & Mutisya, M. (2022). Math pedagogical practices in Kenya and Uganda, and their implications to learning in sub-Saharan Africa. *Journal of Applied Learning & Teaching, 5*(Sp. Iss. 2), 51–58. https://doi.org/10.37074/jalt.2022.5.S2.6

Nicoletti, C. & Rabe, B. (2018). The effect of school spending on student achievement: Addressing biases in value-added models. *Journal of the Royal Statistical Society: Series A, 181,* 487–515. https://doi.org/10.1111/rssa.12304

Nurabadi, A., Bafadal, I., Priyatni, E. T., Juharyanto, & Gunawan, I. (2020). Analysis of the availability of school facilities and infrastructure as an effort to accelerate school quality improvement. *Proceedings of the 6th International Conference on Education and Technology* (ICET 2020). https://doi.org/10.2991/assehr.k.201204.013

Odigwe, F. N. (2020). Assessment of internal revenue generation techniques of public secondary school managers in Cross River State, Nigeria. *Humanities and Social Sciences Letters, 8*(4), 407–417. https://doi.org/10.18488/journal.73.2020.84.407.417

Odigwe, F. N., & Owan, V. J. (2019). Trend analysis of the Nigerian budgetary allocation to the education sector from 2009 – 2018 with reference to UNESCO'S 26% Benchmark. *International Journal of Educational Benchmark, 14*(1), 1–14. https://doi.org/10.5281/zenodo.4458703

Odigwe, F. N., & Owan, V. J. (2022). Principals' management of financial and non-financial resources as correlates of institutional goal fulfilment in secondary schools in Calabar Metropolis. *Global Journal of Educational Research, 21*(2), 123–134. https://doi.org/10.4314/gjedr.v21i2.5

Odigwe, F. N., Offem, O. O., & Owan, V. J. (2018). Vocational training duration and university graduates' job performance in Cross River State, Nigeria. *International Journal of Current Research, 10*(7), 72024–72028. https://doi.org/10.5281/zenodo.4320545

Okolie, U. C., Nwosu, H. E., Eneje, B. C., & Oluka, B. N. (2019). Reclaiming education: Rising above examination malpractices, and its contextual factors on study progress in Nigeria. *International Journal of Educational Development, 65,* 44-56. https://doi.org/10.1016/j.ijedudev.2019.01.001

Olaitan, A. W. (2017). Impact of family structure on the academic performance of secondary school students in Yewa Local Government Area of Ogun State, Nigeria. *International Journal of Sociology and Anthropology Research, 3*(1), 1-10. https://bit.ly/3GW44gD

Ololube, N. P., Onyekwere, L.A., & Agbor, C. N. (2016). Effectively managing inclusive and equitable quality education to promote lifelong learning opportunities (LLO) for all. *Journal of Global Research in Education and Social Science, 8*(4), 179–195. https://www.ikprress.org/index.php/JOGRESS/article/view/2616

Onyeche, N. M. (2018). Alternative sources of funding and management of public universities in the Niger Delta States of Nigeria. *International Journal of Innovative Finance and Economics Research, 6*(3), 66–73. https://bit.ly/3k8OZ2l

Orak, S., Çilek, A., & Yilmaz, F. G. (2020). Adaptation of traditional children's games to social studies course: STEM course design for teachers. *Cypriot Journal of Educational*

Sciences, 15(6), 1422-1438. https://doi.org/10.18844/cjes.v15i6.4318

Owan, V. J., & Ekpenyong, J. A. (2022). Usage of electronic infrastructures and students' learning effectiveness in Nigerian universities: A polytomous logistic prediction. *Ubiquitous Learning: An International Journal, 15*(2), 87–104. https://doi.org/10.18848/1835-9795/CGP/v15i02/87-104

Owan, V. J., & Owan, M. V. (2022). Availability of digital resources and institutional compliance with COVID-19 mitigation measures in a Nigerian university: A descriptive study. *Electronic Journal of Medical and Educational Technologies, 15*(4), Article em2208. https://doi.org/10.29333/ejmets/12411

Owan, V. J., Ekpenyong, J. A., & Asuquo, M. E. (2021). A structural equation model of principals' communication patterns, funds management and school-community relationship. *Journal of Pedagogical Sociology and Psychology, 3*(1), 1–18. https://doi.org/10.33902/JPSP.2020364435

Owan, V. J., Ekpenyong, J. A., Chuktu, O., Asuquo, M. E., Ogar, J. O., Owan, M. V., & Okon, S. (2022). Innate ability, health, motivation, and social capital as predictors of students' cognitive, affective and psychomotor learning outcomes in secondary schools. *Frontiers in Psychology, 13*, Article ID 1024017. https://doi.org/10.3389/fpsyg.2022.1024017

Owan, V. J., Owan, M. V., & Lata, N. (2022). Discharge of pedagogic duties: A bootstrapped structural equation modelling of teachers' use of research materials in school libraries. *Journal of Applied Learning & Teaching, 5*(2), 116–131. https://doi.org/10.37074/jalt.2022.5.2.13

Owan, V. J., Owan, M. V., & Ogabor, J. O. (2023). Sitting arrangement and malpractice behaviours among higher education test-takers: On educational assessment in Nigeria. *Journal of Applied Learning & Teaching, 6*(1), 1–15. https://doi.org/10.37074/jalt.2023.6.1.5

Paletta, A., Basyte Ferrari, E., & Alimehmeti, G. (2020). How principals use a new accountability system to promote change in teacher practices: Evidence from Italy. *Educational Administration Quarterly, 56*(1), 123-173. https://doi.org/10.1177/0013161X19840398

Pastore, F. (2019). Why so slow? The school-to-work transition in Italy. *Studies in Higher Education, 44*(8), 1358–1371. https://doi.org/10.1080/03075079.2018.1437722

Peterson, D. C., & Mlynarczyk, G. S. (2016). Analysis of traditional versus three-dimensional augmented curriculum on anatomical learning outcome measures. *Anatomical Sciences Education, 9*(6), 529-536. https://doi.org/10.1002/ase.1612

Polit, D F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459–67. https://doi.org/10.1002/nur.20199

Prabhakar, G. V., Gantasala, S. B., Madireddy, M., & Upadhyay, P. (2022). School resource inadequacy and school social engagement: Mediating effect of principals' stress. *Journal of Community Psychology, 51*(1), 284-296 https://doi.org/10.1002/jcop.22903

Robert, I. A., & Owan, V. J. (2019). Students' perception of teachers' effectiveness and learning outcomes in Mathematics and Economics in secondary schools of Cross River State, Nigeria. *International Journal of Contemporary Social Science Education, 2*(1), 157–165. https://bit.ly/2VbLzfK

Sanga, I. (2022). Philosophical assessment of youth moral quality in Tanzania in sixty years of independence. *American Journal of Humanities and Social Sciences Research, 6*(2), 192-197. https://bit.ly/3jW6G4Q

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review, 51*(1), 1-17. https://www.jstor.org/stable/1818907

Shing, L. S., & Rameli, M. R. M. (2020). The influence of self-regulation towards academic achievement in English among Malaysian upper primary students. *Universal Journal of Educational Research, 8*(5), 1-11. https://doi.org/10.13189/ujer.2020.081901

Sidi, A. U. (2019). The impact of school facilities on students' level of motivation and academic achievement in senior secondary schools in Northwest Zone, Nigeria. *Prestige Journal of Education, 2*(2), 254-263. https://bit.ly/3vTJDKF

Sinclair, K., & Malen, B. (2021). Student-based budgeting as a mechanism for promoting democratic decision making: Testing the theory of action. *Educational Policy, 35*(5), 781–809. https://doi.org/10.1177/0895904819843601

Strickland, G. L. (2021). *The association between school finance and graduation outcomes for students with disabilities in the state of michigan* (Doctoral Dissertation), Wayne State University. https://digitalcommons.wayne.edu/oa_dissertations/3432/

Suleiman, Y. (2018). Addressing the factors responsible for schooling without learning in primary and secondary schools in Nigeria. *International Journal of Synergy and Research, 7*, 161-178. https://doi.org/10.17951/ijsr.2018.7.0.161-178

Tamminen, K. A., & Poucher, Z. A. (2020). Research philosophies. In D. Hackfort, & R. Schinke (Eds), *The Routledge international encyclopedia of sport and exercise psychology* (pp. 535-549). Routledge. https://doi.org/10.4324/9781315187259-39

Takwate, K. T. (2018). Allocation, availability and maintenance of school facilities as correlates of academic performance of senior secondary school in Anambra State, Nigeria. *International Journal of Scientific and Research Publications, 8*(9), 42-81. https://doi.org/10.29322/IJSRP.8.9.2018.p8142

Ugwuanyi, C. S., Okeke, C. I., & Asomugha, C. G. (2020). Prediction of learners' mathematics performance by their emotional intelligence, self-esteem and self-efficacy. *Cypriot Journal of Educational Sciences, 15*(3), 492-501. https://doi.

org/10.18844/cjes.v15i3.4916

Venieris, G., Naoum, V. C., & Vlismas, O. (2015). Organisation capital and sticky behaviour of selling, general and administrative expenses. *Management Accounting Research, 26*, 54–82. https://doi.org/10.1016/j.mar.2014.10.003

Wang, F., Hauseman, C., & Pollock, K. (2022). "I am here for the students": Principals' perception of accountability amid work intensification. *Educational Assessment, Evaluation and Accountability, 34*(1), 33-56. https://doi.org/10.1007/s11092-021-09368-6

Watson, S. L., Loizzo, J., Watson, W. R., Mueller, C., Lim, J., & Ertmer, P. A. (2016). Instructional design, facilitation, and perceived learning outcomes: an exploratory case study of a human trafficking MOOC for attitudinal change. *Educational Technology Research and Development, 64*(6), 1273–1300. https://doi.org/10.1007/s11423-016-9457-2

Webber, D. A. (2012). Expenditures and postsecondary graduation: An investigation using individual-level data from the state of Ohio. *Economics of Education Review, 31*(5), 615-618. https://doi.org/10.1016/j.econedurev.2012.02.003

Wong, E. Y. C., Ho, D. C., So, S., & Poo, M. C. P. (2022). Sustainable consumption and production: Modelling product carbon footprint of beverage merchandise using a supply chain input-process-output approach. *Corporate Social Responsibility and Environmental Management, 29*(1), 175-188. https://doi.org/10.1002/csr.2193

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal, 11*(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6

Zangirolami-Raimundo, J., Echeimberg, J. O., & Leone, C. (2018). Research methodology topics: Cross-sectional studies. *Journal of Human Growth and Development, 28*(3), 356-360. http://dx.doi.org/10.7322/jhgd.152198

Zhao, Y., & Hong, Y. (2012). Social capital and education attainment: A perspective of social network resources and social closure. *Sociological Study, 42*(5), 47–68. http://dx.doi.org/10.6007/IJARBSS/v9-i13/6247

Zyphur, M. J., & Pierides, D. C. (2020). Making quantitative research work: From positivist dogma to actual social scientific inquiry. *Journal of Business Ethics, 167*, 49-62. https://doi.org/10.1007/s10551-019-04189-6

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

A critical perspective on generative AI and learning futures. An interview with Stefan Popenici

Stefan Popenici[A]
Jürgen Rudolph[B]
Shannon Tan[C]
Samson Tan[D]

[A] Academic Lead - Quality Initiatives, Charles Darwin University, Australia
[B] Director of Research, Kaplan Singapore
[C] Research Executive, Kaplan Singapore
[D] Director of Regional Strategy & Operations (Singapore), Civica Asia Pacific

Keywords

Algorithmic bias;
Artificial Intelligence (AI);
AIEd;
Big Tech;
ChatGPT;
educational technology;
fascism;
generative AI;
higher education;
superintelligence.

Correspondence

Jurgen.Rudolph@kaplan.com [B]

Article Info

Received 6 June 2023
Received in revised form 10 July 2023
Accepted 10 July 2023
Available online 11 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.5

Abstract

We present a wide-ranging interview with Stefan Popenici, a distinguished scholar and public speaker with extensive experience in higher education. Popenici's research focuses on the impact of artificial intelligence (AI) on teaching, learning, quality assurance, and student engagement in higher education. The interview delves into the themes of his book, *Artificial Intelligence and learning futures: Critical narratives of technology and imagination in higher education* (2023), exploring the intersection of AI, intelligence, and societal issues such as eugenics and racism. Popenici critiques the power of tech titans and the belief in technology as a panacea, especially in higher education. The discussion also addresses the identity crisis in higher education, the potential of revisiting Humboldt's 19th-century vision of the university, and the challenges and opportunities presented by the AI revolution. Popenici's insights into the role of AI in assessment, graduate and academic employment, and the future of academic work are particularly illuminating. The interview concludes with Popenici's reflections on his own educational journey and future plans.

Jürgen Rudolph (JR): Thank you so much for making yourself available for this interview for the Journal of Applied Learning and Teaching (JALT). You are a scholar and public speaker with over 25 years of experience in teaching, research and leadership in higher education, with universities in Europe, North America, Southeast Asia, New Zealand and Australia. For your work and strategic leadership in education, the President of Romania knighted you with the Order 'Merit of Education'.

Your research is currently focused on the impact of artificial intelligence in teaching and learning in higher education, and quality assurance and student engagement. We are big fans of your book *Artificial intelligence and learning futures. Critical narratives of technology and imagination in higher education* (Popenici, 2023a). What made you write the book? What are its main theses?

Stefan Popenici (SP): What made me write the book is a complex answer because it comes as a profound sense of panic and dissatisfaction with what I see that is happening now in education. It is ironic that I love technology. I use a lot of technology. My wife complains that I use too much technology and then I have too many gadgets and boxes that she cannot manage. But the obsession with using technology as a silver bullet in education, ignoring some of the most important parts and the lack of reflection over 'What are we actually using? What are we actually doing?' came as a strong motivator to address this in a book rather than a short article.

Mainly, it is this sense of profound crisis for education and for our civil society. This is a very important part of my identity. I believe in a civil society. I'm passionate about intellectual and personal freedom, the sense of a civil society and then the power of education to change lives for the better. I know this sounds like big words, but this is how I grew up. This is how my life developed, and it came with a very profound sense of responsibility. I was lucky. I'm privileged, and I think I have to give back. The book was my way to give back and contribute to the general discussion about what are we actually doing for our present and future. This is a very strong European sense of when you're an intellectual, you have a responsibility for society. This is the part of Europe that I love.

Figure 1. Stefan Popenici with a copy of his *Artificial Intelligence and learning futures*.

JR: Could you tell us a bit more about the main theses of your book?

SP: The main point is looking at the impact of what I find as most consequential, the most influential technology that is going to change education – and that is artificial intelligence. One of the main problems is that we don't stop to think about what we are actually doing. It's a very strange thing for education where you deal with researchers and intellectuals, and there is basically no interest in looking at what are we actually going to use. When you buy a car, you want to know what the car is going to do. When you use a technology as complex and influential as artificial intelligence, you would expect a very serious conversation about all aspects that are shaping this technology. The main point is starting from a fact. Artificial intelligence is a marketing concept. I'm jokingly saying that we are going to use cups of coffee with artificial intelligence, as it is used on everything because it sells. It's not a real thing.

Artificial intelligence is a marketing concept. I'm jokingly saying that we are going to use cups of coffee with artificial intelligence, as it is used on everything because it sells. It's not a real thing.

There are some research groups that went so far as to suggest banning the concept because it's so slippery and open to manipulation. As we speak, we see it in the public discourse. This is a great fight on emotions and then using this concept for marketing purposes, not for anything else but to make even more profits. We're talking about billions of dollars, there's a lot at stake. When you stop to think about this concept, you realize that there are some sources that are very problematic. Educators should stop and think if these roots of the concept of artificial intelligence are not somehow problematic for education. And if they are, what can we do?

So this is what the book is looking at. This is also something that I hope we will touch on later in our conversation – the full impact. Artificial intelligence in education is not new. I'm using an example on purpose in my book about a conference in Europe in 1990. The conference was called something like 'Artificial intelligence in higher education'. It's not a new idea at all, but the full impact is seen, in my opinion, starting with 2023. This is when we realized that education is in a profound crisis and especially higher education is under attack from various ideologies. Universities are under attack by neoliberalism and the obsession to make education a business and reduce all to profits and markets. You have this unfortunate context, and the new technology is coming to *disrupt*. I'm careful with words: Disruption is to destroy. So when you have a disruptive force that is going to change entirely the landscape in a crisis, results are going to be very problematic. This is what I think is the second main thesis of the book – that we're going to see massive changes, and we have to start paying attention to these challenges.

Shannon Tan (ST): Could you illuminate the unsavoury connections between the concepts of intelligence and artificial intelligence with eugenics and racism that you discuss in the first section of your book?

SP: This is an important part of the book because it is a part that is universally ignored. It's impossible to turn on the news and miss artificial intelligence. It's going to be mentioned somehow. It's going to destroy the world, do something extraordinarily important, move us all to Mars, whatever! The problem is that the concept of intelligence, which is at the core of artificial intelligence, is tainted by a certain view of the world. The way we understand intelligence today is, unfortunately, shaped by the group of thinkers and researchers that looked at intelligence as a dimension that can and should be measured. This is such an important part of that conversation about artificial intelligence that I felt that it must be very well-documented.

When we speak about intelligence, we speak in general and almost universally from the perspective opened by Francis Galton. It's not the only perspective, by the way. If you look at indigenous cultures – and I'm on Larrakia land, 60,000 years of continuous history – they look at intelligence in a very different way from academia. But Galton looked at intelligence as something that should be measured, and that is ranking human beings in a certain order that was based on a concept that he invented: eugenics. Eugenics did not originate in Nazi Germany.

It's originating in this unfortunate development in human history where Francis Galton came up with the racist idea that intelligence is linked to races that are superior in terms of what he called intelligence. What he called intelligence was only what he could measure, and the next step was that because we have a ranking of intelligences based on races, we have to practice racial hygiene.

Figure 2: Francis Galton (right), aged 87, at Fox Holm, Cobham, with his biographer, the statistician Karl Pearson. Source: Wikimedia Commons (n.d.), public domain.

This idea transpires to this point in Silicon Valley. I purposely documented very well how one of the founding fathers of Silicon Valley, William Shockley's ideas of white supremacism and eugenics, are absolutely astonishing. It's important to keep in mind that he ran for office in the United States with these ideas. These guys were not hiding these ideologies. They're quite proud and organized international conferences at University College London and Stanford University. We talk about the most prominent institutions where these ideas shaped the way intelligence is seen by artificial intelligence.

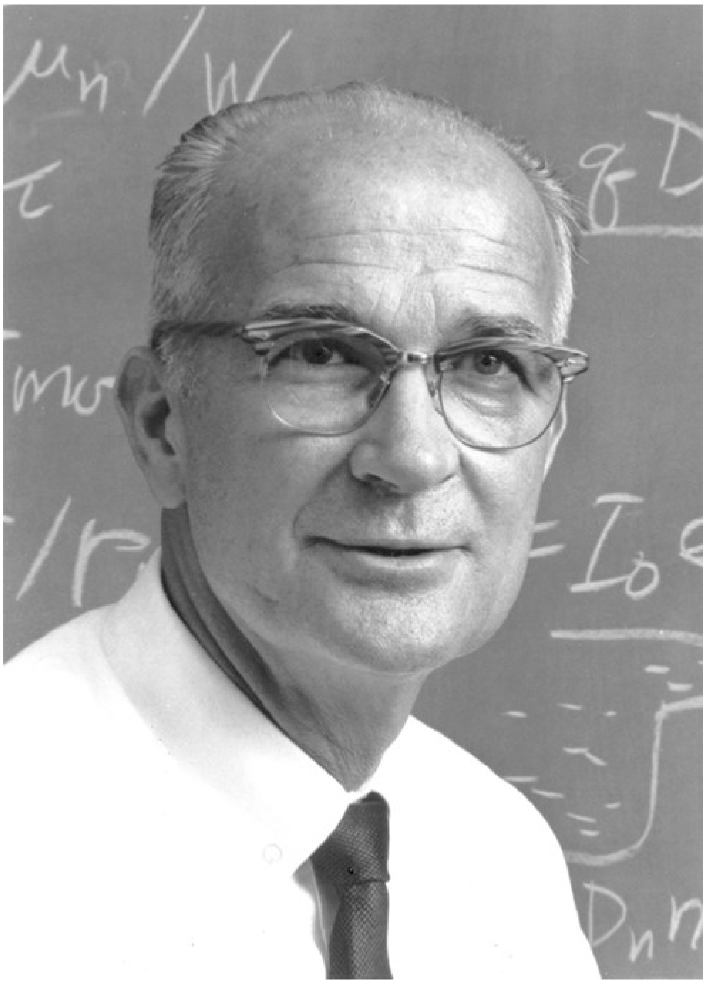


Figure 3. William Shockley. Source: Painter (1975), public domain.

Shockley not only considered that Whites are intellectually superior, but he proposed to create a welfare system with financial incentives to get rid of what he called genetically disadvantaged groups – of course, Blacks and other minorities. This is what Meredith Broussard (2023) documented so well in her latest book, *More than a glitch: Eugenics, racism and discrimination* are at the core of artificial intelligence. It's not a glitch, it's by design, and this is very important.

I can give you just one example: In an article on technology and the positions of women, John McCarthy (2006) wrote that it's a mistake to think that women are equal to men; they are inferior. This is the guy who created the concept of artificial intelligence. This is a part that we must not ignore if we are serious about that conversation. Banning or ignoring AI is a mistake. Equally bad is to ignore that this narrow view of intelligence is very problematic.

JR: What I found really shocking when reading your book was that the Nazis were looking at some of the things that were going on in the US, and they thought that was too extreme.

SP: This is a crucially important topic, and I can explain in a nutshell why. One of the main problems with the world we have today, in my opinion, is the rise of fascism and white supremacy all over the world. This is a mortal danger for our societies, and I think it's easy to argue that we had a wrong approach to looking at ourselves as the human race. I look at history because I think it's important to understand where we are today and what we are going to have tomorrow.

One of the main problems with the world we have today, in my opinion, is the rise of fascism and white supremacy all over the world. This is a mortal danger for our societies.

We had the First World War, and then the world completely missed the lessons. They couldn't understand anything that happened. Then, we had the Second World War, and unfortunately, what happened after that, in my opinion, is that we fell on easy explanations that are very problematic and false. One of the explanations was that this horrendous example of dehumanization – and how wrong ideas create monsters – was reduced to one nation and to one narrow geographical region. This is failing to understand why we had genocides happening in Myanmar, Rwanda and other parts of the world.

More importantly, it fails to look at the truth. The truth is that the first forced sterilisation in the world did not happen in Germany. It happened in the United States. The first time Zyklon B – the gas used in Auschwitz for what was called horrendously 'the final solution' to wipe out an entire race – was used was not in Germany. The first time it was used was in the United States to clean – and this is tragically symbolic – Mexicans and other foreigners. Unfortunately, at that point, there was one German scientist, and then he came up with this idea: 'Oh, using gas to clean aliens, that's an interesting idea!' Then, when the Nazis became so monstrous, they just increased the dosage.

Where in the history of the United States can you see that? Nowhere in the public discourse. Before the Nazis became so extreme, they were always monstrous. But there were stages before they reached the final and most disturbing stages. This is documented very seriously in a book that is significantly called *Hitler's American model* (Whitman, 2017). The Nazi party sent a delegation to the United States to see how Americans solved the problem of Blacks in the United States and to learn from them how to deal with the Jewish problem in Germany. Now the horrendous lesson is that the Nazi delegation came back from the United States with the message that 'we are civilised people. We cannot do what they do in the United States'. That's documented in an archive, it's not a sad metaphor for what happened. This is how extreme the Jim Crow laws were. It's enough to read James Baldwin about the experience of Blacks in the United States to understand why people found that unacceptable and outrageous (e.g. Baldwin, 2001).

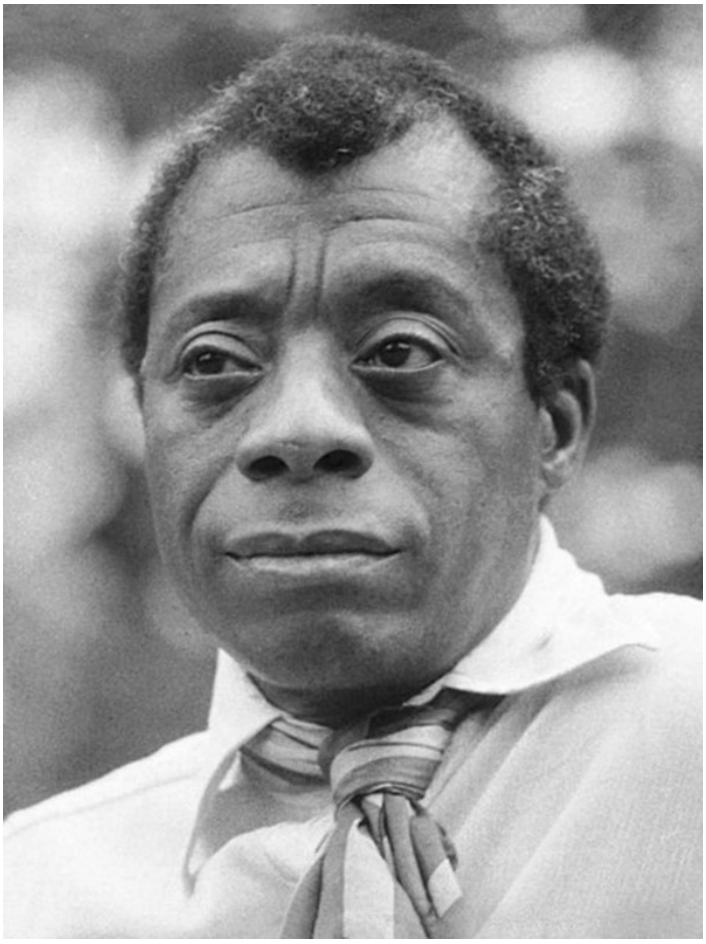


Figure 4: James Baldwin. Photograph: Warren (1969), public domain.

When we look at these problems, we should not fall into this reflex of pointing the finger at others, and say 'Oh, it's just that group, it's just that nation, it's just that party or just that set of ideas that are marginal. We shouldn't care'. I think it's important to look at these challenges and risks that somehow seem to be part of human nature and deal with them courageously. Most importantly, universities are the space that is most suitable and responsible for dealing with that. Universities don't care about this; that's the reality. There is no conversation. You read *Times Higher Education* or *The Chronicle of Higher Education*. You read research papers, and you don't see this part of the conversation. We dismiss it as philosophical, ideological, 'it is not real life'. Well, before, it was real life. It was very ideological and philosophical, as I explained with eugenics and that led later on to the very real concentration and extermination camps.

That's very real. You can't get more real than that when you kill people en masse. By the way, technology should answer to what happened. Because going back to the Nazis, there's a very important lesson. I'm not religious. I'm not Jewish. I'm not even sure if I can identify myself with a certain nationality because I travelled too much. It's not about something that is personal because it's linked to my culture. It's personal because it's linked to my status as a human being.

There is an important lesson in what happened in the Second World War when the Nazis decided to apply the final solution. One main problem they have – which I mentioned too succinctly, unfortunately, in the book – was that the scale of killings was so massive that they could not organize it. Who came to help? IBM. The book titled *IBM and the Holocaust* (Black, 2001) is a massive doorstopper because it is extraordinarily well documented. This is what technology also did: help the Nazis get rid of human beings. It should be a serious soul-searching of the role of technology and then what technology without values and without serious thinking can do. It's not doing good things, unfortunately.

JR: I agree it's very important not to forget history. I'm not pointing any fingers at anybody else but at myself because that is my history. You were saying that you find it difficult to identify yourself with any nationality. I feel the same. But there's this huge guilt because of German history, and I don't think that's a bad thing. I was not personally involved because I'm too young. Even my father was too young to be involved, but my grandfathers were involved, and I think it's important to never forget that.

SP: When I was working in Romania, I started many projects in education. I worked my entire life in education. One of the projects I started was a national project called Education Against Racism, Discrimination and Anti-Semitism. It started with racism against gipsies. Then I reached the point of anti-Semitism. I'm a nerd: I go to archives, and I look at facts. At that time, Romania had the universally accepted narrative that Romania was a safe haven for the Jews fleeing Nazism. The truth is that the Holocaust happened in Romania in the most horrendous circumstances one can imagine.

It is important to note that the Holocaust was a European project; it was not a German project. In general, it was accepted by the whole world. Before other nations say, 'oh, that was a European problem', I don't think any continent is in the position to point any fingers if they look at their own recent history. I'm not even talking about long history. That's why I refuse to link this with nationality because it's not accurate. It's simply wrong. It's about the responsibility of human beings, and for me, what is important is the responsibility of educators. It all started with bad education.

Samson Tan Yong Tiong (STYT): I want to just keep listening to you talking about that very strong link between racism and human biases being passed on to technology itself. This is something I'm very concerned about as well. In 2015/2016, when there was already a lot of comparison between the development of AI in education in China and the US, they were installing facial recognition cameras in the classroom, and they were trying to find out about student responses to the lessons. We realized that those AI-powered facial recognition algorithms that were developed in China seemed to be able to pick up the students' responses much better across the board. Even when they applied them in the US, they were able to recognize a person who is White, Black or of any other colour, compared to the software that was developed in the US. Why was that the case? Possibly because of the biases of the algorithms that were developed in the US itself. They somehow built in those biases that weaken the software's ability to recognize people who are not white.

When I read your book, you talked about your concerns about the power of the tech giants. They are mostly dominated by US companies. They construct algorithms within a black box – we don't know what's going on inside the algorithms. Earlier, you also referred to what Meredith Broussard (2023) terms techno-chauvinism, IBM's role in the Holocaust, and what's happening in higher education. Could you elaborate a bit more on this? I really want to find out about your thoughts here.

SP: I wrote the book as someone interested in technology from an educational perspective. I don't claim expertise in engineering, though I read as much as I can in terms of research papers and books presented by engineers. In general, I trust what they are saying when they have real expertise. I apply my academic scepticism to look at what they present. The black box principle is very simple. We know what gets in. We know the kind of data and information we produce, and we see the results. What is happening inside, we have no idea, and to paraphrase Meredith Broussard (2023) again, this is not a glitch.

A couple of weeks ago, I was in a meeting with a top executive of one of the Big Five in Australia. I expressed my astonishment that Australia couldn't manage to bring a case against Meta [formerly known as Facebook] to court because Facebook simply said: 'You have no jurisdiction, we are Americans, we don't care about your stuff.' You can see the sense of impunity they have after all the scandals and disasters, including Cambridge Analytica. They still don't have any social responsibility or serious legal responsibility.

We are impressed by them having to pay millions of dollars. But it's not even change for these companies. They consciously exploit innumeracy. People can't make sense of big numbers because mathematics is not as strong in our education. (That's another discussion.) But this black box is cultivated. This executive in the meeting, when he took the floor, he said: 'So you want artificial intelligence in Australia?' when the discussion was specifically about the privacy of data and the importance of how this data is managed. 'Why do you want it? You want to speak in Australian lingo?' The dismissive and ridiculous arguments show something else behind all these types of reactions that should not exist in a serious conversation. And that was a very serious conversation with decision-makers in Australia. This principle of the black box is always defended. Google is not making clear why you show up and why you disappear from their rankings.

OpenAI made clear that they are going to be totally opaque. If you read what they've said at the beginning, it was all about the fiduciary duty to humanity (Rudolph et al., 2023a). 'We don't care about money. We don't care about anything but transparency, humanity and serving the world well'. They got $10 billion in January. So much for 'we don't care about money'.

Then, this black box principle is very important. It's at the core of the kind of education we want to create. They discriminate based on race, and social status and then you realize that tomorrow I can be one of those discriminated against. Tomorrow, you may be guilty of living in a less affluent area. When you do that, you may be immediately the victim of one of the algorithms that is acting on your life and shaping your life. The rates you pay at the bank, the kind of credit and the kind of healthcare you get: the algorithm is deciding this. It's not only that. It's the black box principle that you don't know how the algorithm is working, and you have no idea how the decisions are made. It's even worse than that: you have no possibility of recourse when it's saying you don't deserve healthcare.

By the way, it happened, and I can give you some examples. One is an example from Europe. It is an algorithm that was used in Spain to decide the kind of help women suffering sexual abuse and then living with serious threats to their lives would receive. They used an algorithm to rank these threats. In 2021, because there was no help assigned by the algorithm, 71 women were killed. We talk about people losing their lives because it was decided by an algorithm that the police should not attend to these cases. There are numerous cases of people in the United States who were arrested and put in prison because an algorithm decided that they were guilty just because they lived in a poor neighbourhood.

There are cases of women in an American Hospital who got no medical care because an algorithm decided that they didn't need that much care. Later on, it was discovered that the algorithm was discriminating in favour of affluent White women. When you do this in education, there is the risk of discrimination against those who most need our help and attention and that we can benefit from. Just look at human history and see how many of the great inventors and artists and then people who really pushed the world ahead came from disadvantaged backgrounds. Beethoven might be imprisoned if decided by an algorithm, and this is just one example. I can give you thousands.

The second part that you destroy is education. This should be evident for anyone going through significant, meaningful education. When you are constantly under surveillance, you kill education, you kill the sense of connection, you kill the sense of trust. How can I trust you if you keep me under surveillance all the time? The saddest part of the space of higher education now is that it is guided by what we call evidence-based decisions. If you're familiar with the field, you realize that it is evidence-based as long as the evidence serves a certain ideological position. When the evidence shows that this is wrong, oh, we forget about the evidence. Just look at the research on open spaces. It shows that it is killing productivity. Just what do you see in universities? Where is the evidence that this is working? It's the same with technology. I'm using common sense, easy-to-see examples, but when you go into details, you realize that. Research shows that surveillance is changing human beings' behaviour. When you do it to kids in schools, the kind of impact can be devastating, and you just killed education that is meaningful. You pass on information, and you train the same way you train dogs to bring a ball, but that's not education.

Research shows that surveillance is changing human beings' behaviour. When you do it to kids in schools, the kind of impact can be devastating, and you just killed education that is meaningful. You pass on information, and you train the same way you train dogs to bring a ball, but that's not education.

When you look at artificial intelligence, techno-chauvinism or solutionism (a term coined by Morozov (2013)), is a perfect example. In order for artificial intelligence to be perceived as universal and all-encompassing, the trick was to narrow down what we understand by intelligence. The second trick is to look at life as a set of problems that can be solved. Well, life is more than that. You can solve all the problems, and you have your heart crushed in love, and it's all going to fall apart. That's not a problem to solve. That's about emotions. That's about love. That's about humanity. It still matters. Emotions still matter. It's just not a problem to be solved. You cannot reduce this so badly. When you apply this colonialism of problem-solving, and you say only technology can solve that, you ignore how the world is going. That's a criminal mistake.

You ignore simple lessons that are connected to our previous points. Let's look at societies where technology was working perfectly. When you look at Nazi Germany, whether we like it or not, it was the most advanced nation on Earth in terms of technology. We can think about exploring the moon and Mars and all that because Hitler started the project on rockets. By the way, Americans took all the Nazis who used slaves and just moved them to the United States, where they continued their research. The point is that from a technological perspective, Nazi Germany was not doing badly at all. They were the most technologically advanced. They had the best weapons and the best technologically-trained people. Technology is not solving all the problems. When you reduce all problems to technology, and then technology can solve all, you create monsters. We have a long history to prove that. When we have lessons that are too painful to contemplate, then it's important to have these discussions now rather than when disasters happen.

Technology is not solving all the problems. When you reduce all problems to technology, and then technology can solve all, you create monsters. We have a long history to prove that.

STYT: Your exposition reminded me of something that I was trying to explain to a group of STEM teachers. They were talking about how to explain this bias that you put into the algorithms that are problematic to all sorts of things, especially in an educational context. I was fascinated by your real-life examples from Spain and elsewhere. I used more of a science fiction approach because the STEM teachers were supposed to be targeting younger students. I referred to Marvel, where there is Captain America dealing with Hydra. Hydra developed algorithms to target people who are against them. If you develop algorithms to target anyone that opposes you, you can take them out using that kind of sophisticated weapon. Obviously, that's in a science fiction environment.

But the point that I mentioned to the teachers was that these are the kind of things that are happening in society quietly behind closed doors in the black box. The danger of it is that we get targeted without even knowing that we're being targeted. It's not only happening in Marvel science fiction movies, but it's already intruding into our lives. I'm a convert when it comes to cautioning about the blind faith that technology can solve all problems. This is something that we need to be very aware of, not only in education but by and large. Now a lot of people are thinking that just because of generative AI, they can solve a lot of problems that we face today.

SP: Across humanity, we've been tempted to look at technology as something that is going to give us some certainty. So we reduce life to something that we can finally control, and then science and technology are going to give us that. It's nothing new in that we've always believed that the latest technology is going to give us the solution to control the world. It's a very dangerous thing, and the example I was going to use is a very real example and it can be found in the book (Popenici, 2023a). Stanislav Petrov didn't believe that all solutions coming from technologies were good. The problem was that the latest technology used by Soviet Russia showed that Soviet Russia was under attack by the Americans, and all nuclear rockets were armed. There was this guy who said, 'no, this doesn't make sense'. Petrov risked his life – I lived in a communist dictatorship; believe me, that's not a metaphor! He risked his life, and he said, 'no, I'm not going to start this'. We wouldn't be here if humanity was at that moment based on this logic that technology knows best because it's going to give us solutions. Technology without human control can spell the end of that. It's not going to be 'I don't believe in that'. If we are going to end our race and then destroy the Earth, it is going to be us. It's not going to be anything other than us, so no technology is going to do this better than ourselves.

JR: Meta is moving away from the black box idea that OpenAI and Alphabet are following because they are making their algorithms publicly available (Weatherbed, 2023). I think this is not out of the goodness of their hearts, but they are trying to catch up. Meta obviously has a lot of examples of using AI, which were quite dismal and disastrous. But I nonetheless thought that it was interesting that they were suddenly being more open than OpenAI on which Elon Musk commented that they are not open anymore (Rudolph et al., 2023a).

One very quick follow-up question: I love Beethoven, and you mentioned that he could have been in jail if he had lived in our time. Why?

SP: During much of Beethoven's life, he lived in relative poverty. Imagine a world without Beethoven.

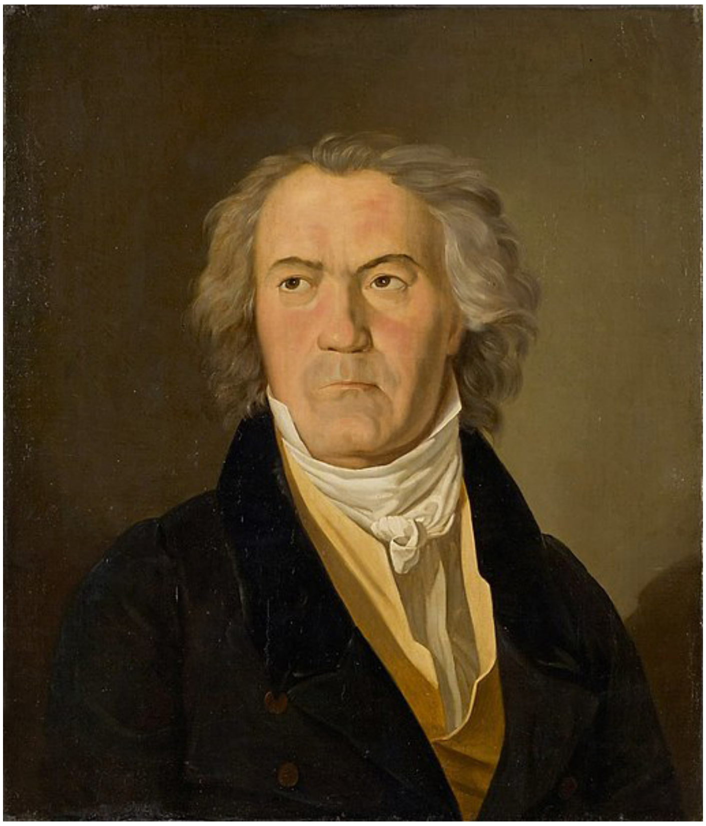


Figure 5: Ferdinand Georg Waldmüller's 1823 painting of Ludwig van Beethoven. Beethoven's (1770 – 1827) financial situation was often precarious. Complicating matters further, Beethoven struggled with health issues, including deteriorating hearing, which affected his ability to perform and earn income from concerts. He frequently accumulated debts. Facing legal actions from creditors did not prevent him from creating some of his greatest masterpieces (Swafford, 2014).

Imagine a world without so many thinkers. We all like Michelangelo (1475 - 1564) and Leonardo Da Vinci (1452 - 1519). With the kind of surveillance we have today, we wouldn't have any Leonardo's because he was breaking all the rules of his time. Leonardo conducted dissections when there were severe penalties for that. Then he came up with ideas that were outrageous.

Michelangelo also conducted dissections (Eknoyan, 2000). Then again, you can imagine Michelangelo being very young in prison rather than giving us La Pietà and then giving us the Sistine Chapel. That's the kind of thing we have to consider, especially when we talk about education. It's not a marginal thing.

Going back to your comment, Jürgen, you can blame me for my scepticism. I would believe it when I saw it. What I see so far about the Big Five tech companies [Alphabet, Amazon, Apple, Meta, and Microsoft] is a lot of PR and noise. I can give you the latest fact. Bard was released by Google as the new AI solution (see Rudolph et al., 2023b) – for transparency and openness and the love of God, whatever. When you look at where it was not released, that's 180 nations. They did not release it in the European Union. Why? Because the European Union is asking them to be more transparent and more responsible. So surprise, surprise, when you look at facts, when you draw the line, that's what you can see. We have a long history of smoke and mirrors used by unchecked power that never ends well. This is what we have with these big tech companies. They don't answer to anyone. They are not elected, not checked. That's a problem.

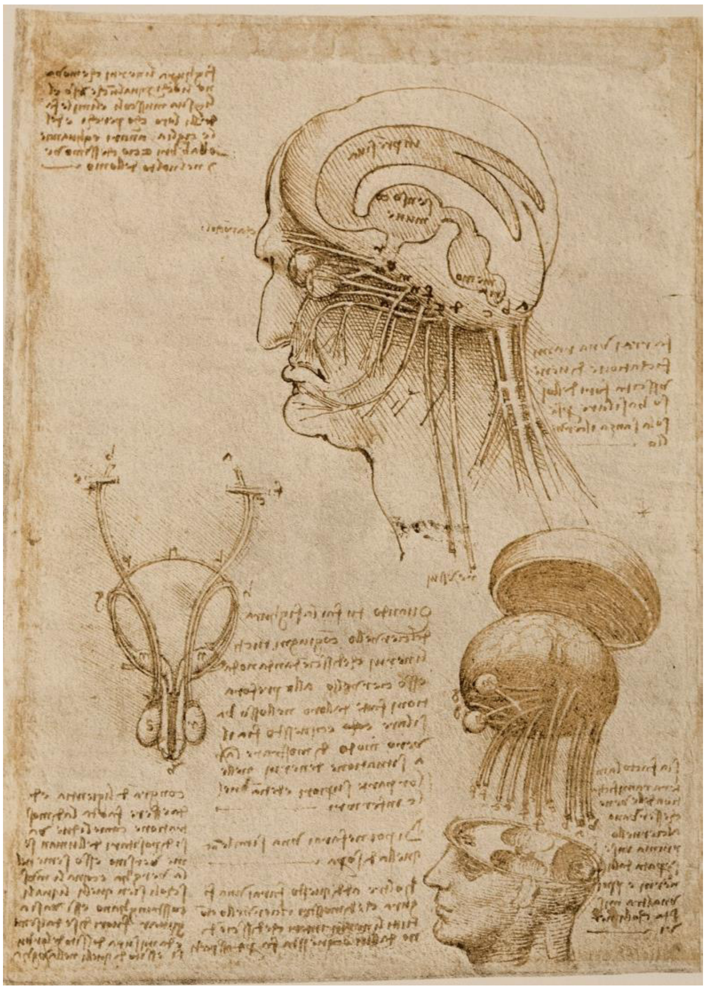


Figure 6. Leonardo da Vinci's Sketch of the Human Brain and Skull (1510). Leonardo had a deep interest in understanding the human body and was known for his anatomical studies. While the Catholic Church considered the practice of dissection as sacrilegious and immoral, Leonardo believed that it was essential for the advancement of medical knowledge (Jose, 2001). Penalties for performing dissections on human cadavers during the Renaissance ranged from fines and imprisonment to more severe punishments, including public humiliation and ex-communication from the church (Isaacson, 2017).

We have a long history of smoke and mirrors used by unchecked power that never ends well. This is what we have with these big tech companies. They don't answer to anyone. They are not elected, not checked. That's a problem.

JR: In the second section of your book, you diagnose that higher education is undergoing an identity crisis. In your view, rampant anti-intellectualism, the Americanisation of higher education, the audit culture and the metrification of academic life run counter to educational and human

Figure 7. Michelangelo's Madonna della Pietà (1498–1499). La Pietà, a dolorous image of Jesus and Mary at Mount Golgotha, is a key work of Italian Renaissance sculpture. Source: Traykov (2008).

values such as the love for learning, beauty, and passion. In *Dark academia*, Peter Fleming (2021) provides a historical overview of four shifts that the university has undergone. The first paradigm shift in higher education was epitomized by Wilhelm von Humboldt's 19th-century vision. He conceptualized the university as a place that harmoniously blends research and teaching within the bounds of academic freedom, aiming to nurture students into independent thinkers and global citizens. The next transformative phase began in the 1960s, known as the academic revolution, which democratized access to university education on an unprecedented scale.

However, this led to a backlash and the emergence of the neoliberal university in the mid-1980s, characterized by increased administrative control and the quantification of academic performance. The fourth evolution, triggered by the pandemic, has further propelled the transformation of universities into corporate-like entities, straying even further from Humboldt's model. Despite the shortcomings of Humboldt's idea, marked by elitism and the prevalence of white male privilege, do you think it would be feasible to revisit and adapt his concept while simultaneously eliminating its intrinsic class, race, and gender prejudices? Or is the end nigh? Fleming (2021, p. 19) cautioned, "Beleaguered by managerial bloat, business bullshit and a Covid-compromised economic environment, the idea of the modern university may soon come to an end". Would you concur with this historical overview (see Fleming et al., 2021), and is there any ideal of higher education that we can meaningfully refer to in light of the apparent AI revolution?

SP: Your question points to some very important aspects and steps in the evolution of higher education. Fleming's book (2021) is excellent, and I really enjoyed reading that. However, the most important step is missing, in my opinion, because it all started at the end of the Second World War. The full commercialisation of education started when all these consultants, experts, generals and military guys on the winning side, especially in America, realized that they were out of a job because there was no war anymore. So they found very convenient jobs in international organizations, and they came with a certain view of the world. Amongst these international organizations are OECD and the World Bank; they shape the world and education according to their views. In the 1950s, and then especially after the '60s, people forgot the horrors and the fact that Nazis were pretty good capitalists. Technology and capitalism were not their problems; they were good at them.

There is a great danger in turning capitalism and then technocracy into a religion. It's a very dangerous path. People resisted that, but after the '60s and especially in the '70s, you see the twist of neoliberal ideas gaining ground, also in education. After that, you have the very unfortunate '80s, where you have Reagan in the United States and Thatcher in the UK. They come with this disastrous ideology that never worked.

This is another thing that is magical for me because neoliberals are supposedly good with money. The deficit started in the United States with Reagan's ideas. What is so wrong in looking at the evidence? The evidence is that neoliberalism destroyed the fabric of society, the nature of education, and healthcare. It started to erode the foundations of civil society. Even economically, it wasn't working. It made the rich richer and the poor poorer. The budget went into deficit. This is what happened both in the UK and in the United States. In terms of the misery created in the UK by Thatcher, you just have to read what people at that time were writing, including American diplomats. Even Henry Kissinger, one of the most strident supporters of Thatcher, noted that Britain at that time was a country in disaster; in private conversation with the US President, he observed that "Britain is a tragedy... it has sunk to begging, borrowing, stealing" (Kissinger, 1975).

Way before COVID-19, we have a crucial moment in the history of academia. There was the 1994 meeting in Marrakesh, Morocco, organized by the World Trade Organization, and they came up with this great idea: a General Agreement on Trade in Services (GATS). I'm being sarcastic: it's not great; it's terrible. They come up with this weird idea to include education in trade agreements, and this is causing a fundamental shift in the way we look now at education.

Fast forward to 1999 in Seattle. It was very violent when they organized the WTO conference. What the protesters were saying is entirely forgotten. What happened was that education services were included officially for the first time in human history as a tradable commodity. This shift is minor for people looking from the outside, but it is enormously consequential. You don't make a difference in terms of profits, money and markets from fridges and cars to teachers' and students' education. I'm being serious, no irony or sarcasm: it is important for any educator to attend an international education fair because it's like a cattle fair. You realize that you remove the word students and then it's like selling cattle. It is dehumanized and horrifying. Learning is not part of the discussion. It is all about profits and markets. You replace students with anything you want, and you realize where the problem is. Human learning is not happening this way, and this is what changed education entirely. I fully agree with what Fleming said: hit after hit came over education (Fleming, 2021; Fleming et al., 2021).

There is no surprise that in the United States, you see books banned. That's a very concerning sign. I lived in a dictatorship where I was reading books in secret. I got in trouble in high school because I was asked what it is to be a patriot, and I was naively honest. I was called into the principal's office, and I was threatened that I'll be thrown in prison with my family.

JR: Oh my God!

SP: I was thinking they had no idea what I was thinking about. I read a forbidden book; it wasn't a bad book. It wasn't a toxic book. It was just not aligned with the party in power. That's it. You have the attack: burning books is not far, going against teachers, going against intellectuals. They are all fascist tendencies. I call it fascism because this is a serious threat to the world. And academia is at the core of that; it is under threat, and it is under threat since the WTO said: 'Forget about what you're doing. All that matters, in reality, is how much money you bring and how much money you give back. What are your books showing? You have the right balances. This is what matters in reality.' And the intellectual conversation just ended. It becomes dangerous when the managers have no respect, no concern and no understanding of why education is important. Why are these discussions important? They are deciding your future if you can pay your bills next month. It becomes very dangerous for anyone responsible or minimally realistic to engage in an honest manner.

I use the example of the crisis of academia and what is happening in reality. A report that was published a while ago, called Google Academics Incorporated (Tech Transparency Project, 2017), shows how these big companies, unethically and potentially criminally, buy influence and target academia on purpose. What this report shows that the best and most respected scholars and universities in the world are part of this research game where they publish research. They don't disclose any conflict of interest, and then they say, 'Oh yeah, this is great; it's going to help a student. It's going to help whatever.' But they are paid in reality by those who sell that technology, and this is where the problem starts.

This is a very serious, well-documented research. Another thing that is used by big corporations is to drag you down and destroy your life through lawsuits when you dare say something against them. Even if they know that they are going to lose, they know that you will go bankrupt and then your life will be destroyed. This report (Tech Transparency Project, 2017) is so important because it's uncovering that the space of honest intellectual conversation about the social and educational implications of what is happening in the world is tainted and that you have no protection.

When I wrote this book (Popenici, 2023a), I thought about the consequences. It can be laughable, but it's not. I was thinking about 25 years in higher education. This is where I actually started to work, and I worked across the world in Southeast Asia, North America, and Europe. In all these parts of the world, you see how every day is bringing a decline. What Fleming (2021) is saying that we are close to the end, I think, happened a while ago. I don't think we have universities as we imagine them. We still imagine them, and I go back to one of the 'saints' of the conservative movement, Disraeli, a great politician and a great intellectual. I can't be blamed that I'm using someone from the right. Benjamin Disraeli said at the end of the 19th century [in a speech at the House of Commons on 11 March 1873] that a "university should be a place of light, of liberty, and of learning". We all know, if we are honest with ourselves, that learning is pushed to the margins of the conversation in universities. The most disadvantaged parts of our university are schools of education and the least relevant, and also probably the most dull. Go to law! Go to business! That's where you see the power. That's where you see the influence. 'Education is creating teachers, they should be happy that they still have a job.'

A "university should be a place of light, of liberty, and of learning" (Benjamin Disraeli).

In terms of how students learn, I can quote you something that was published in The Guardian (Cassidy, 2023). They quote a student in Melbourne, and this is part of a report put together by Monash University: "International students are considered cash cows, not humans" (Cassidy, 2023). That's a real problem! Dismissing it is easy. The reality is that we have to admit that how much students learn is not at the core of what universities are doing now. Learning is just part of the mission statement, and that's why I said artificial intelligence is coming at a point of a very serious crisis for universities.

What do we have academics for? To analyse the impact of artificial intelligence with courage, intellectual vigour and substance; to warn society, this is what you're going to deal with. If you are too honest, then goodbye, research funds. You don't have research. You may not have a job. You say some controversial things that one executive in your university may be personally upset about because it's this new religion of technocracy. It's a real religion, and you have zealots with religion, and if you dare question the religion, you may end up like Giordano Bruno. You can be burned. That's a very serious context where we have this problem. We are part of the moment of the end. We have to decide now, in my opinion, what is next and who can survive in terms of institutions of thinking. Is there going to be a reaction from civil society? Is it going to be a political movement, to think? Universities are not able to attract the best and the brightest because they're just not paid; it's very hard.

This new religion of technocracy is a real religion, and you have zealots with religion, and if you dare question the religion, you may end up like Giordano Bruno. You can be burned.

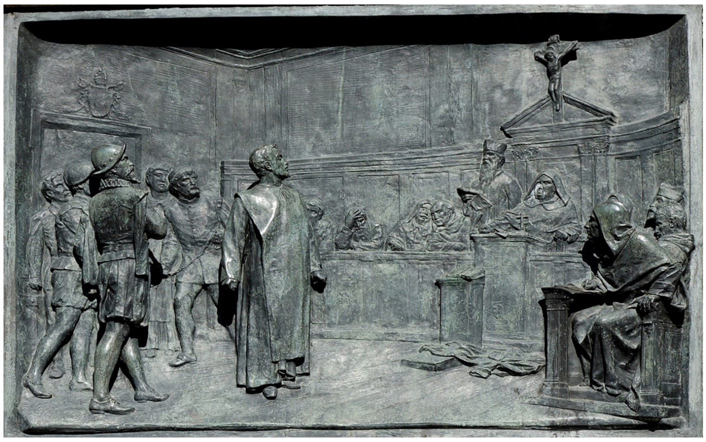


Figure 8: The trial of Giordano Bruno (1548 – 1600) by the Roman Inquisition. Bronze relief by Ettore Ferrari (1845-1929), Campo de' Fiori, Rome. Source: Relief Bruno Campo dei Fiori n1. (n.d.), public domain. At his trial, the Church authorities convicted Giordano Bruno, and he was taken to Campo de' Fiori, stripped, tied to a stake with a metal plate clamped over his tongue, and burned alive. His books were banned and placed on the Index of Prohibited Books.

It's a climate of fear, control and surveillance. I spoke recently about COVID-19 with a respected academic. I don't even think the country is relevant, and he said, 'Before I left my university during COVID, we had people from Human Resources jumping into our meetings on Teams and saying we are here to keep you under observation.' This is the culture of control. We know it's happening, and it's present. The climate of fear and intimidation, including intellectual intimidation, is very real. It's cultivated as a method of management. This is how you keep people in line: you keep them afraid. I'm lucky because I think there are things that are more important in life than your career. This comes with a certain recklessness. You can call it the advantage of being shot at during a violent revolution and getting away alive. You realize that maybe it's something more important than your next paycheck and your career. Fleming is spot on when he's talking about all the bullshit, all the managerialism and all the nonsense that comes now with academic life. But this is a very dangerous moment for all of us, and I'm not sure that politicians and civil society are aware of the great risk we face.

This is a very dangerous moment for all of us, and I'm not sure that politicians and civil society are aware of the great risk we face.

STYT: Thank you, Stefan, for providing the fascinating historical context of how the whole of education has become marketized. It reminded me of my own doctoral studies on education reform across the world, during which I argued against the idea of marketization, which happens in the context of the indiscriminate use of technology in education.

JR: Is there anything about Humboldt that is worth rescuing?

SP: I think it's very important to look back at that moment. I make this point in a book that I put together with a colleague on German education (Nickl et al., 2020). What

I tried to explain in the introduction of that book within German studies is the way we understand modern higher education since WTO happened and since rankings became a thing. All these things are more recent than we realize. We think this model has been around forever, but it wasn't at all. Higher education used to be free, and now we talk about graduates starting their life with $100,000 in debt. That was not the thing in our lifetime. It's a very recent thing.

The modern idea of higher education is a creation of concepts that were common in the German space, and scholars from Harvard got in contact with that. The ideal of *Lernfreiheit* (freedom of study) was immensely influential. In the late 19th century the concept was taken by American scholars and adopted by Harvard University, and changed the landscape of higher education since then. This idea reshaped the academic direction of Harvard to such an extent that in 1897 the only compulsory course left was the freshman rhetoric, while all other disciplines became open to be selected by students. Since then, students had the option to create their own academic and learning pathways. This is how the modern idea of a university was created. Of course, you see the need to improve on it, as the Humboldtian model was a discriminatory model that favoured elites. You have to remove that. Going back to Disraeli, higher education was the idea of having a place for intellectual nobility. It wasn't about the social class. It was about the intellectual class. I grew up very poor. It was a poor society because of Communism, and I was the poorest in the poor society. That's not a nice thing. I'll never resonate with the idea of elitism because I know what it is like to be on the other side. I know from experience and it follows you your entire life.

It's not that elitism that should be grabbing our attention in the Humboldtian model; it should be this idea of the "aristocracy of the intellect". That was an expression that George Steiner (and Max Weber) used. Disraeli said that a university is a place for the cultivation of the intellect for innovation, for research – this is the place of liberty and light. There is no liberty in reality in universities now.

You can go like this if you question the new mantra. I remember questioning the wisdom of putting all the money into one thing. This is what happened in Australia: some universities put their entire budgets into the next big thing – Massive Open Online Courses (MOOCs). The University of Virginia fired its leader because she wasn't fast enough to adopt the next big thing – MOOCs. I was thinking: this is madness! We already have MOOCs. It's called iTunes U – it is not a new thing.

It is not going to change structurally and fundamentally the universities. But the argument from the zealots from this new religion was reflected in an editorial written by an executive of an Australian University (Barber, 2013). Now that we have this free flow of information – this was in 2013 – universities had just one challenge: to sell their campuses to real estate developers. That was the idea, and I was thinking, 'Oh God, another manager discovers the Internet'. It is not a new thing! We had this thing for a while. This is not how things work, and this is not what a university is about. It is not about selling packages of stuff.

This is going back to Humboldt. This is what we have to rebuild if we are going to have a future, our civil societies, and this is my dream – to live in a society that is civil and free.

You can grow and develop your own identity, and you don't live with constant fear and surveillance, and the only way is to recreate these spaces that were imagined at that time by Humboldt. I think it's a very important project that, for a variety of reasons, must be revisited and not only Humboldt. It's a sum of ideas that we had floating at that time in that European space, and it was put together by Humboldt with the modern idea of the university.

STYT: I want to find out more about your explorations of the relationship between intelligence, imagination and AI. You remind me of how George Siemens et al. (2022) wrote about human intelligence versus artificial intelligence. At that point in time, we were still not so close to what we are seeing today in 2023. You argue that higher education institutions' key challenges are not technological but political, educational and cultural. Earlier, you talked about the need to be courageous. I particularly resonate with you that we, as academics, need to have courage in these very challenging times as we move forward. I think you partially answered the question already. Could you please elaborate a bit more beyond what you mentioned earlier?

SP: The first important point is to go back to look at what is human intelligence and what is artificial intelligence. Then, if we have the patience to look at this label 'artificial', it is not something that is positive. The first thing that comes up is 'fake', 'made up'.

It is in the name; the name is warning us before anything else that it is artificial. Don't take it as the whole thing. All findings have shown that the kind of skills and abilities that are captured by artificial intelligence are very different and very limited in comparison with human intelligence. This is very important to keep in mind when we talk about education and about what artificial intelligence can do.

I can give you one funny example that is true because artificial intelligence was developed from the very beginning in collaboration with the military. Since Turing, by the way, since he came up with the inception, the label was not yet created. That came in 1956. He came up with the idea, but since then, it has been a military project. We have to think about that it wasn't an educational project. It was a military project, and then that came with a very certain focus and a very certain set of capabilities that were limited to military use: identifying patterns, looking at pictures and identifying patterns of Soviet bases and so on. That was what it was used for. It was not used to solve human problems and human challenges. That's why artificial intelligence is presented now in a way that is going to put humans out of business. Well, why don't we have a solution for COVID-19? It is a virus that we can see under the microscope. Just use this magical thing and get rid of COVID-19 because I don't want to get sick. Not even vaccines work so completely. They work, of course, but they don't get rid of it entirely. It's important to look at what key challenges humanity now faces.

When I look around, I don't see challenges with technology. I have magical technology to turn on my light. I'm just saying, 'turn on the light', and it's turning on. "Play the music", it's playing and so on. From a technological perspective, my car is magical: it is doing stuff that I never imagined a decade ago that it was going to do. We don't have a big challenge in technological terms. I think we are doing quite well but look seriously at what is dangerous: nuclear annihilation. We have a war at the heart of Europe. That's a real challenge, in my opinion. So we have wars, we have genocide, we have the rise of fascism. We have a crisis that is puzzling the minds of experts in climate change. Those are serious challenges, they are existential, not minor challenges. There are economic challenges and inequity challenges. The kind of society is proposed where the rich get immensely rich. Again, we are talking about innumeracy. It is the same as illiteracy. It's just that people cannot comprehend the kind of money some people make. As they sip their coffee they make a couple of hundreds of thousands of dollars – in just a few minutes – working people are starving. And we have international students who say, 'I came to study, and my relationship with food is changed because I can't feed myself.'

That's not a minor challenge. I think this is where we have challenges, and we are looking at technology. What can we do next? This is incomprehensible to me. I have to admit, it's not the kind of serious challenge. The majority of people in the world breathe in polluted air. And then our problem is technology. I simply don't get it. Polluted air is linked with Alzheimer's disease, public health and with survival. In some societies, when healthcare costs become a burden for society, that's a very serious problem. And then the solutionism of technology is simply not delivering on its promises. Remember IBM's Watson? It was the magical solution not too many years ago, we don't talk so much about that. It came into healthcare promising a solution for cancer (O'Leary, 2022).

I have bad news: it didn't happen. [All laugh.] This is a mechanism that is important to keep in mind: you have this new technology, and it's this magical thing that is going to solve cancer, and then you have a wave of media stuff, and then millions of voices in the public space, dealing with the same mythology and the same religious feelings of how this is going to work and after that, we forget. It's what actually was promised. Now I read that artificial intelligence is going to solve climate change.

> ***This is a mechanism that is important to keep in mind: you have this new technology, and it's this magical thing that is going to solve cancer, and then you have a wave of media stuff, and then millions of voices in the public space, dealing with the same mythology and the same religious feelings of how this is going to work and after that, we forget.***

Until then, artificial intelligence is a serious problem for clean water. That's what we have in reality happening right now because you have increased computing power, and this requires clean water. Clean water is a problem across the world, and the latest studies show that lakes across the world are shrinking. This is our reserve of drinking water, and this is disappearing fast. Artificial intelligence is using what is crucial for our survival, and the promises that it is going to somehow magically solve our problems are questionable. There is no researcher worth their name saying that AI will solve climate change. Serious challenges are pushed aside by propaganda. Unfortunately, I lived almost two decades in a dictatorship: it's like a vaccine. I believe in vaccines. They help you because you create antibodies, and for me, it's the antibodies for propaganda. I just don't believe it, and I can smell it immediately because it's following the same formula.

I visited recently a country with my wife, and then my wife said, 'I don't know why and you may think this is crazy, but I felt like this is Communist Romania.' And I said, 'because fascism is the same, it doesn't matter what flavour, the shade can be red, brown, or green. It's fascism, and that's why you have this feeling.' This is how propaganda works. It is common that we have to look beyond propaganda if we want to deal with solutions.

ST: Now we're going beyond your very important book. ChatGPT has fired the public imagination with a vengeance. Within a couple of months, the AI chatbot has hit more than 100 million users. There is an increasing number of rival chatbots (Rudolph et al., 2023b). What are your personal experiences and impressions of the new generation of large-language-model-based chatbots?

SP: Well, thank you so much, Shannon. You don't have simple questions, right? [All laugh]. The first thing that I can say about this is that ChatGPT, as you said, is one of the things in a long list of shiny things created by artificial intelligence. But I think that ChatGPT is a very important development because it's based on generative artificial intelligence. The advantage of ChatGPT is that it came as a mirror for education, and it shows where we are because it shows that we are completely unprepared for reality. We lose perspective on what matters when we have this engine of mediocre text.

When you look at what GPT and other large language models create in terms of generative texts, they are the most advanced. If you understand how artificial intelligence works, you realize that this is the future of generative text, and you put that text, no matter how good it is, next to what James Baldwin was writing, for example. James Baldwin is coming with fire, ideas, creativity, emotions, and change. That's entirely missing from AI text. It is plausible; the text, syntax and grammar are great but the message is all mediocrity.

We created a space in education of generalized mediocrity because it's better to have mediocrity when you want to control a system that is focused on production and profit. So we managed to create this, and then you have a certain model and view of education where it becomes normal to have a class of 1,500 students. That's not higher learning. You can find a different label. You check how students learn through a standardized test; that's not learning at all. In this context, ChatGPT and other similar engines come and head straight at the core of the problem because you get that immediately in 30 seconds. That's expected and glorified mediocrity. That's what you have automatically, and it is available to everyone. The way institutions of education reacted shows also the profound incapacity of serious thinking about education. The intellectual endeavours are extracted and dead, and now you have the kneejerk reaction. The ridiculous, laughable reaction of decision-makers in education was 'we are going to ban it'.

> *The way institutions of education reacted shows also the profound incapacity of serious thinking about education. The intellectual endeavours are extracted and dead, and now you have the kneejerk reaction. The ridiculous, laughable reaction of decision-makers in education was 'we are going to ban it'.*

I'm not against technology at all. I'm for a responsible use of technology with a critical perspective on technology. My first reaction was, 'You think you are going to ban this? Then you're going to have to ban the internet.' Because this is right next door, and this is exactly what happened. Now we have a couple of hundreds of AI apps released per month, if not per day, God knows. You have so many things that you don't know. Which one should I ban first? Now Microsoft and Google give you this as part of their normal navigation. So what are these schools going to do? Just months ago came the idea of banning it. You're going to ban Google? You're going to ban Teams? You are going to ban Skype? I spoke with someone yesterday on Skype. You have a permanent presence, your AI assistant. I can ask that thing to write an essay for me that I can submit as my assignment.

The problem we have is that we are completely unprepared due to our glorifying of technology. The very amusing thing, and it is laughable, is that technology is showing us how far we are from what we should do. The risk is that we are going to lose our legitimacy entirely. It's a massive challenge because we turned assessment into this industrial process of mass assessment, with no quality, no look of originality, and need of substance. This is what you have to submit; use citations; use good grammar, and good syntax, and you don't do massive mistakes. It's good to go. You graduate. You're good. It's fantastic when we turn the whole system to this; it's just that we lost the plot, and then it is a disgrace. Technology is showing us how much we are at risk. It is striking at the core of education. This is a consequential moment. This is going to change entirely what we are doing for good or for bad.

JR.: How do you foresee will ChatGPT and other generative AI change higher education? For instance, when hand-held calculators became common about half a century ago, there was lots of moaning that this would erode students' maths skills. And at some point in time, spell checkers were banned, and their use by students was regarded as cheating. So is ChatGPT akin to the introduction of the calculator or spell-checker, or is it something more revolutionary?

> *The problem we have is that we are completely unprepared due to our glorifying of technology. The very amusing thing is that technology is showing us how far we are from what we should do. It's a massive challenge because we turned assessment into this industrial process of mass assessment, with no quality, no look of originality, and need of substance. This is what you have to submit; use citations; use good grammar, and good syntax, and you don't do massive mistakes. It's good to go. You graduate. You're good.*

SP: It is very important to deal with this problem. ChatGPT and generative artificial intelligence are structurally and completely different, and it's a very simple fact. Calculators are dealing with things that are not on itself sufficient to pass an exam. Spell check just helps you to write better English or whatever language you speak. It's not changing much in terms of text. Now you have a tool that creates the whole text of the assignment for you, and we have shifted the whole endeavour of education to assessment. This is the core of what we are doing. It is at the point that some universities just outsource teaching and learning to videos that are pre-recorded from five or seven years ago, and tutoring, if there is any tutoring. This has been documented well (Smee, 2023). The use of media for teaching is happening around the world, and tutoring is left to people almost off the street: no skills required, no knowledge, no nothing. It's just pretending that students have some presence there. The real weight is on assessment.

No assessment can be replaced entirely by ChatGPT, and generative large language models. Because assessment is not asking you to come up with anything of substance, it's not asking you to come up with original ideas, God forbid. You have a big problem. It's fundamentally different. This is not spell check. This is hitting the model at the heart. It's a spike in the heart of the model of education as we have it today, and it's going to be a massive change. It's truly revolutionary, not because it's going to bring something necessarily better. It is revolutionary because it's going to ask institutions: what are you actually doing?

> *Generative AI is a spike in the heart of the model of education as we have it today, and it's going to be a massive change. It's truly revolutionary, not because it's going to bring something necessarily better. It is revolutionary because it's going to ask institutions: what are you actually doing?*

I did nationwide research on student motivation for learning. It turned out immediately that students find motivation for learning central. There is no surprise. There is a long literature showing that student motivation for learning is crucial for the quality of learning and the way they see their academic careers. In terms of motivation for learning, if we don't change our project of education, this is striking again at the core of motivation for learning. Because if you reduce learning to assessment and the assessment can be outsourced by students to just write a sentence and think a bit about the text you have no motivation. Why would I do that? Why would I learn anything? Because I can just give it this AI solution. The kind of implications for universities are massive.

For society, let's imagine that you have engineers not interested to learn anything, and they build bridges. I wouldn't use that bridge built by graduates who used generative AI to pass their exams. It’s going to fall apart. I wouldn't go to nurses who do that, and so on and so forth. It is very serious. I fail to see any politicians so far in Europe, the United States, or Oceania who are actually looking at these aspects. There is no discussion about how is this going to impact students’ motivation to learn. How is this going to impact the nature of our universities?

The real concern is how can we jump faster than China to do artificial intelligence. How can we adapt better than this country and that country? There is a fear of missing out on the global scale and without thinking about what we are actually doing. It’s just ‘we have to do it faster than others’. It's a concerning moment.

ST: ChatGPT and other chatbots (like Bing Chat, Bard and Ernie) are the latest shiny thing in the long history of AI in education. It may be too early to determine ChatGPT’s (and GPT-4’s) place in that history, but what are your preliminary thoughts? How should higher education institutions deal with generative AI, in your view?

SP: I partially covered this previously. I'll just focus on institutions of higher education because it’s going to be immensely consequential for universities. It is a revolutionary moment, it is the mirror that is put in front of us. Using generative AI is the kind of mediocre thing you do, and this is not higher learning; this passing assessment is done by a statistical model of predicting what word and then sentence comes next. The mirror is saying: “you should be ashamed of yourselves, guys!” This is what we have. This is the message.

But how universities should deal with that, in my opinion, should be a step back and question the new religion to have a serious discussion rather than the fear of missing out. What is our model of teaching and learning? Is teaching and learning still important?

I look at teaching and learning and my personal experience. I shouldn't use an anecdote, but I confess to this guilt, I'm terrible. When I wrote in 2017 about artificial intelligence in higher education (Popenici & Kerr, 2017), universally, the feedback from peer reviewers came that this topic doesn't exist in our field. Why don't you write about learning analytics? Not because they really care about learning analytics, I suspected, but because learning analytics was the fashionable topic. It was what the whole field was dealing with. But there was no interest in this, and now you have a stampede of experts in artificial intelligence. I think I'm going to get lessons in artificial intelligence from the cleaner at my university. ’This is what it is, Stefan, and this is how you should use it.’ It's just everybody, and I'm grateful that they’re suddenly interested in my topic of research. I'm grateful that my book came at the perfect time. That's very lucky.

But the point for institutions is that they’re still lacking any serious concern about the impact on learning and teaching. You can see this in the literature; in the kind of research projects approved or already running. The interest in artificial intelligence in higher education is not on the impact of learning and teaching, and unfortunately, this is the most significant impact. A university in the United States did a study and assessed how jobs are going to be replaced by artificial intelligence (Felten et al., 2023). Then they identified the top occupations exposed to language modelling where in universities, teachers of various disciplines are going to be replaced by these things.

It's a stab in the heart by a group of experts. It shows if you think about it, if you read literature, if you look at facts and if you look at studies like this, you realize that the most affected space is going to be learning and teaching, and it is still marginal for research, academic discourse and politicians. This is what should be at the centre of what we are talking about because if we don't have learning and good education, we don't have a future.

It's not a compliment: one of the countries that I genuinely love and admire is Singapore. I genuinely love and admire it because education is at the core of that country's project. In my country, higher education is really concerning. There is no interest in substance, and there's no interest in crucial areas like learning and teaching. How is this going to impact students? How is this going to impact teachers? How is this going to impact our model of education? 1,000 students in a classroom, is this making sense in the new context? So this kind of questions should be critically examined.

STYT: Talking about the political aspect of higher education, I just recalled something that I lamented to Jürgen a few weeks ago. I was commenting that the US Senate hearing was grilling TikTok CEO, Chew Shou Zi. I was watching it and thought they were barking up the wrong tree. You have your backyard on fire with an AI crisis, and they are not doing anything about that. Well, soon enough after that, they had this Senate hearing with OpenAI CEO Sam Altman. They were so aggressive towards Chew but so civil with Altman, almost like doves. They should be very suspicious when you have a businessman like Sam Altman asking the government to enact stricter laws in managing the growth of generative AI. It was either to stamp out competition and enact laws that favour OpenAI or to ask for a way in which he can manipulate the legislature that will favour OpenAI. Obviously, politicians know very little about AI technologies and their impact.

SP: Altman is a very shrewd and cynical operator. I remember he was quoted in Forbes a couple of years ago. And then he said, “AI will probably most likely lead to the end of the world, but in the meantime, there’ll be great companies” (meaning, companies that are making a lot of money) (cited in Popenici, 2023b). [All laugh.] This is the idea. When you look at what he's doing, he's bringing a double-digit billion

US dollar amount into his company, he is making a lot of money. Obviously, when he's bringing a lot of money to this mind-boggling extent, it shows that this is his main interest and his main value. I'm just being logical here. You have to take what he's saying with some healthy suspicion, and here is my problem with what he was saying. He is saying that you have to regulate us is a very typical distraction of big tech because they don't observe any rules of common sense, decency, and laws.

You see what Silicon Valley startups are always doing and how there is no concern for workers. There is no concern for those who are going to be affected. It's just what happened. We start to film people around the world and then take photos of all streets and stuff. Is this legal? ‘We don't care. We are just going to do it.’ This is the modus operandi for these companies. I've never seen social responsibility in reality associated with their endeavours. So when Altman is calling for regulation, it's a bit too late. You scrubbed the Internet, you don’t care about copyright. These texts were created by someone, and they’re creating now value for you personally. He didn't care about regulation. He used poor people in Kenya to be exposed to some of the most horrifying and traumatic content (see Rudolph et al., 2023b), and he didn't care about the consequences of that work on their life. Suddenly he’s concerned about regulation.

I think that in reality it's more a concern about distracting from specifics because I can do the same. It's just when my students, let's imagine, would say, ‘Stefan, your teaching sucks because you don't have time for us; you skip over topics; you don't care about your assessments’. And then my reaction can be, ‘the field of higher education is problematic’.

Let's go back to what was the discussion. Let's go back to the specifics. Let's talk about it before it ends the world. How is this going to impact the workforce and learning? All the things that I mentioned before. Is this a trick to pass responsibility back to the regulators when it's already too late? So what he's doing in reality is just, ‘oh, we are the good guys here. It's you who should do the work of regulating us’. You didn't care until now. You break all the rules of common sense and of humanity (think about the poor workers in Kenya as an example). Now, after it’s a fait accompli, they say, ‘come, regulate us’. I'm sorry, but I don't believe it. It's very significant that this is where we are, and we miss the specifics.

JR: I’m wondering if you have any kind of final thoughts on assessment in Samson's question?

SP: This is important because, as I said, we created this space of mediocrity. This is where assessment has the greatest weight and importance in our educational project. Now large language models strike at the core. Assessments are crucial in any scenario. Imagine the future and how we organize higher education. Of course, we should have assessments that are more authentic. I'll go back to what I mentioned very briefly before. Higher education was not always ‘modern’ in the 20th century. It was different. We have to make some very important choices if we are going to create meaningful assessments.

In my own education, seminars were a very important part of assessments. Your work in seminars etymologically comes from the Latin *seminarium*, meaning seedbed. You plant the seeds of ideas, critique, engage with the text, and become part of the learning process as an informed contributor and participant in the conversation about the topic. This is removed when you reduce assessments to multiple-choice questions or an essay with no meaning. But when you think about assessments in a more personal, significant way, ChatGPT is becoming marginal. It is becoming, as it should be, an assistant.

Your pocket calculator is not taking the crucial role that it is playing now, and then I don't blame students for using it because, first, it's tempting; it's doing the kind of meaningless work they are asked to do in a couple of seconds. Why not? I mean, they have their own lives. They have their own challenges. They deal with this faster and more efficiently. (By the way, it has better syntax and grammar because we don't teach grammar in Australia.)

This is why it's important to look at assessments, but looking at assessments to deal with large language models such as ChatGPT requires a rethink of the project of education that we have. Lecturers discovered that in one class, 44 students used ChatGPT. My reaction was ‘only 44?’ [All laugh.] ‘Or you're not really good at catching them.’ Seriously, it's just that, or you got the laziest who just dropped entirely what ChatGPT gave and then did not even bother to look over it and think a bit about the text. Anyway, 44 students used that panic and the reaction we had was to ban it. I said, ‘no, you can't ban it. It's not realistic. It's ridiculous’. The solution is to use a different approach. My solution is to ask students to use ChatGPT for this particular assignment. It was a literature review. And after that, I'd say, ‘the next part of your assignment is to see what is missing.’ Well, this is an easy trick because in order to see what is missing, you have to see what is there. So you have to read, and after that, you ask them to see what ChatGPT gave you and what is wrong. There are many hallucinations (made-up stuff) and factual errors.

In order to find out what is wrong, you have to know what is good. So you ask them to learn. But this is the problem: in time, we are going to have large language models that are going to cover this because they are going to have fewer errors. This is a punch. It is not going to solve forever the problem, and it's an illusion to think that it is going to solve for the long term the problem. We have to keep in mind that we have ChatGPT-3.5 for only half a year. Six months changed the conversation entirely in education. This is the pace of change. If we don't change structurally and substantially the way we look at education, we are going to have a process where students are going to use ChatGPT-like technology to submit their assignments. Lecturers overworked with 1,500 essays are going to submit these to AI for assessment, and then you end up with a process where nobody's learning anything.

JR: Exactly. [All laugh.]

SP: And I wouldn't go on that bridge again.

JR: Thanks for saying this so well. We may arrive at a situation where the lecturer creates the assessment using ChatGPT, then the students write the assessment using ChatGPT, and then the lecturer will mark the assessment using ChatGPT, and nobody learns anything.

We may arrive at a situation where the lecturer creates the assessment using ChatGPT, then the students write the assessment using ChatGPT, and then the lecturer will mark the assessment using ChatGPT, and nobody learns anything.

SP: This is a very real scenario. Apple just banned some employees from using generative AI. When I was discussing this, I thought that's a very shrewd corporate talk because they say 'some employees', and you think, 'oh they're very flexible'. Some employees are free to use ChatGPT. Basically, they're free to use whatever they want, but people with anything of substance are banned to use ChatGPT. Why? Because they know that people use it, and universities should be aware that not only students but academics use this to create their own content and their own assessments and then their own stuff. So the process that you just summarized, Jürgen, is very real. It's not a stretch of the imagination, it's happening.

JR: What will generative AI do to graduate and academic employment? Daniel Susskind (2021) recently wrote a book titled *A world without work*, and Aaron Bastani (2019) a manifesto, wonderfully titled *Fully automated luxury communism*. What are important skills and competencies for graduates to become and remain employable? How do you see the future of academic work in light of generative AI?

SP: I'm familiar with Susskind's book. The future of work and unemployment is an area that frustrates me because there is no concern about these corporations making immense profits on the kind of impact in people's lives, the kind of social discontent and social tensions this is going to cause. It's definitely going to impact massively on the future of our graduates. They are going to face an even more difficult employment situation when they graduate. I have to be honest: I can't figure out how this is going to look. It depends very much on the kind of society we will have. If we are going to cultivate responsibility and civil society, then we can hope that this is going to be managed somehow. But if we are going to go to highly extractive practices and see authoritarian impulses developing even more, then for individuals graduating now, the future is bleak. I have more of a dystopian view of the future of work, and this is truly concerning.

When it comes to skills and competencies, universities again dropped the ball badly. Because the whole logic of running universities was reduced to markets and profits. As I said earlier, the most important parts of the university are considered to be business and law. It is what it is. There's nothing wrong with that. It's wrong that that focus was used against the humanities, and we are going to see some consequences. So far, what we see is that some of the most successful employees working in AI come from the humanities. I think it was in the Washington Post presenting a profile of someone working with artificial intelligence, large language models, and she was making a lot of money. And then she said, 'I have no idea about computers. I'm just a user' in the sense that she was not an engineer. Of course, she had an idea as a user, but not coding. She was a graduate of humanities. And that made her excellent at dealing with the kind of challenges posed by using large language models.

It was the Cinderella of academic life; humanities were less important for universities. I'm not going to defend humanities against anything else. I think they have a very important role, and I'm thinking this 'anti-stance' is very damaging and at the core of the problem. I'm not going to follow the same logic. I think it's important to have excellent business schools, excellent engineering schools and excellent humanities. Without any of them, I can't imagine progress. But in the book (Popenici, 2023a), I use an example that is fascinating to me about what universities are doing without thinking of consequences. This comes from an accident in research.

It was the story of two researchers. One researcher who studied something found a disproportionate representation of terrorists among graduates of engineering, and he found that's a very interesting coincidence. He started exploring that. He joined forces with another researcher, and they wrote a book. Long story short, they wrote a book called *Engineers of Jihad* (Gambetta & Hertog, 2018). When I read the first time about their research, my immediate hypothesis was, of course, that they have more terrorists with a background in engineering because they know how to make damaging stuff. Interestingly, they're not the builders, they're the ideologues. They come with the ideas, they are the masterminds, not the builders of stuff that is killing people. It's also fascinating that after that, the next step for their research was to look at other terrorist movements, not only fundamentalist Islam.

They found again, in the extreme left and in the extreme right, a disproportionate presence of engineers. This is the argument I made earlier with the missed lesson of Nazi Germany that technology without values, without humanity, is very dangerous. But the universities went in that direction at full speed. 'We don't need humanities. We don't need philosophy, only practical stuff'. The first time I heard that 'we don't care about ideas; we care about practical stuff' obsessively was in Communist Romania in the worst years of the dictatorship. It's a massive lie, but that lie was exactly at the heart of that narrative. That's exactly what they sold. 'We don't waste time thinking, we do stuff'. Well, it fell apart and led to poverty. The whole communist bloc, because they were doing stuff without thinking, that's why they fell apart. And all these things should matter.

Technology without values, without humanity, is very dangerous.

This kind of thing is the essence of my answer: Thinking skills, values, creativity, and the ability to think critically. When I lived in a dictatorship, the most obsessive word used in the worst years of communism was "democracy". Everything was democratic. [All laugh.] It was all about citizens. If I hear again about critical thinking, I think I'm going to develop an allergy because you hear the words, and you don't see it in practice. This is not how it works.

If you hear the words 'critical thinking', but there is no critique, and once you see a genuine critique, you're accused of being the enemy of the people, that's the real problem. We have to cultivate the genuine capacity to put a question mark and come up with our set of arguments, engage with another set of arguments and debate the idea. We don't do that because we care about passing knowledge. We don't cover the set of skills in reality. Mostly, of course, I'm talking about general problems. This is the beauty. This is why I'm still in university. Because you find exceptions, you still find extraordinary minds, you find extraordinary people passionate about their students. But I don't want to live in a system where this is an accident. I want to live in a system that works together to advance society, not to think, 'Oh, you know what happened today. I found someone who's interested in the students learning something'. It shouldn't be extraordinary.

JR: What you say about the use of language is very apt. This is why I'm such a huge fan of Orwell's (2021a, b) *Nineteen eighty-four* and *Animal farm*. The whole idea of doublespeak, it's still extremely powerful.

STYT: We are now looking a bit more into the future. Nick Bostrom, a philosopher at Oxford University, has written a book about *Superintelligence*. He is cautioning that after computers have achieved Artificial General Intelligence (AGI – which essentially means that they can think and act like humans), superintelligence may be close. This would mean that machines would be potentially exponentially more intelligent than us humans. One possible outcome would be a humanity extinction event. A more benevolent one, perhaps, would be that computers keep us as pets. And the movie *The Matrix* immediately comes to mind. What are your thoughts on this?

SP: I read his book a while ago, and I was struck by the number of assumptions he's making. Because he's obviously a very smart man, I don't think that these are mistakes. I think they're intentional. The fundamental mistake is to equate artificial intelligence with human intelligence. And when you ignore this difference, you can reach the idea of superintelligence.

I can give you an example that I found funny because this is a topic that is a bit frustrating for me, and I'll explain why. The example comes from the army. They have been the most passionate users of artificial intelligence since 1956, and they asked some military guys to beat artificial intelligence. They used the most advanced, as you can imagine, and the artificial intelligence was defending a perimeter, and soldiers were tasked to beat artificial intelligence and bridge the perimeter. Do you know how they managed to do that? They actually found a solution in playing video games, and all of them beat artificial intelligence when one of them found and used a cardboard box, while artificial intelligence was looking for a silhouette.

It was not identifying. It was just a box. Another soldier was dressing as a tree or whatever. Of course, you can finetune the artificial intelligence, and you see moving things. But that's not the point. The point is that artificial intelligence operates very differently from human intelligence because you don't have to finetune a human that is standing guard and say, 'no one should pass', and then if it's a box that human is going 'oh, it's just a box walking, that's fine.' We operate differently; we are different. Superintelligence is a myth. That is a distraction from the real problems of artificial intelligence, it is not a real thing.

I don't have the space, and I probably don't have the skills, but I am reading now a book that is called *The myth of artificial intelligence* (Larson, 2021). Erik Larson, who wrote the book, is a computer scientist and tech entrepreneur. He worked for DARPA, the Defense Advanced Research Projects Agency of the United States Department of Defense, responsible for the development of AI for use by the military. So he knows what he's talking about. He's unpacking specifically the myth of superintelligence of artificial intelligence, it is not a real thing. It's not how artificial intelligence works. It's not a real possibility, and it's very unscientific.

Another big problem with the whole idea is the biggest distraction of dystopian versus utopian views – these extremes of 'super AI is going to wipe out the human race' versus the utopian thing that you find mostly in higher education: 'Oh yeah, the next big thing is going to solve all our problems'. Hold on! We have even more problems than before. What's in between is missed, and this is the most important part. If artificial intelligence is going to destroy our modern model of learning and teaching and is going to undermine substantial learning, what kind of future are we going to have? This is what is going to wipe out the human civilization.

It wasn't any technology that wiped out the Roman empire and as a civilization. It was themselves with corruption and stupidity. That's what wipes out civilization, and we should pay attention to what's in between this utopian and dystopian discourse. When people spend all their energy on mostly impossible projects and miss the day-by-day manipulations and problems, we have a very propagandistic reflex to use a distraction.

ST: Could you please tell us more about your own schooling and university education? You grew up in socialist Romania, and admirably, you speak English, French, and Italian, in addition to Romanian. Your Bachelor, Master and PhD (in education sciences) degrees are all from the University of Bucharest in the post-Ceaușescu era. Could you please reflect on your own education? Were there any formative experiences that influenced you to become an academic?

SP: That's going to put me in a space that I avoided, not necessarily intentionally. I just don't find reasons to talk about that part of my identity. I grew up in Communist Romania, and in the 70s, Romania was the most open country in the

Communist bloc. That's why the President of the United States visited for the first time a country in the Communist bloc, Romania. And after that, the leader of communist Romania became, at an accelerated pace, increasingly crazy and authoritarian. I shouldn't say crazy. It's not professional, but he was totally crazy. [All laugh.] And very authoritarian.

Figure 9. Nicolae Ceauşescu and Kim Il Sung during the party and state visit to North Korea. Source: Fototeca online a comunismului românesc (1970).

The idea was to create in Europe a country like North Korea. It was one of the most extreme dictatorships that you can imagine, and that shaped my identity a lot because I remember that as a teenager, I passionately hated Communism. I grew up in a house full of books. My father was a book hoarder. When I met my wife, she said, 'I've never found a house with so many books.' It was thousands and thousands of books everywhere. I grew up reading, and that was lucky because, at that time in Romania, you didn't need to be rich to have books. They were quite cheap. Also, my father was a librarian, and I was sneaking into the part of the library with forbidden books that were locked. But I had the key, and I had access to a wide range of ideas. That time was very important for me, and the most formative part was that I had to drop out of day classes because I couldn't afford to go to them. I had to work, so that was the perfect mix for me for daily contact with real life and daily contact with intellectual life. I don't think now, as an adult, that you should have just one. You should have both for a proper understanding of how the world is going.

In 1989, it was a bloody revolution. Because I passionately believed that the regime should go, I was on the streets, and that was a very intense experience. They're shooting to kill, and I remember that people were shot and killed next to me. So I graduated from high school, and I was working, and I realized that at one point in '92, I had to go to university. It was very difficult because we had limited access. There was a quota, and university places were very limited. Most students had private tutors because one of the things in Communist regimes was that you were not equal at all. They didn't care about weaker people. When I decided that I hated the regime, the next thing for me was to go to the public library and borrow books from Mao, Marx, and Lenin. I'll never forget the look on the face of the librarian. She looked at me like, 'you're young, and you lost your mind completely'. [All laugh.] But the next thing I realized reading was that the most aggressive propagandists (called *Politruk* – political appointees) had no idea about Communism whatsoever. So that was really interesting. It was just a dictatorship, fascism with the red shade.

When I went to university, I was lucky because the university was very traditional. But Ceaușescu banned psychology and pedagogy. What was called education sciences and sociology was also banned. They're considered dangerous disciplines, and as you can imagine, academics managed one way or another to flee from the country. So I went to university at the moment when many good professors were back in this newly re-established faculty. Then, many academics from abroad came back from the UK, Germany and parts of the world that were completely foreign to me. They came with a very different way of teaching and dealing with us. We lived in a generation that was just out of a revolution with very naive ideas but full of passion. One naive idea was that we could change this country; we could make it better.

I grew up in this constantly revolutionary mode, I hated communism. I was a terrible student in high school. I was the best student in university. I was living as a student in a traditional university, and most of my professors were coming from a very traditional background. And it drove them crazy to have the best student in that generation with very long hair. [All laugh.] It was completely against their idea of what a good student should look like, and then when I finished university, I had short hair like at present.

There was a lot of enthusiasm for learning in my generation at that time: 'We can make it better through education', and that's why it's an important project. I still keep this at the core of my beliefs. It was a very different model of education. We had groups, and we had seminars, and then we had viva voce exams. We had the real personalised education where you can actually see people in front of you, and then these people can actually hear you. In seminars, what was the model? Attending seminars was compulsory. Many unfortunate colleagues dropped out because you had to attend seminars. But if you attended the seminars without reading the books – not the book, the books you have to read for every seminar – you were a subject of ridicule, and you lost face. You didn't want to be in that position, so you had to be knowledgeable, and then you had to engage. That's why you were there. So we had vicious debates, and probably they'll call the police, students yelling at each other with arguments. They were very passionate, I remember the first year was Chomsky versus Piaget, and that was a very passionate debate. But we were talking about this, and we really engaged with that, so that stayed with me. That's genuine education; you engage with content. It is not about ticking boxes at the end, and this is your assessment. It's what you do with knowledge and how you can use it for your part to contribute to society.

Figure 10: Stefan Popenici in the early 1990s.

Now I live in a privileged position. I think I am privileged from all points of view, and I think that the responsibility is to care about those who are facing new challenges, the new graduates. This is how it reflects on my own education. It was a formative time, it was very important for me. This is why I became an academic, it was the idea that I have to give back.

STYT: When you mentioned about 1989, I remember watching on TV everything that was unfolding in Europe. It was surreal to me. Those were very tumultuous years, with lots of changes after that.

SP: It was surreal being there as well.

JR: It sounds extremely frightening what you were sharing, Stefan, about people being shot and you being on the streets. That was very brave. My perspective is quite different because I was very lucky to be born in former West Germany, and so I didn't go through the velvet revolution in East Germany or the bloody revolution in Romania. Could you tell us a bit more about your future plans? And is there anything we did not cover that you would still like to talk about?

SP: I think we covered a lot, and I think we covered the most important parts of the topics. I don't think we missed any points. In terms of future plans, the biggest plan is a new book, it is called *Education in the age of artificial intelligence*. It is about narratives of humanity, higher learning and the challenge of artificial intelligence. It's still focused on what I consider to be the greatest challenge for universities. This book is already keeping me awake at night because I'm thinking about how I can address this. Because of this, I think I'm going to reduce my public speaking events because I try to be entirely focused on that.

Thank you so much for your work and for your time, and for your excellent questions that made me think and stay engaged. Thank you, Jürgen, Shannon and Samson!

JR: On behalf of us, thank you very much for this fantastic interview, Stefan!

References

Baldwin, J. (2001) [1953]. *Go tell it on the mountain*. Penguin.

Barber, J. (2013). The end of university campus life. *ABC Radio National Australia*. http://www.abc.net.au/radionational/programs/ockhamsrazor/5012262

Bastani, A. (2019). *Fully automated luxury communism*. Verso Books.

Black, E. (2001). *IBM and the Holocaust: The strategic alliance between Nazi Germany and America's most powerful corporation*. Crown Publishers.

Bostrom, N. (2017). *Superintelligence. Paths, dangers, strategies*. Oxford University Press.

Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability bias in tech*. MIT Press.

Cassidy, C. (2023, May 23). At the brink: How international students in Australia are ending up homeless and hungry. *The Guardian*. https://www.theguardian.com/australia-news/2023/may/23/at-the-brink-how-international-students-in-australia-are-ending-up-homeless-and-hungry

Eknoyan, G. (2000). Michelangelo: Art, anatomy, and the kidney. *Kidney International, 57*(3), 1190-1201.

Felten, E. W., Raj, M. & Seamans, R. (2023). *How will language modelers like ChatGPT affect occupations and industries?*. SSRN: https://ssrn.com/abstract=4375268 or http://dx.doi.org/10.2139/ssrn.4375268

Fleming, P. (2021). *Dark academia. How universities die*. Pluto Press.

Fleming, P., Rudolph, J., & Tan, S. (2021). 'Never let a good crisis go to waste'. An interview with Professor Peter Fleming on dark academia, the pandemic and neoliberalism. *Journal of Applied Learning and Teaching, 4*(2), 110-120. https://doi.org/10.37074/jalt.2021.4.2.14

Fototeca online a comunismului românesc. (1970). Nicolae Ceauşescu and Kim Il Sung during the party and state visit to North Korea [Photograph]. *Wikimedia Commons*. https://commons.wikimedia.org/wiki/File:CeausescuKim1971.jpg

Gambetta, D., & Hertog, S. (2018). *Engineers of Jihad: The curious connection between violent extremism and education*. Princeton University Press.

Isaacson, W. (2017). *Leonardo da Vinci*. Edizioni Mondadori.

Jose, A. M. (2001). Anatomy and Leonardo da Vinci. *The Yale Journal of Biology and Medicine, 74*(3), 185-195.

Kissinger, H. (1975). Cold War: Kissinger conversation with President Ford ("Britain is a tragedy") [1975 Jan 8, declassified 1984]. *Margaret Thatcher Foundation*. https://www.margaretthatcher.org/document/110510

Larson, E. J. (2021). *The myth of Artificial Intelligence: Why computers can't think the way we do*. Harvard University Press.

Leonardo da Vinci (1510). Surgeon Educational Resource of Leonardo Da Vinci's Sketch of the human brain and skull [digital image]. *Wikipedia*, https://en.wikipedia.org/wiki/File:Leonardo_Da_Vinci%27s_Brain_Physiology.jpg

McCarthy, J. (2006). *Technology and the position of women*. Stanford Computer Science Department: Stanford University, http://jmc.stanford.edu/commentary/future/women.html

Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. Public Affairs.

Nickl, B., Popenici, S., & Blackler, D. (2020). *Transnational German education and comparative education systems*. Springer International Publishing.

O'Leary, L. (2022, January 31). How IBM's Watson went from the future of health care to being sold off for parts. *Slate*, https://slate.com/technology/2022/01/ibm-watson-health-failure-artificial-intelligence.html

Orwell, G. (2021a). *Nineteen eighty-four*. Hachette UK.

Orwell, G. (2021b). *Animal farm*. Oxford University Press.

Painter, C. (1975). William Shockley [Photograph]. *Stanford News Service*. https://commons.wikimedia.org/wiki/File:William_Shockley,_Stanford_University.jpg

Popenici, S. (2023a). *Artificial intelligence and learning futures: Critical narratives of technology and imagination in higher education*. Routledge.

Popenici, S. (2023b). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning & Teaching, 6*(2). https://doi.org/10.37074/jalt.2023.6.2.4

Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12*(1), 1-13.

Relief Bruno Campo dei Fiori n1. (n.d.). In *Wikipedia*. https://en.wikipedia.org/wiki/Giordano_Bruno#/media/File:Relief_Bruno_Campo_dei_Fiori_n1.jpg

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9.

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Siemens, G., Marmolejo-Ramos, F., Gabriel, F., Medeiros, K., Marrone, R., Joksimovic, S., & de Laat, M. (2022). Human and artificial cognition. *Computers and Education: Artificial Intelligence, 3*, 100107.

Smee, B. (2023, March 7). 'No actual teaching': alarm bells over online courses outsourced by Australian universities. *The Guardian*. https://www.theguardian.com/australia-news/2023/mar/07/no-actual-teaching-alarm-bells-over-online-courses-outsourced-by-australian-universities

Susskind, D. (2021). *A world without work*. Metropolitan Books

Swafford, J. (2014). *Beethoven: Anguish and triumph*. Faber & Faber.

Tech Transparency Project. (2017, July 11). Google Academics Inc. *Tech Transparency Project*. https://www.techtransparencyproject.org/articles/google-academics-inc

Traykov, S. (2008). Michelangelo's Pietà in St. Peter's Basilica in the Vatican [Image]. *Wikipedia*. https://en.wikipedia.org/wiki/Piet%C3%A0_(Michelangelo)#/media/File:Pieta_de_Michelangelo_-_Vaticano.jpg

Waldmüller, F. G. (1823). Beethoven [Painting]. *Wikimedia Commons,* https://commons.wikimedia.org/wiki/File:Beethoven_Waldmuller_1823.jpg

Warren, A. (1969). *James Baldwin* [Photograph], https://commons.wikimedia.org/wiki/File:James_Baldwin_37_Allan_Warren_(cropped).jpg

Weatherbed, J. (2023, June 29). Meta reveals plans for increased transparency in Facebook and Instagram algorithms. *The Verge,* https://www.theverge.com/2023/6/29/23778068/meta-facebook-instagram-social-media-algorithms-ai-transparency

Whitman, J. Q. (2017). *Hitler's American model. The United States and the making of Nazi race law.* Princeton University Press.

Wikimedia Commons. (n.d.). *Photograph of Karl Pearson and Sir Francis Galton,* https://commons.wikimedia.org/wiki/File:Karl_Pearson;_Sir_Francis_Galton.jpg

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Chatbots and AI in Education (AIEd) tools: The good, the bad, and the ugly

Augustine Osamor Ifelebuegu[A]

Peace Kulume[B]

Perpetua Cherukut[C]

A Institute of Education and Life-long Learning, Victoria University, Kampala; Faculty of Science and Technology, Victoria University Kampala, Uganda

B Faculty of Science and Technology, Victoria University Kampala, Uganda

C Institute of Education and Life-long Learning, Victoria University, Kampala, Uganda

DOI: https://doi.org/10.37074/jalt.2023.6.2.29

Abstract

As the application of Artificial Intelligence (AI) continues to permeate various sectors, the educational landscape is no exception. Several AI in education (AIEd) applications, like chatbots, present an intriguing array of opportunities and challenges. This paper provides an in-depth exploration of the use and role of AI in education and research, focusing on the benefits (the good) and potential pitfalls (the bad and ugly) associated with the deployment of chatbots and other AIEDs. The opportunities explored include personalised learning, facilitation of administrative tasks, enriched research capabilities, and the provision of a platform for collaboration. These advantages are balanced against potential downsides, such as job displacement, misinformation, plagiarism, and the erosion of human connection. Ethical considerations, particularly concerning data privacy, bias reinforcement, and the digital divide, are also examined. Conclusions drawn from this analysis stress the importance of striking a balance between AI capabilities and human elements in education, as well as developing comprehensive ethical frameworks for AI deployment in educational contexts.

Keywords: Artificial Intelligence; AI in Education (AIEd); chatbots; data security; education; ethics; personalised learning; privacy; research.

Introduction

During the COVID-19 pandemic, the use of technology in education became a lifeline for many institutions across the globe. Since then, there has been a significant growth in the use and application of technology in education. Rapidly expanding today is a new era of education and research aided by artificial intelligence chatbots following the launch of OpenAI's ChatGPT in November 2022. They are infiltrating the education sector, reshaping traditional approaches to teaching and learning. One of AI's potent implementations in this domain is the use of AI chatbots employed to interact with users, deliver information, clarify doubts, and facilitate a host of other learning-related tasks (Adiguzel et al., 2023; Ifelebuegu, 2023; Pinzolits, 2024).

Before delving into the specifics of AI and its educational implications, it is crucial to understand what chatbots are and how they work. A chatbot is an artificial intelligence program and a human-computer interaction model (Adamopoulou & Moussiades, 2020; Bansal & Khan, 2018). In simpler language, chatbots are software that imitates human-like communication with the user in a chat. They are built upon the transformer-based language model architecture. They can generate a text that closely mirrors human language by predicting the likelihood of a word following a given set of words. They are trained on a diverse range of internet text, allowing them to exhibit broad, context-aware capabilities. Advanced natural language processing algorithms power them and they have the capacity to simulate human-like interactions, making them a valuable tool for various educational and research endeavors. For instance, in education, they can be used to facilitate personalised learning, provide immediate feedback, and manage administrative duties, among other things (Chen et al., 2023; Ifelebuegu, 2023; Popenici, 2023). They can assist with data acquisition, participant interaction, and more in research. Several chatbots are currently employed for literature search, review, content analysis, scientific writing, and revision.

Despite the optimism, the use of chatbots engenders a number of concerns. Potential challenges include issues of privacy and data security, ethical dilemmas, and the risk of over-reliance on technology. Therefore, it is essential to strike a balance between leveraging the potential of this technology and mitigating its drawbacks (Crawford et al., 2023). This study will investigate the good, the bad and the ugly implications of AI and chatbots in education and also highlight the future perspectives. The study will employ an exploratory research strategy and qualitative research methodology to acquire data for analysis and expert interpretation. The research findings will provide insights, perspectives into the applications of AI chatbots in teaching, learning and research, as well as potential developments and ethical challenges.

Methods

This study is characterised as exploratory and interpretive in nature, aligning with the philosophy of interpretivism. This philosophical approach posits that human behaviour and meaning are products of social construction and subjectivity. It asserts that knowledge and understanding can only be achieved through interpretation and the creation of meaning (Creswell, 2014; Kooli, 2023). In this research method, the investigator strives to comprehend the subjective experiences and viewpoints surrounding the observed phenomena, along with their associated meanings and interpretations. The central focus lies in the social construction of reality, with the researcher aiming to grasp the world through the lens of the phenomena observed (Creswell, 2014).

The study relies exclusively on qualitative data. A thorough analysis of collected secondary and qualitative data is conducted using a thematic analytical framework. This framework is employed to derive themes that align with the study's objectives and inquiries. Given the continuous advancements in artificial intelligence tools and chatbots, this study aims to provide an in-depth exploration of the multiple applications of AI chatbots in education while also addressing the challenges and emerging ethical dilemmas. As such, the primary research approach adopted is exploratory, geared towards investigating the novel phenomenon of using chatbots and other AI educational tools in both teaching, learning and research contexts. The outcomes of this research endeavor are expected to offer valuable insights, prospective viewpoints, and potential developments, all while contributing to the ongoing discourse surrounding the ethical dilemma associated with the transformation of education and research through AI systems and chatbots.

The research procedure followed three primary steps, as previously outlined by Kuhail et al. (2023), which include (1) establishing the review protocol, encompassing the research queries, the approach for addressing them, the search strategy, and the criteria for inclusion and exclusion; (2) executing the study by hand-picking relevant articles, evaluating their quality, and amalgamating the outcomes; and (3) presenting the discovered insights.

Research questions

To address the research objectives, four research questions were developed. The first research question (RQ1) investigates the various existing and emerging AI tools and chatbots used in teaching, learning and research. The second research question (RQ2) explores the beneficial applications of chatbots in education and research (the good). The third research question (RQ3) investigates the negative impacts of AI use in teaching and research (the bad). The fourth research question (RQ4) evaluates the potential ethical challenges associated with the use of AI and chatbots in education and research (the ugly).

Literature search strategy and data collections

An exhaustive and methodical examination of electronic databases was performed. Google Scholar, ScienceDirect and Scopus were selected due to their comprehensive coverage of educational and AI-related literature (Pinzolits, 2024; Tlili et al., 2022, 2023b). Moreover, specialised scientific journals in the field of pedagogy and artificial intelligence were also reviewed. These include the Journal of Applied Learning and Teaching, International Journal of Artificial Intelligence in Education, Education and Information Technologies, Computers & Education, Artificial Intelligence Review, Smart Learning Environments, etc. These were chosen due to their focus on AI and education or special editions with a focus on AI tools and pedagogy. The search time ranges for the articles is from 2018 to 2023. Also, to ensure state-of-the-art information on AI and pedagogy, online sources and blogs were reviewed. Only articles published in English were included, and those related to AI tools and chatbots in education and research.

The keywords used during the literature search include "AIEd", "chatbots in education", "AI tools for education", "ChatGPT in education", "AI-assisted learning", "personalised learning" and "learning technologies". The listed keywords were used individually and in combination during the search to help maximise the spread and depth of the results (Pinzolits, 2024).

Limitations

In this paper, the authors' expert knowledge in education technologies and pedagogy enters into play, thereby potentially influencing their perspectives. The methodology employed was intended to ensure the most exhaustive and objective review possible while recognising the inherent limitations. Some relevant studies may have been missed despite the exhaustive search strategy, particularly those published in languages other than English or in less accessible databases.

Results and discussion

AI in Education (AIEd) tools

The first research question (RQ1) examines the various AI tools and chatbots used in teaching, learning, and research, both existing and emergent. AI tools and chatbots are already changing the way teaching and research are done in higher education (Calonge et al., 2023; Ifelebuegu, 2023; Pinzolits, 2024). AIEd applications like chatbots have the potential to improve the teaching, learning and research processes in higher education (Kooli, 2023). For example, chatbots can assist learners, teachers, researchers and administrators alike to enhance their efficiency and effectiveness (Kooli 2023; Kuhail et al., 2023; Pinzolits 2024). Baker and Smith (2019) classified the various AIEds into three categories, namely learner-facing (LF), teacher-facing (TF) and system-facing (SF) tools. Based on the literature review, the authors propose a fourth category called research-facing (RF) AIEDs. Table 1 summarises some of the general, as well as institution-specific chatbots and AIEd tools that find applications in various educational activities.

Table 1. Commonly used chatbots and other AIEds tools.

Generally Available AI Tools and Chatbots		
AI Tools and Chatbots	**Use Case and classification**	**Link/Organisation**
AI-Writer	Generates articles from a simple headline RF, LF, TF	https://ai-writer.com/
Bard	Bard is an AI powered tool that allows creation and collaboration of ideas, boost productivity and bring your ideas to life. Its potential ranges from creating/responding to emails/ interview questions, designing briefs to planning. LF, TF, RF, SF	https://bard.google.com/
Bing Chat	An AI-driven chatbot for seamless communication allowing users to issue requests and receive actionable insights and provide guide to planning. LF, TF, RF, SF	https://www.bing.com/?/ai
Botsify	Allows educators to create their own chatbots for teaching. Teachers can design chatbots for their courses, enabling personalised learning experiences for students. LF, TF, RF, SF	https://botsify.com/
ChatGPT	A game-changing AI-powered tool for several applications including personalised learning, literature review, study support and writing LF, TF, RF, SF	https://chat.openai.com/ for ChatGPT 3.5 and various versions of GPT-4
Chat PDF	Assists researchers in reading and comprehending complex academic papers. Chat PDF employs artificial intelligence to provide a conversational interface, allowing researchers to pose queries about the paper and receive real-time responses. A chat-based interface facilitates the retrieval of information and answers from PDF files. LF, TF, RF, SF	https://www.chatpdf.com/
Claude	Offers vast data processing capabilities, creative writing and assistance in code generation and debugging. LF, TF, RF, SF	https://claude.ai/login
Consensus	Consensus is an AI-powered search engine that also provides answers supported by scientific evidence. It employs AI to extract key findings from peer-reviewed sources and present them in an easily consumable format. RF, LF, TF	https://consensus.app/
Copy.ai	Generates creative writing samples used as examples of various writing styles. LF, TF, RF, SF	https://www.copy.ai/
Duolingo	Web-based and mobile AI-powered platform that provides a gamified approach to language learning, making it engaging and accessible for users of all ages. LF	https://www.duolingo.com/
Ebbot by Learnify	Ebbot helps with administrative tasks in educational institutions. It can answer common questions about course details, schedules, and other relevant information SF	Learnify https://ebbot.com/
Elicit	Elicit is a freely available AI-powered research assistant that employs language models to locate pertinent academic papers, even in the absence of an exact keyword match. In addition, it can summarise and extract key information, assist with a variety of research duties, and integrate with citation managers. LF, RF	https://elicit.org/
Ernie	Integrates factual information from various sources into pre-trained language models. Allows dialogues, content creation, as well as ability to reason with knowledge and generate multiple modes of output. LF, TF, RF, SF	Baidu https://yiyan.baidu.com/welcome
Essay Writer	Guides students in structuring their essays logically and suggests areas for improvement. RF, LF, TF	https://www.the-good-ai.com/ https://collegeessay.org/ai-essaywriter
Jasper AI	Tool to write and generate images using AI RF, LF, TF, SF	https://www.jasper.ai/art
Jenni AI	Assists students in brainstorming and organizing thoughts prior to writing. LF, TF, RF	https://jenni.ai/
Laser AI	Laser AI is an application that streamlines systematic reviews, specifically Living Systematic Reviews. Its semiautomated data extraction module reduces extraction time without sacrificing quality. RF, LF	https://laser.ai/
Lateral AI	Lateral is an app propelled by artificial intelligence that offers text search, organisation of findings, simple sharing, and document viewing. LF, RF	Lateralio https://ai-finder.net/ai/lateralai
Ludwig	This chatbot helps students and writers with the English language. Users can input sentences to get contextual examples from reliable sources, aiding in understanding and usage. LF, TF, RF, SF	Linux Foundation AI & Data https://ludwig.ai/latest/
PaperPal	An AI tool that assists in the editing of academic texts in line with technical and language quality standards. It uses AI to ensure clarity, coherence, and adherence to academic writing standards. Performs grammar checks and provides real-time corrections. LF, TF, RF, SF	https://paperpal.com/
Plaito	By leveraging AI, Plaito brings the benefits of one-on-one tutoring providing understanding, confidence, clarity, and empowerment to students. LF, RF	https://www.plaito.ai/
Querium	An AI-powered tutoring platform that provides students with personalised learning experiences. It uses machine learning algorithms to assess students' strengths and weaknesses, providing targeted instruction and real-time feedback. LF, TF	https://www.querium.com/
Quillbot	A tool to improve writing, check for grammatical errors, assist with citations and paraphrase previously published text. LF, TF, RF, SF	https://quillbot.com/
Research Rabbit	Research Rabbit is a freely available digital platform that streamlines the literature search and administration for researchers. It integrates with citation managers and provides personalised recommendations, interactive visualisations, and collaboration options. RF, SF	https://www.researchrabbit.ai/
Scholarcy	Scholarcy is an AI application that summarises scholarly content, extracts structured data and knowledge summaries, and reduces the amount of time needed to extract essential information from an article. RF, LF, TF	https://www.scholarcy.com/
SciSpace	AI tool that can be used to read and interrogate a research paper LF, RF, TF	https://scispace.com/
Wisio	A platform propelled by AI that streamlines the scientific writing process. It provides customised text recommendations, citation extraction, translation, and English correction tools. RF, LF, TF	https://www.wisio.app/

Wordtune	Aids students in improving sentence structure and in expressing their ideas more clearly RF, LF, TF	https://www.wordtune.com/
Writely AI	AI-powered writing assistant that helps students enhance their writing skills and produce high-quality content. It utilises natural language processing and machine learning algorithms to provide real-time suggestions for improving grammar, style, and clarity. LF, TF, RF	https://www.writelyai.com/
Examples of organisation-specific chatbots		
ALEKS (Assessment and Learning in Knowledge Spaces)	Provides personalised learning experiences for students.	Carnegie Mellon University https://www.aleks.com/
Deakin University's Genie:	Voice-activated tool helping students with various aspects of university life, from finding resources for assignments to locating lecture halls.	Deakin University Australia
Pounce Chatbot	Georgia State University developed Pounce, an AI-powered chatbot designed to assist students during the enrollment process.	Georgia State University

The chatbots and AIEds represented in Table 1 all find applications in various areas of education.

Literature search and article analysis

A literature review is a crucial component of any research endeavour, but it can be a daunting, time- and energy-intensive undertaking. Researchers utilise diverse databases such as Google Scholar and Scopus to locate, peruse, and evaluate hundreds or thousands of research papers, synthesise existing knowledge, and determine knowledge gaps and research opportunities. Today, AI tools such as Research Rabbit, Scholarcy, SciSpace Lateral, Chat PDF, Elicit, WritelyAI, and Consensus can assist researchers in locating, organising, summarising, analysing, and writing research papers. By summarising a research paper's main points, Chat PDF, for example, can save the researcher a substantial amount of time. ChatGPT, Bard, and other comparable applications can also assist with literature searches and research questions.

Academic writing

Academic writing can be a daunting and tiring process. Reports and papers can take a long time to complete, and achieving excellence requires a substantial amount of effort. However, AI tools can assist students, instructors, and researchers with academic writing. Tools such as Jasper, QuillBot, PaperPal, Worldtune, Jenni, Copy.ai, and AI Writer (see Table 1) can facilitate the writing and editing of texts significantly. Chatbots such as ChatGPT, Claude, Bard, etc., are also effective writing and revision tools.

Administrative tasks

By automating administrative duties such as addressing frequently asked questions, scheduling appointments, and sending out notifications, an AI tool can aid in streamlining day-to-day business operations. The Georgia University Pounce Chatbot and the Deakin University Genie are examples of AIEd that perform administrative duties by responding to student inquiries.

The good

The second research question (RQ2) investigates the beneficial applications of chatbots in education and research (the good). The 'good' part of the paper outlines the advantages that AI chatbots bring to education and research.

AI chatbots hold great promise as an educational tool, particularly in the field of higher education. Recent studies and research have explored the opportunities and challenges of integrating AI chatbots into educational settings (Calonge et al., 2023; Holmes & Tuomi, 2023; Mohammadkarimi, 2023; Nemorin et al., 2023). AI chatbots hold extraordinary potential as educational tools, particularly in their ability to serve as virtual tutors. ChatGPT, for example, has been the subject of several recent studies exploring its potential use in education (Lo, 2023; Grassini, 2023). A study conducted by Kasneci et al. (2023) examined the benefits and risks of implementing ChatGPT as a virtual tutor in education. Willems (2023) also discussed the wider ethical implications of using large language models like ChatGPT in universities. The potential benefits of integrating Chatbots into education are significant (Tlili et al., 2023a; Adiguzel et al., 2023).

The majority of AI-powered chatbots in education assist students and instructors with a variety of tasks, including answering inquiries, providing study materials, monitoring progress, and more. These chatbots can be integrated into educational platforms and learning management systems for student support, instruction, and overall learning enhancement. Using natural language processing and machine learning, these educational chatbots can converse with students and provide individualised, real-time assistance. The sections that follow examine specific applications where they are particularly useful.

Personalised learning

AI chatbots can provide individualised instruction and feedback based on the needs and progress of each student. They are, therefore, revolutionising the concept of personalised learning, a methodology designed to cater to each student's unique needs, abilities, and learning preferences (Ifelebuegu, 2023). ChatGPT's and other AI chatbots' capability to process natural language input and generate human-like text make them a potent tool for individualised education. They can engage in dialogues, answer a wide range of questions, and provide detailed explanations, thereby acting as a virtual tutor available 24/7 for students (Hew et al., 2023; Limna et al., 2023). This immediate accessibility empowers learners to clarify doubts or understand complex concepts at their own pace without feeling time-pressured as they might in traditional classroom settings.

A key feature of personalised learning is the ability to adjust the level of difficulty based on the learner's proficiency. Here, chatbots shine by gauging the complexity of the questions posed by students and tailoring their responses accordingly. The Botsify chatbots, Plaito and Querium (see Table 1) are examples of applications for personalised learning. In language learning, for example, they can provide simpler explanations for beginners while offering more complex answers for advanced learners (e.g. Duolingo app for language learning). Chatbots can also generate practice problems or quizzes, enabling students to test their understanding and receive immediate feedback. Such interactive learning stimulates engagement and promotes active learning, which research suggests can significantly enhance knowledge retention (Gill et al., 2024; Rudolph et al., 2023).

Furthermore, the tool can adapt to each learner's interests, making the learning experience more engaging and relevant. For example, if a student is interested in space exploration, chatbots can contextualise mathematical problems or scientific explanations within that theme, thereby fostering a deeper connection to the material.

In summary, chatbots, as a tool for personalised learning, hold great promise. They combine the power of AI with the principles of individualised education to provide a flexible, accessible, and engaging learning experience. However, it is essential to remember that it should supplement, not replace, the invaluable role of human educators in nurturing curiosity, empathy, and critical thinking in students.

Homework assistance

The increasing sophistication of AI chatbots presents a valuable opportunity for enhancing the educational experience. One emerging role of AI in education is serving as a homework assistant, providing round-the-clock support to students (Chan, 2023; Limna et al., 2023). Chatbots can assist students in a variety of subjects, offering explanations, facilitating problem-solving, and giving examples for clearer understanding (Radziwill & Benton, 2017). In mathematics, for instance, it can guide students through problem-solving processes, demystifying complex principles. GPT4 (see Table 1) helps answer questions for children and assist them with homework. For language-based assignments, it can suggest essay ideas or help refine grammar and sentence structure. One of AI chabots' most significant advantages is their availability. Unlike human tutors, who are bound by time constraints, chatbots can be accessed at any time, making it a flexible tool for students who study outside conventional hours or balance multiple responsibilities (Ifelebuegu, 2023; Ray, 2023).

Furthermore, AI chatbots can foster self-directed learning. They encourage students to find solutions themselves, nurturing critical thinking skills and promoting independent problem-solving (Bruff et al., 2013). It also allows students to revisit concepts or problems without fear of judgement, promoting a conducive and stress-free learning environment (D'Mello, 2016).

However, it is essential to note that while AI chatbots are a powerful tool, they are not without flaws. They can occasionally provide incorrect or misleading information, which underscores the importance of supplementing it with other reliable educational resources. Additionally, the potential misuse of AI tools, such as using them to generate complete essays, must be addressed to ensure academic integrity (Ifelebuegu, 2023).

In conclusion, when used responsibly, chatbots can serve as an effective homework assistant, supplementing traditional learning methods with a flexible, supportive, and personalised approach, and students must understand that the tool should be used for assistance and not as a means to do their work for them.

Answering queries

AI chatbots are versatile tools capable of answering a wide range of queries, whether they pertain to academic topics, technical issues, or general information. As a large language model trained on extensive data, they have the ability to understand and process natural language queries, thereby providing responses that closely mimic human-like conversation. In educational settings, chatbots serve as a valuable resource for learners, educators, and researchers alike. Students can use them as a virtual tutor to ask questions about complex subjects or clarify concepts that they might find challenging (Kasneci et al., 2023; Kuhail et al., 2023). For educators, chatbots can answer queries related to lesson planning, curriculum development, or even pedagogical strategies, offering insights drawn from a broad knowledge base. Researchers can employ them to understand complex topics, navigate scholarly literature, or brainstorm ideas (Olujimi & Ade-Ibijolla, 2023)

In addition to their utility in academic contexts, chatbots also serve a broader purpose as information assistants. They can provide information on a vast array of topics, from answering trivia questions to explaining current events or scientific phenomena. This makes them handy tools for anyone seeking immediate, convenient access to information. Chatbots' potential as a tool for answering queries is vast, offering a highly accessible, interactive, and broad-ranging resource for information seekers in various contexts.

Content generation

AI chatbots' ability to generate content is multi-faceted and versatile. They can be harnessed in the educational domain for the creation of learning materials such as worksheets, quizzes, and even lesson plans. This can be particularly beneficial for educators, saving them time and allowing them to focus on their core teaching activities (Huang & Liang, 2021; Lo, 2023). In the realm of research, they can be used to generate summaries of complex papers, abstracts, or literature reviews. This function can support researchers by simplifying the process of digesting extensive amounts of information (Hill-Yardin et al., 2023; Sarrison, 2023).

Furthermore, in content marketing or blogging, chatbots like ChatGPT, Bard and Ernie can help generate ideas, draft articles, and even suggest edits. It can be an excellent tool for brainstorming and outlining, which can speed up the content creation process (De Bruyn, 2020). However, while a chatbot can be a valuable tool for content generation, it is essential to remember that its output needs careful supervision. It is not entirely flawless, and its output should be reviewed for accuracy, coherence, and appropriateness.

Language learning

Primarily, chatbots can serve as language practice partners, available round the clock. Learners can engage in conversation with the AI, practising their speaking and comprehension skills in a low-pressure environment. They offer instant feedback, allowing learners to correct mistakes and enhance their language proficiency over time (Huang et al., 2022; Jeon et al., 2023). Duolingo (see Table 1) is a classic example of a commonly used language learning AIEd tool.
Additionally, chatbots and other AIEd tools can provide grammatical explanations and context-based examples, assisting learners in understanding intricate language rules. they can also help learners expand their vocabulary by introducing new words and phrases within a conversational context, making them easier to remember.

Moreover, AIEds like Duolingo and ChatGPT's capacity for multi-language dialogue makes them useful tools for learners of diverse languages. Their ability to provide translations can be particularly beneficial for beginners who are developing their foundational vocabulary and comprehension skills (Belda-Medina & Calvo-Ferer, 2022; Jeon et al., 2023). However, while chatbots offer numerous advantages, it is crucial to use them responsibly and in conjunction with other language learning methods. It is also necessary to keep in mind that they represent AI and are not human, and may not perfectly capture nuances of cultural context or colloquial language use.

Administrative tasks

Educational institutions can use chatbots to automate various administrative tasks, such as answering frequently asked questions, scheduling appointments, or sending out notifications. Georgia University's Pounce chatbot and Deakin University's Genie (see Table 1) are examples of currently used chatbots for institution-specific administrative tasks. Ebbot from Learnify also help educational institutions with basic to more complex administrative tasks.

These AIEd tools can automate many such tasks, leading to enhanced efficiency and productivity. For instance, they can be utilised for drafting emails or other forms of communication, utilising their natural language processing capabilities to construct coherent, professional, and contextually appropriate messages (Kooli, 2023; Ifelebuegu, 2023). In the realm of schedule management, chatbots can assist in setting up meetings, sending reminders, or creating task lists. Integrating it with calendar applications or project management tools can help keep track of important dates and deadlines, ensuring smooth administrative functioning. Furthermore, they can be valuable assets in customer service, capable of answering frequently asked questions or providing information about products or services. This can significantly reduce the response time and improve the customer service experience (Rasul et al., 2023). However, it is important to note that while chatbots can effectively handle many administrative tasks, human supervision is necessary to ensure accuracy and to handle tasks that require human judgment or decision-making.

Research

Researchers can use AIEd like chatbots to summarise articles, generate hypotheses, or even draft sections of a paper. However, due to potential inaccuracies, it is essential to thoroughly check and verify any outputs from the AI. The emergence of OpenAI's ChatGPT has opened up new possibilities for the use of AI in research. With its extensive training on diverse datasets, ChatGPT and other AIEd tools can offer a vast amount of information and ideas, making it a valuable tool for researchers in various disciplines (Kooli, 2023).

One of the most significant ways chatbots can assist in research is by aiding in literature review and content analysis. Researchers can use the model to summarise long texts, articles, or papers, saving considerable time and effort. They can also provide an overview of multiple articles, helping researchers grasp the broader narrative or trends in their research domain (Tlili et al., 2023a). Another area where chatbots can prove useful is generating ideas and brainstorming. Researchers can ask chatbots to generate ideas or hypotheses on a particular subject, leveraging its broad knowledge base to foster creativity and explore diverse perspectives.

Furthermore, some chatbots can be employed to draft research proposals or write preliminary versions of research papers. It can suggest structures, write introductions, or even create abstracts, which can then be further refined by the researchers (Davies, 2016). They can also aid in explaining complex concepts or methods, particularly in areas such as statistics or computational methods, making them more accessible for researchers who may not be experts in these areas.

However, while AIEds like chatbots offer substantial benefits, it is crucial for researchers to use them responsibly and not rely on them blindly. Outputs should always be cross-checked for accuracy, and key decisions, especially ethical ones, should be made by the researchers themselves.

Inclusive education

Chatbots can be used to create more inclusive learning environments (Chen et al., 2023; Han & Lee, 2022). For example, some chatbots can provide additional support for students with disabilities or learning difficulties or assist in translating materials for students who speak different languages. While AI applications like ChatGPT and Duolingo have numerous potential benefits in education and research, it is crucial to use them responsibly and in a way that complements rather than replaces human input. The role of educators remains paramount in providing students with a comprehensive, empathetic, and enriched learning experience. Similarly, in research, human expertise, critical thinking, and ethical judgment are essential and cannot be replaced by AI.

Collaboration

Chatbots can act as a platform for collaboration among students. In group projects, they can assist with brainstorming sessions by generating ideas or suggesting solutions based on the given input, facilitating a more dynamic and productive collaboration. In an educational setting, chatbots can facilitate collaborative learning by serving as a tool where students can interact, share ideas, and solve problems together. For instance, they can be used to pose questions or provide prompts that encourage group discussions, fostering critical thinking and problem-solving skills (Brynjolfsson & McAfee, 2014; Tlili et al., 2023a; Gill et al., 2024). They can also assist in collaborative projects by generating ideas, drafting collaborative documents, or managing tasks. They can serve as a neutral platform where ideas are pooled and refined, promoting a sense of equal contribution among team members (Davenport & Ronanki, 2018).

Furthermore, chatbots can enable international collaborations by bridging language gaps. With the multilingual capabilities of some of them, they can translate conversations in real time, making interactions seamless and inclusive.

Professional development for educators

Educators can utilise AIEd, such as chatbots, for their own learning and development (Molala & Mbaya, 2023). They can interact with the tool to deepen their knowledge in certain areas, generate ideas for class activities, or even assist in administrative tasks such as grading or scheduling. Continuing Professional Development (CPD) is crucial for educators to keep up with evolving educational trends and pedagogies. Chatbots can be a valuable tool in facilitating this process. Firstly, chatbots can provide personalized learning opportunities for educators. By interacting with the AI, teachers can learn at their own pace and focus on areas they wish to improve or explore, ranging from subject matter expertise to teaching strategies. Additionally, they can serve as a source of up-to-date information. Given their extensive training on a vast corpus of text, they can provide insights into recent educational trends, research, and pedagogical techniques, helping teachers stay abreast of the latest developments (Jama et al., 2023).

Additionally, chatbots can assist in the creation and evaluation of instructional materials. They can generate content for lesson plans, quizzes, or worksheets, and provide feedback on existing materials, making the process more efficient and effective (Holmes et al., 2022). However, while chatbots can be a useful tool for CPD, they should be used responsibly, with educators verifying information and applying professional judgment.

Interdisciplinary learning

With their ability to generate information across a wide array of topics, chatbots can promote interdisciplinary learning, allowing students to see connections between different fields of study (Zhu et al., 2023). They can provide relevant information from various disciplines, thereby fostering a holistic approach to learning. OpenAI's ChatGPT and many more chatbots present unique opportunities for promoting interdisciplinary learning, an approach that integrates knowledge and methods from different disciplines, providing students with a more comprehensive understanding of complex real-world problems. Primarily, chatbots, with their vast knowledge base across diverse subjects, can help link concepts from different fields, facilitating an integrative learning experience. Students can query the AI on how concepts from one discipline relate to another, fostering cross-disciplinary thinking. Chatbots can also aid in developing interdisciplinary learning materials; for instance, they can generate content that blends concepts from different subjects or provides interdisciplinary exercises or problem sets. In summary, Chatbots can play a key role in promoting interdisciplinary learning, fostering an integrated understanding of knowledge and nurturing students' ability to tackle complex, multifaceted problems.

Self-paced learning

Chatbots can support self-paced learning by allowing students to interact with it whenever they want. Its 24/7 availability makes learning more flexible and accessible, allowing students to learn at their own pace and on their own schedule. AI applications like ChatGPT, Bard, Ernie, etc., can significantly enhance self-paced learning, a learning model where learners control the speed and the sequence of their learning (Opara et al., 2023). Firstly, they can provide learners with 24/7 access to information and help. Students can interact with the AI at any time, asking questions or seeking explanations as needed, which enables learning at their own pace. Secondly, they offer personalised learning pathways. A chatbot can adapt its responses based on previous interactions, tailoring the information to the learner's existing knowledge and understanding. This can foster deeper and more effective learning (Kaplan & Haenlein, 2016). Additionally, chatbot capabilities extend beyond content delivery. They can generate practice problems, provide immediate feedback, or even guide reflective practices, reinforcing learning and supporting mastery of concepts (Kaplan & Haenlein, 2016).

Motivation and engagement

The novelty and interactive nature of chatbots like ChatGPT can boost student motivation and engagement. They can make learning more fun and interesting and thereby encourage students to participate more actively in their education (Deng & Yu, 2023; Kuhail et al., 2023). Firstly, the chatbots' interactive nature can enhance learners' engagement. They can provide immediate responses to queries, fostering an active learning environment where learners can explore concepts at their own pace and triggering intrinsic motivation. Secondly, AIEd applications can offer personalised learning experiences, tailoring their interactions to individual learner's needs and interests. This personalisation can enhance the relevance of learning content, fostering a deeper connection with the subject matter and increasing motivation (Kooli, 2023). Furthermore, ChatGPT can introduce elements of gamification, such as challenges or quizzes, into the learning process. These elements can create a fun and competitive environment, boosting learner engagement and motivation (Shim et al., 2023).

Feedback provision

Chatbots can be utilised to provide instant feedback on certain tasks, such as multiple-choice quizzes or language exercises. This immediate response can enhance the learning process by allowing students to quickly identify and correct their mistakes. Chatbots can provide instant feedback to learners, correct mistakes, explain correct answers, and offer strategies for improvement. The immediacy of this feedback is beneficial for learning, as it allows learners to adjust their understanding and strategies promptly (Baskara, 2023). Moreover, the AI can offer personalised feedback based on the learner's performance. By tailoring the feedback to the learner's needs and progress, chatbots can help improve individual learning outcomes. Also, feedback provided by AI tools and chatbots is non-judgmental. This objective nature can create a safe learning environment where learners feel comfortable making mistakes and taking risks, which can foster a growth mindset (Lo, 2023; Ray, 2023). However, while automated feedback from AI, like chatbots, can be valuable, it should be supplemented with human feedback. Educators can provide context-specific feedback and address socio-emotional aspects of learning that AI systems might miss (Holmes et al., 2022).

Mental health support

While not a replacement for professional help, chatbots like ChatGPT can be used as a first step in providing support for students who may be dealing with stress or anxiety. They can provide basic advice, suggest coping strategies, or simply act as a non-judgmental listener. However, it is critical to ensure that students are directed to appropriate professional resources for further assistance (Aminah et al., 2023; Farhat, 2023; Rathnayaka et al., 2022).

Chatbots can provide a safe and non-judgmental space for students to express their feelings and concerns. Since they can respond to queries and prompts with empathy and understanding, students might find it easier to open up and discuss their mental health challenges with it. Chatbots can offer resources, strategies, and advice for managing stress, anxiety, and other common mental health issues. They can also direct students to professional help when necessary.

Admission process support

The process of guiding students through the enrolment procedure is a crucial aspect of operating a successful educational institution. Since the process varies from institution to institution, the applicants find it tiresome. Everyone desires straightforward and speedy solutions, so assisting your students in obtaining these will increase conversions. AI-powered chatbots that can be trained and programmed to understand a prospect's admission lifecycle can greatly facilitate the admissions process (Fitria et al., 2023; Nguyen et al., 2021). Without involving a human, these algorithms administer an entrance exam, monitor student performance, shortlist those who qualify, inform them about the next steps and course options, and answer all of their queries. This expedites the procedure by eliminating the typical waiting time required by a human agent.

In conclusion, it is essential to emphasise that while chatbots can be a useful instrument in many fields of education and research, they cannot replace human interaction and discernment. The utilisation of AI tools must always be governed by responsible practices that prioritise the welfare and growth of students. In addition to the numerous benefits, it is essential that we comprehend the potential risks and ethical considerations as we continue to investigate the integration of AI in education.

The bad

The third research question (RQ3) investigate the negative impacts of AI use in teaching and research (the bad). The 'bad' part of the paper investigates the challenges and limitations associated with the use of AI chatbots in education and research. These include concerns related to privacy and data security, as chatbots often handle sensitive personal information. The paper also raises issues of equity and accessibility since the benefits of this technology may not be available to all due to digital divides. While chatbots and other AIEd can be valuable tools in education and research, they are not without potential downsides. The following are some areas where the use of chatbots and AI tools could be problematic.

Reinforcement of bias

Chatbots like ChatGPT are only as unbiased as the data they are trained on. If the training data contains biases, the model can potentially learn and reproduce these biases. In an educational context, this could result in students receiving biased or skewed information. Chatbots are usually trained on a vast corpus of internet text data. While this allows the AI to generate human-like text, it also presents a significant risk of bias reinforcement (Kooli, 2023; Talanquer, 2023).

Firstly, chatbot training data includes content from the internet; they are exposed to biases that exist in these texts. These biases can be based on race, gender, religion, and more. When generating responses, the AI may unknowingly perpetuate these biases, thus reinforcing them (Caliskan et al., 2023). Secondly, AI systems like chatbots are sensitive to the input they receive. If users interact with the AI using biased language or ideas, it could adapt to these biases, further propagating them in its responses (Zhao et al., 2020). While efforts have been made to "debias" AI, these measures can only reduce, not completely eliminate, bias. Moreover, the lack of transparency in AI algorithms, often referred to as the 'black box' problem, makes it difficult to understand and correct bias in AI responses (Ray, 2023).

The reinforcement of bias by AI has serious implications for education and research. It can lead to the perpetuation of stereotypes, misinformation, and discriminatory practices, undermining the goals of fairness, equality, and objectivity in these fields (Eubanks, 2018). Therefore, while chatbots can be valuable tools in education and research, it is crucial for users to be aware of their potential for bias reinforcement. Continuous monitoring and adjustment of the AI's responses, as well as educating users about AI bias, can help mitigate this risk.

Overreliance

With the emergence of AI in education, however, there are also concerns about its potential negative effects on students. Some experts contend that AI-based learning platforms could inhibit critical thinking and reduce human interaction, which is essential for learning (Mhlanga, 2023; Zanetti et al., 2019). The convenience of using chatbots and similar AIEd applications can lead to overreliance, which can be detrimental. Students may become overly dependent on these AI tools for learning, potentially limiting their problem-solving and critical thinking skills. Similarly, educators may be tempted to overuse AI tools for administrative tasks or content creation, leading to decreased personal interaction and human connection in education. When students lean heavily on AI for answers, they may not develop the necessary skills to independently analyse and solve problems.

Additionally, overreliance on AI could negatively impact the social aspects of learning. Education is not just about information transmission but also about human interaction, socialisation, and collaboration. If AI becomes the primary source of learning, these essential aspects could be undermined (Brynjolfsson & McAfee, 2014). While AI systems like chatbots provide valuable support for education and research, overreliance on these tools can pose challenges to critical thinking, social learning, and the cultivation of a fair and unbiased learning environment.

Misinformation

Misinformation is another significant concern in the use of AIED applications like chatbots in education and research. Chatbots rely on extensive data, including vast amounts of internet text, to generate responses. However, the internet is replete with misinformation, and this could potentially seep into the AI's outputs (Adetayo, 2023). For instance, if a chatbot encounters factually incorrect or misleading data during its training, it could inadvertently propagate these errors in its interactions with users. If learners or researchers rely on this information without cross-verifying, it could lead to the spread of misinformation.

Furthermore, the potential for chatbots to generate new content based on the patterns it has learned adds another layer to this problem. This feature, while innovative, could result in the creation and dissemination of unfounded information. To mitigate these risks, it is crucial to combine the use of AI with critical thinking and information literacy skills. Users must be encouraged to scrutinize AI-generated content and cross-reference information from multiple sources.

Plagiarism

The use of chatbots and other AIEd applications in education and research raises serious concerns about plagiarism. Although these tools can be instrumental in helping students understand complex concepts or aiding researchers in generating ideas, the ease with which they generate human-like text could potentially encourage plagiarism (Chaka, 2023; Ifelebuegu, 2023; Kleebayoon & Wiwanitkit, 2023; King, 2023). In an academic context, plagiarism is the presentation of someone else's work or ideas as one's own without proper attribution. If students use text generated by a chatbot such as ChatGPT in their assignments without proper citation, it could be considered plagiarism. Also, chatbots could facilitate 'contract cheating' where students submit AI-generated work as their own. This is a significant concern as it undermines the educational objectives of fostering original thinking and academic integrity (Mohammadkarimi, 2023).

On the flip side, AI can also be used to detect plagiarism, using algorithms to match text patterns and identify potential instances of academic dishonesty. However, it is imperative that educators reinforce the importance of academic honesty and proper citation practices in the digital age.

Data privacy

Data privacy is a critical concern in the deployment of AI-powered chatbots like ChatGPT in education and research. These systems often require access to extensive user data to personalise and enhance their services. However, the collection, storage, and use of these data bring about significant privacy implications. In the context of education, personal data might include students' grades, learning patterns, personal interests, and even their social interactions. When an AI tool is granted access to this information, there is an inherent risk of data breaches or misuse. If data is not adequately protected, sensitive information could be exposed, leading to significant harm to the individuals involved (Baskara, 2023; Yang et al., 2023). Moreover, the use of data in AI systems can raise questions about consent. It is essential that users, including students and educators, fully understand what data are being collected, how they are being used, and how they can control their data. Without informed consent, the use of AI in education could violate fundamental privacy rights (Gupta & Jain, 2023).

Depersonalisation of education

The increased use of AI in education could lead to a decrease in human interaction, which is a crucial part of learning. It could lead to less personal, more standardised education and reduce the development of social skills and emotional intelligence in students (Kuhail, 2023; Tlili et al., 2023a).

Accessibility and digital divide

Accessibility is an essential consideration in deploying AI systems like chatbots in education and research. These systems have the potential to democratise education by making high-quality learning resources accessible to all, irrespective of geographical location or socioeconomic status (Bostrom & Yudkowsky, 2014). However, despite this potential, disparities in access to technology can lead to a digital divide, where individuals without access to the necessary technology are left behind. This could exacerbate existing educational inequalities, creating a divide between those who can benefit from AI-enhanced learning and those who cannot.

Furthermore, the accessibility of AI systems for individuals with disabilities is another significant issue. AI tools must be designed to be accessible and inclusive, considering the needs of all potential users. This includes providing features such as text-to-speech for visually impaired individuals or predictive text for those with motor impairments (Khowaja et al., 2023). In conclusion, while AI, like ChatGPT, can be a powerful tool, these limitations need to be acknowledged and managed in educational and research contexts.

The ugly

The fourth research question (RQ4) evaluates the potential ethical challenges associated with the use of AI and chatbots in education and research (the ugly). The 'ugly' part delves into more severe concerns about the use of AI chatbots, focusing on the ethical dilemma of AI use in education and research. The potential for misuse of AI chatbots in education and research settings is examined. This section also highlights the need for comprehensive ethical guidelines and effective regulatory measures to ensure the responsible use of AI chatbots in education and research.

Ethical implications

The use of AI-powered chatbots raises important ethical questions. Issues around fairness, accountability, and transparency must be addressed. Moreover, there is a risk of AI systems perpetuating or even amplifying biases present in their training data, potentially leading to biased outcomes or discriminatory practices.

The integration of chatbots in education presents various ethical dilemmas, some of which have been covered in a previous section. First, there is the issue of data privacy. As students interact with chatbots, vast amounts of personal data, including learning behaviours and preferences, are collected. How these data are stored, used, and potentially shared poses significant privacy concerns. Unauthorised access or misuse could have repercussions on a student's academic and personal life.

Next, the authenticity of interactions comes into play. Relying on chatbots for educational feedback might deprive students of genuine human interaction, which is crucial for emotional and social development. It could also influence students' perceptions of what constitutes meaningful communication (Mohammadkarimi, 2023).

Additionally, the potential biases in AI and chatbots, stemming from their training data, can inadvertently reinforce stereotypes or skewed perspectives. In an educational context, this could lead to misinformation or narrow-minded thinking. Addressing these ethical dilemmas requires careful consideration, ensuring that while technology aids education, it does not compromise students' growth, security, or values.

Threats to jobs

Job displacement due to AI is a pressing concern in today's rapidly advancing technological landscape. As artificial intelligence systems become more sophisticated, they increasingly perform tasks previously handled by humans—often more efficiently and at a lower cost. Automation and AI can streamline repetitive tasks, analyse vast datasets with precision, and even perform complex operations in fields like finance, medicine, and manufacturing. While this offers substantial economic benefits and productivity gains, it simultaneously poses challenges for the workforce. Many low-skilled jobs, particularly those involving routine, repetitive tasks, are at the highest risk of being automated. This shift has the potential to exacerbate income inequalities, as those without the skills to navigate an AI-augmented job market might face unemployment or underemployment.

In education, roles such as administrative staff, learning support, and even some teaching roles could be affected. For instance, if an AI system like ChatGPT can handle student inquiries effectively, answer routine questions, and provide personalised learning resources, the need for some human roles might be reduced (Khogali & Mekid, 2023). However, it's crucial to note that AI is not likely to replace educators entirely. The role of a teacher involves far more than providing information as it includes fostering a positive learning environment, providing emotional support, and nurturing critical and creative thinking, tasks that AI is currently incapable of fully replicating (Chan & Tsi, 2023). Therefore, while AI may alter the landscape of jobs in education and research, it also provides opportunities for new roles and requires upskilling and reskilling for effective integration and usage in these sectors.

The erosion of human connection

The integration of AI systems like chatbots into education and research also raises concerns about the potential erosion of human connection. Education is inherently a social endeavour, and the human touch in teaching and learning processes plays a crucial role in fostering empathy, understanding, and collaboration (Kamalov & Gurrib, 2023). AI, while proficient at managing information and facilitating personalised learning, cannot replicate the nuances of human interaction, such as body language, tone of voice, or emotional understanding. The overreliance on AI systems might result in a diminished emphasis on these human elements of communication and social interaction in an educational setting (Abbas et al., 2023).

Moreover, the role of educators extends beyond imparting knowledge; they also provide emotional support and mentorship and foster a sense of community, aspects that an AI like ChatGPT currently cannot fully replicate. While AI systems can provide numerous benefits, they should be used as a supplement to human interaction, not a replacement. Balancing the use of AI tools with the need for human connection and interaction is key to ensuring a rich, engaging, and holistic educational experience.

Technical failures

Technical failures in the context of AI chatbots in education can pose significant disruptions to the learning process. Chatbots rely on complex software, servers, and often cloud-based infrastructures. When any component of this intricate system faces issues, the chatbot can malfunction or become entirely unresponsive. For instance, a server outage can render a chatbot inaccessible, depriving students of essential learning resources at crucial moments, like just before an exam or assignment deadline. Moreover, bugs in the chatbot's code or issues in its underlying AI algorithms can lead to the dissemination of incorrect or misleading information, which, if undetected, can impede accurate learning. These technical failures not only hinder academic progress but can also erode trust in digital learning tools. Students and educators might become reluctant to rely on such tools, fearing unpredictability. In worst-case scenarios, consistent technical issues can exacerbate educational inequities, especially if alternative resources are not readily available to all students.

The potential 'ugliness' of using AI in education highlights the need for caution, regulation, and continuous evaluation. It is crucial that the deployment of such technologies is guided by ethical principles and a deep understanding of the possible long-term implications. While AI tools like ChatGPT offer many exciting opportunities for education, it is essential to ensure that their use enhances rather than detracts from the core values of education.

Conclusion and future perspectives

In conclusion, artificial intelligence, exemplified by chatbots and other AIEDs, is experiencing an unprecedented surge in its incorporation into the educational and research sectors. Multiple dimensions of the education sector are being transformed by these tools. These advancements offer numerous benefits to students, educators, and researchers alike. They personalise the educational experience by democratising access to immense stores of information, expediting administrative duties, and opening the door to novel and more thorough research methodologies. These beneficial changes—the good—are transforming pedagogical strategies and research paradigms.

On the other hand, there are concerns that require critical consideration. As AI becomes more pervasive, concerns regarding job displacement grow. The potential exists for these tools to inadvertently spread false information or reduce education to rote memorisation based on AI responses. In addition, an excessive reliance on digital tools may diminish the irreplaceable value of human connection and mentoring in the learning process, thereby introducing the bad and the ugly. Moreover, when the ethical dimension is considered, the challenges of data privacy become paramount. With AI systems trained on enormous datasets, the unintentional reinforcement of societal biases in educational tools becomes an urgent concern.

As we chart the course for the future of education, it is crucial that we strike a balance between the potential benefits of AI and its potential drawbacks by considering how we can leverage AI's potential while mitigating the associated risks. A blend of human and AI collaboration may present the best path forward, combining the benefits of AI with the creativity, empathy, and contextual understanding that characterises human interaction. Critical to the responsible use of AI is ongoing research and dialogue, not only among technologists but also among educators, policymakers, and learners. Comprehensive frameworks for the ethical use of AI in education, clear policies on data privacy, ongoing professional development for educators, and digital literacy education for learners are all essential components of responsible AI use.

References

Abbas, N., Ali, I., Manzoor, R., Hussain, T., & Hussaini, M. H. A. (2023). Role of Artificial Intelligence tools in enhancing students' educational performance at higher levels. *Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 3*(5), 36-49. https://doi.org/10.55529/jaimlnn.35.36.49

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications, 2,* 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Adetayo, A. J. (2023). Artificial intelligence chatbots in academic libraries: the rise of ChatGPT. *Library Hi Tech News, 40*(3), 18-21. https://doi.org/10.1108/LHTN-01-2023-0007

Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology, 15*(3), 429. https://doi.org/10.30935/cedtech/13152

Aminah, S., Hidayah, N., & Ramli, M. (2023). Considering ChatGPT to be the first aid for young adults on mental health issues. *Journal of Public Health, 45*(3) 615 - 616. https://doi.org/10.1093/pubmed/fdad065

Baker, T., Smith, N. A., Sheehan, K., Ward, K., Waters, A., Berditchevskaia,A., Van Den Berg, C., Campbell, N., Candsell, O., Casasbuenas, J., Cinnamon, J., Copeland, E., Duffy, E., Hannon, C., John, J., Grant, J., Klinger, J., Latham, M., Macken, C., … Ward-Dyer, G. (2019). *Educ-altion rebooted? Exploring the future of artificial intelligence in schools and colleges*. www.nesta.org.uk

Bansal, H., & Khan, R. (2018). A review paper on human computer interaction. *International Journal of Advanced Research in Computer Science, 8*(4), 53.

Baskara, F. R. (2023). Chatbots and flipped learning: Enhancing student engagement and learning outcomes through personalised support and collaboration. *IJORER: International Journal of Recent Educational Research, 4*(2), 223-238. https://doi.org/10.46245/ijorer.v4i2.331

Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences, 12*(17), 8427. https://doi.org/10.3390/app12178427

Bostrom, N., & Yudkowsky, E. (2014). The ethics of Artificial Intelligence. In K. Frankish & W. M. Ramsey (Eds.), *Cambridge handbook of Artificial Intelligence.* https://doi.org/10.1017/CBO9781139046855.020

Bruff, D. O., Fisher, D. H., McEwen, K. E., & Smith, B. E. (2013). Wrapping a MOOC: Student Perceptions of an Experiment in Blended Learning. *Journal of Online Learning and Teaching, 9*(2), 187.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183-186. DOI: 10.1126/science.aal4230

Calonge, D. S., Smail, L., & Kamalov, F. (2023). Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics. *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://doi.org/10.37074/jalt.2023.6.2.22

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://doi.org/10.37074/jalt.2023.6.2.12

Chan, C. K. Y., & Tsi, L. H. (2023). *The AI revolution in education: Will AI replace or assist teachers in higher education?* https://doi.org/10.48550/arXiv.2305.01185

Chan, W. (2023). AI chatbots and liberal education. *AI & Society: Knowledge, Culture and Communication*, 1-3. https://doi.org/10.1007/s00146-023-01736-8

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers, 25*(1), 161-182. https://doi.org/10.1007/s10796-022-10291-4

Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice, 20*(3), 02. https://doi.org/10.53761/1.20.3.02

Creswell, J. W. (2014). *A concise introduction to mixed methods research.* SAGE publications.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108-116

Davies, M. (2016). Automated essay scoring and the future of educational assessment in medical education. *Medical Education, 48*(10), 950-962. https://doi.org/10.1111/medu.12517

De Bruyn, A. (2020). *AI for marketing and product innovation: Powerful new tools for predicting trends. Connecting with customers, and closing sales.* Wiley.

Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability, 15*(4), 2940. https://doi.org/10.3390/su15042940

D'Mello, S. (2016). Emotional learning analytics and affective computing: Potential, principles, and practices. In C. Lang., G. Siemens., A. F. Wise., D. Gašević., & A. Merceron (Eds.), *Handbook of learning analytics* (pp. 120-129). Society for Learning Analytics Research.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

Farhat, F. (2023). ChatGPT as a complementary mental health resource: A boon or a bane. *Annals of Biomedical Engineering,* 1-4. https://doi.org/10.1007/s10439-023-03326-7

Fitria, T. N., Simbolon, N. E., & Afdaleni, A. (2023). Chatbots as online chat conversation in the education sector. *International Journal of Computer and Information System (IJCIS), 4*(3), 93-104. https://doi.org/10.29040/ijcis.v4i3.116

Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., ... & Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots. *Internet of Things and Cyber-Physical Systems, 4*, 19-23. https://doi.org/10.1016/j.iotcps.2023.06.002

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences, 13*(7), 692. https://doi.org/10.3390/educsci13070692

Gupta, D. G., & Jain, V. (2023). Use of Artificial Intelligence with ethics and privacy for personalized customer services.

In *Artificial Intelligence in customer service: The next frontier for personalized engagement* (pp. 231-257). Springer International Publishing.

Han, S., & Lee, M. K. (2022). FAQ chatbot and inclusive learning in massive open online courses. *Computers & Education, 179*, 104395. https://doi.org/10.1016/j.compedu.2021.104395

Hew, K. F., Huang, W., Du, J., & Jia, C. (2023). Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions. *Journal of Computing in Higher Education, 35*(1), 40-68. https://doi.org/10.1007/s12528-022-09338-x

Hill-Yardin, E. L., Hutchinson, M. R., Laycock, R., & Spencer, S. J. (2023). A Chat(GPT) about the future of scientific publishing. *Brain, Behavior, and Immunity, 110*, 152-154. https://doi.org/10.1016/j.bbi.2023.02.022

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education, 57*(4), 542-570. https://doi.org/10.1111/ejed.12533

Huang, C., & Liang, P. (2021). Write to rank: Deep reinforcement learning for automatic report generation. In *Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: Human language technologies*.

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237-257. https://doi.org/10.1111/jcal.12610

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://doi.org/10.37074/jalt.2023.6.2.2

Jamal, A., Solaiman, M., Alhasan, K., Temsah, M. H., & Sayed, G. (2023). Integrating ChatGPT in medical education: Adapting curricula to cultivate competent physicians for the AI era. *Cureus, 15*(8). https://doi.org/10.7759/cureus.43036.

Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, 104898. https://doi.org/10.1016/j.compedu.2023.104898

Kamalov, F., & Gurrib, I. (2023). *A new era of Artificial Intelligence in education: A multifaceted revolution.* https://doi.org/10.48550/arXiv.2305.18303

Kaplan, A. M., & Haenlein, M. (2016). Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. *Business Horizons, 59*(4), 441-450. https://doi.org/10.1016/j.bushor.2016.03.008.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khogali, H. O., & Mekid, S. (2023). The blended future of automation and AI: Examining some long-term societal and ethical impact features. *Technology in Society, 73*, 102232. https://doi.org/10.1016/j.techsoc.2023.102232

Khowaja, S. A., Khuwaja, P., & Dev, K. (2023). *ChatGPT Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) evaluation: A review.* arXiv preprint arXiv:2305.03123.

Kleebayoon, A., & Wiwanitkit, V. (2023). Artificial intelligence, chatbots, plagiarism and basic honesty: Comment. *Cellular and Molecular Bioengineering, 16*(2), 173-174. https://doi.org/10.1007/s12195-023-00759-x

Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability, 15*(7), 5614. https://doi.org/10.3390/su15075614

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies, 28*(1), 973-1018. https://doi.org/10.1007/s10639-022-11177-3

Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching, 6*(1), 65 – 74. https://doi.org/10.37074/jalt.2023.6.1.32

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Mhlanga, D. (2023). *Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning*. SSRN Electronic Journal. DOI: 10.2139/ssrn.4354422

Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://doi.org/10.37074/jalt.2023.6.2.10

Molala, T. S., & Mbaya, T. W. (2023). Social work and Artificial Intelligence: Towards the electronic social work field of specialisation. *International Journal of Social Science Research and Review, 6*(4), 613-621. https://doi.org/10.47814/ijssrr.v6i4.1206

Nemorin, S., Vlachidis, A., Ayerakwa, H. M., & Andriotis, P. (2023). AI hyped? A horizon scan of discourse on artificial intelligence in education (AIED) and development. *Learning, Media and Technology, 48*(1), 38-51. https://doi.org/10.1080/17439884.2022.2095568

Nguyen, M. T., Tran-Tien, M., Viet, A. P., Vu, H. T., & Nguyen, V. H. (2021, November). Building a chatbot for supporting the admission of universities. In *2021 13th International Conference on Knowledge and Systems Engineering (KSE)* (pp. 1-6). IEEE.

Olujimi, P. A., & Ade-Ibijola, A. (2023). NLP techniques for automating responses to customer queries: A systematic review. *Discover Artificial Intelligence, 3*(1), 20. https://doi.org/10.1007/s44163-023-00065-5

Opara, E., Mfon-Ette Theresa, A., & Aduke, T. C. (2023). ChatGPT for teaching, learning and research: Prospects and challenges. *Global Academic Journal of Humanities and Social Sciences, 5*. https://doi.org/10.36348/gajhss.2023.v05i02.001

Pinzolits, R. (2024). AI in academia: An overview of selected tools and their areas of application. *MAP Education and Humanities, 4,* 37-50. https://doi.org/10.53880/2744-2373.2023.4.37

Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://doi.org/10.37074/jalt.2023.6.2.4

Radziwill, N. M., & Benton, M. C. (2017). *Evaluating quality of chatbots and intelligent conversational agents.* https://doi.org/10.48550/arXiv.1704.04579

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41 – 56. https://doi.org/10.37074/jalt.2023.6.1.29

Rathnayaka, P., Mills, N., Burnett, D., De Silva, D., Alahakoon, D., & Gray, R. (2022). A mental health chatbot with cognitive skills for personalised behavioural activation and remote health monitoring. *Sensors, 22*(10), 3653. https://doi.org/10.3390/s22103653

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems,* (3), 121-154. https://doi.org/10.3390/s22103653

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364 – 389. https://doi.org/10.37074/jalt.2023.6.1.23

Sarrion, E. (2023). Using ChatGPT for Text Content Creation for Businesses. In *Exploring the power of ChatGPT: Applications, techniques, and implications* (pp. 99-106). Apress.

Shim, K. J., Menkhoff, T., Teo, L. Y. Q., & Ong, C. S. Q. (2023). Assessing the effectiveness of a chatbot workshop as experiential teaching and learning tool to engage undergraduate students. *Education and Information Technologies*, 1-24. https://doi.org/10.1007/s10639-023-11795-5

Talanquer, V. (2023). Interview with the Chatbot: How does it reason?. *Journal of Chemical Education, 100*(8), 2821-2824. https://doi.org/10.1021/acs.jchemed.3c00472

Tlili, A., Padilla-Zea, N., Garzón, J., Wang, Y., Kinshuk, K., & Burgos, D. (2022). The changing landscape of mobile learning pedagogy: A systematic literature review. *Interactive Learning Environments*, 1–18. doi:10.1080/10494820.2022.2039948

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023a). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Tlili, A., Huang, R., Mustafa, M. Y., Zhao, J., Bozkurt, A., Xu, L.,Wang, H., Salha, S., Altinay, F., Affouneh, S.,& Burgos, D. (2023b) Speaking of transparency: Are all Artificial Intelligence (AI) literature reviews in education transparent? *Journal of Applied Learning and Teaching, 6*(2). Advance online publication. https://doi.org/10.37074/jalt.2023.6.2.15
Willems, J. (2023). ChatGPT at universities–the least of our concerns. Available at SSRN 4334162.

Yang, J., Chen, Y. L., Por, L. Y., & Ku, C. S. (2023). A systematic literature review of information security in chatbots. *Applied Sciences, 13*(11), 6355. https://doi.org/10.3390/app13116355

Zanetti, M., Iseppi, G., & Cassese, F. P. (2019). A "psychopathic" Artificial Intelligence: The possible risks of a deviating AI in Education. *Research on Education and Media, 11*(1), 93-99. https://doi.org/10.2478/rem-2019-0013

Zhao, J., Zhou, Y., & Mao, K. (2020). Learning Gender-neutral Word Embeddings. EMNLP.

Zhu, G., Fan, X., Hou, C., Zhong, T., Seow, P., Shen-Hsing, A. C., ... & Poh, T. L. (2023). Embrace opportunities and face challenges: Using ChatGPT in undergraduate students' collaborative interdisciplinary learning. https://doi.org/10.48550/arXiv.2305.18616

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Enough of the chit-chat: A comparative analysis of four AI chatbots for calculus and statistics

David Santandreu Calonge[A]
Linda Smail[B]
Firuz Kamalov[C]

A Department of Academic Development, Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates

B Associate Professor, College of Interdisciplinary Studies, Zayed University, Dubai, United Arab Emirates

C Associate Professor, School of Engineering, Applied Science and Technology, Canadian University Dubai, United Arab Emirates

DOI: https://doi.org/10.37074/jalt.2023.6.2.22

Abstract

This article presents a comparative analysis of four AI chatbots with potential utilization in the fields of mathematics education and statistics, namely ChatGPT, GPT-4, Bard, and LLaMA. Our objective is to evaluate and compare the features, functionalities, and potential applications of these platforms within the domains of calculus and statistics. By examining their strengths and limitations, this study aims to provide insights into the selection and implementation of AI chatbots in calculus and statistics to enhance student learning. The results of the comparative analysis reveal that, while not perfect, GPT-4 outperforms ChatGPT, Bard, and LLaMA as a learning tool in calculus and statistics. Findings also reveal that chatbots may have a positive transformational impact on higher education.

Keywords: AI chatbots; Bard; calculus; ChatGPT; comparative analysis; GPT-4; Large Language Models (LLMs); LlaMA; statistics; student learning.

Introduction

Calculus and statistics are vital subjects that require effective teaching and learning methods to enhance students' engagement and comprehension. With the advancements in artificial intelligence (AI) and natural language processing (NLP), AI chatbots have emerged as promising tools for supporting students in higher education.

Kuhail et al. (2023) argued that chatbots provide a "cost-effective solution" (p. 2) to personalize learning activities, support educators, and "develop deep insight into learners' behaviour" (p. 1). AI chatbot platforms have gained significance in higher education (Singh Gill et al., 2023; Sok & Heng, 2023; Rudolph et al., 2023a; Tlili et al., 2023; Okonkwo & Ade-Ibijola, 2021; Hwang & Chang, 2021; Sandu & Gide, 2019). Moreover, the literature suggests that chatbots have the potential to enhance students' learning experience (s) in mathematics education (Castevecchi, 2023; Wardat et al., 2023) and statistics (Lee & Yeo, 2022), offering innovative solutions for learning, problem-solving, and concept clarification. They can provide personalized support, immediate feedback, interactive problem-solving, and adaptive instruction, fostering engagement and improving learning outcomes.

While there exist several studies that consider the performance of AI chatbots in mathematics problem solving, they are limited in two ways: (i) no notable analysis of Bard and LLaMA, and (ii) no analysis in statistics. This article fills the gap in the literature by evaluating and comparing four popular AI chatbot platforms, namely ChatGPT (GPT-3.5), GPT-4, Bard, and LLaMA 13-B, with a focus on their applicability and potential benefits in calculus and statistics. By examining their unique features and applications, this study aims to assist students (and educators) in selecting appropriate AI chatbot platforms to enhance their learning (and teaching) experience(s) in calculus and statistics.

Background

Benefits of using chatbots in higher education

There are several potential benefits to using chatbots in higher education (Kamalov et al., 2023). One of the main benefits is the ability to provide students with access to personalized and on-demand learning support. With chatbots, students can ask questions and receive immediate constructive feedback, which can help to reduce the workload on educators and improve the overall learning experience for students.

Another benefit of using chatbots is the ability to scale educational services (Neumann et al., 2021). Chatbots can handle large numbers of student inquiries simultaneously, which can be particularly useful in large classroom settings or in situations where there is a high demand for educational support. This can also help decrease educators' workload and ensure that all students have access to the (individual) support they need to succeed. Findings from a study by Chen et al. (2023) revealed that chatbots had tremendous potential to help students "learn basic content in a responsive, interactive, and confidential way" (p. 1).

Additionally, chatbots have the potential to improve the efficiency of educational delivery (Huang et al., 2022). Educators can create customized learning pathways for students, which can help to ensure that students are receiving the most relevant and effective support. This can help to improve student outcomes and reduce the overall time and resources required to complete a course of study.

Drawbacks and challenges of using chatbots in higher education

While there are several potential benefits to using chatbots in higher education, there are also some drawbacks, limitations, and challenges (i.e., ethical (Popenici, 2023; Kamalov et al., 2021)) that need to be considered (Rasul et al., 2023; Rudolph et al., 2023b). Limna et al. (2023) argued, for instance, that chatbots such as ChatGPT had "caused immense concerns in education", particularly to those disciplines that "rely heavily on written assignments" (p. 3). One of the main drawbacks is the inability to fully replicate the experience of interacting with a human educator (Chen et al., 2023; Santandreu Calonge et al., 2023; Kamalov et al., 2023). This could lead to a loss of personal connections and a reduction in the quality of educational support.

Another challenge of using chatbots in higher education is the potential for harmful bias (Rasul et al., 2023; Kooli, 2023). AI systems can be biased if they are trained on biased data. This can lead to the amplification of existing biases and the exclusion of certain groups of students. It is important for educators to be aware of this potential issue and to take steps to mitigate it, such as by ensuring that chatbots are trained on a diverse and inclusive dataset. Therefore, continuous improvement and evaluation of the AI model are crucial.

A final challenge of using chatbots in higher education is the potential for technical issues (Yang & Evans, 2019). Chatbots rely on complex algorithms and sophisticated machine learning models, which can be prone to errors and glitches. This can disrupt (a) the learning experience for students and (b) the teaching experience for educators if used in the classroom as a learning and teaching activity, therefore reducing the effectiveness of chatbots as an educational tool.

To evaluate and compare the mathematical problem-solving abilities of Large Language Models (LLMs), we selected four: ChatGPT (GPT-3.5), GPT-4, Bard, and LLaMA 13-B. The choice of those four LLMs was made to ensure diversity in the study, as each AI model has its own architecture and learning mechanisms. The selection included two LLMs that were primarily designed for generating human-like text (ChatGPT and GPT-4), one LLM designed for language-related tasks (LLaMA), and one LLM that was designed to provide detailed explanations (Bard).

We investigated the following research question: Which of the four chatbots is more accurate and less verbose for statistics and calculus prompts? Kabir et al. (2023) indicated, for instance, that 52 per cent of ChatGPT answers to 517 Stack Overflow questions were incorrect, and 77 per cent were verbose.

Pros and cons of each chatbot for helping students understand calculus and statistics

ChatGPT

ChatGPT is a chatbot developed by OpenAI that is based on a large language model. It allows the user to control the conversation in terms of length, format, level of detail, style, and language. While the main purpose of the chatbot is to simulate human conversations, it can perform a wide range of tasks, including writing computer programs, composing music, answering test questions, writing poetry, and others. ChatGPT has achieved enormous popularity within a very short period, gaining over 100 million users in less than 3 months of its initial release (Rudolph et al., 2023b).

The basic version of ChatGPT is based on the GPT-3.5 model, which is a generative pre-trained transformer developed by OpenAI. GPT-3.5 is a transformer model that is first trained on large swaths of publicly available text as a general-purpose language model. Then, the model is further fine-tuned for conversational applications using a combination of supervised and reinforcement learning methods. Since GPT-3.5 is trained on unfiltered text, it is vulnerable to bias and misinformation. In addition, ChatGPT suffers from 'hallucinations' – incorrect answers that sound plausible (Rudolph et al., 2023b).

Given its capabilities, ChatGPT has been utilized in various educational domains (Lee, 2023; Qadir, 2023; Santandreu Calonge et al., 2023; Wardat et al., 2023). Wardat et al. (2023) showed that ChatGPT has the potential to provide students with mathematical knowledge. At the same time, the authors cautioned about its weaknesses in certain topics, such as geometry. The accuracy and effectiveness of ChatGPT solutions depend on the complexity of the equation, input data, and the instructions given to the chatbot. Ellis and Slade (2023) presented ChatGPT's capabilities in statistics and data science education, providing examples of how ChatGPT could help in developing course materials. A recent survey of 110 students enrolled in a mathematics course showed that students quickly adopted the ChatGPT tool, exhibiting high confidence in their responses and general usage in the learning process, alongside a positive evaluation (Sánchez-Ruiz et al., 2023). On the other hand, the development of lateral competencies was a cause for concern.

Pros

- Wide knowledge base: ChatGPT has been trained on a diverse range of topics, including calculus and statistics so that it can provide relevant information and explanations.

- Conversational nature: Students can engage in an informal dialogue with ChatGPT, asking questions and seeking clarifications, which can enhance their understanding and interest.

- Availability: ChatGPT is readily accessible through various platforms (including smartphones), making it convenient for students to seek help anytime, anywhere.

Cons

- Limited context understanding: ChatGPT might occasionally provide incorrect, incomplete, or irrelevant information due to its inability to fully grasp the context of a specific calculus question.

- Lack of visuals: Graphical representations and visual aids are often crucial in understanding calculus and statistics concepts, which ChatGPT cannot provide directly.

GPT-4

GPT-4 is a more advanced version of the GPT-3.5 language model developed by OpenAI. GPT-4 is commercially available for users under the name ChatGPT Plus. The main difference between the two versions of GPT is the size of the models, where GPT-4 consists of a much larger number of parameters than its predecessor. Although GPT-3.5 and GPT-4 show similar performance on most routine conversation tasks, the latter achieves significantly better performance on more advanced tasks, including solving mathematics questions (OpenAI Blog, 2023). For example, GPT-4 achieved over 40% percentile on the AP Calculus exam, while GPT-3.5 achieved 0%. Recent findings by Abramski et al. (2023) show that GPT-4 produces a five-fold semantically richer, more emotionally polarized perception with fewer negative associations compared to older versions of GPT. A large-scale study based on 4,550 MIT exam questions in mathematics, computer science, and electrical engineering showed that GPT-3.5 can solve a third of the problems, while GPT-4 is able to achieve a near-perfect score (Zhang et al., 2023).

Pros

- Improved contextual understanding: GPT-4 is expected to have better contextual comprehension compared to previous models, which may result in more accurate and complete responses.

- Enhanced knowledge base: GPT-4 could be trained on an updated and larger dataset, allowing it to offer more comprehensive and up-to-date information on calculus and statistics.

- Potential for more specialized models: GPT-4's architecture might be used as a basis for domain-specific models that focus solely on calculus and statistics, providing more targeted assistance.

Cons

- Potential for errors: Although GPT-4 may have a better contextual understanding, it is still a language model and can make mistakes or generate inaccurate information.

Bard

Bard is a generative artificial intelligence chatbot developed by Google. Its current version is based on the PaLM large language model, which is a transformer-based model consisting of 520 billion parameters. Bard was released to compete with the rival ChatGPT. It garnered lukewarm reception due to initial mishaps. Unlike the GPT models, Bard has direct access to the internet. A study by Plevris et al. (2023) showed Bard performs better than ChatGPT on math problems that are available online, while it underperforms on original questions. Evaluation of the mathematics performance of Bard on the mathematics test of the Vietnamese National High School Graduation Examination showed that it lagged ChatGPT (Nguyen et al., 2023). Despite the backing of Google, Bard is a relatively underutilized software with very few applications and studies in the field of education.

Pros

- Tailored for education: Bard is an AI language model specifically designed for educational purposes, including teaching subjects like calculus and statistics (Kamalov et al., 2023).

- Curriculum alignment: Bard can align its explanations and guidance with specific curricula, ensuring that students receive targeted assistance based on their educational needs.

- Pedagogical approach(es): Bard incorporates instructional strategies to enhance learning, such as providing step-by-step explanations, interactive examples, and adaptive feedback.

Cons

- Limited knowledge outside of educational content: Bard's expertise might be focused on educational topics, potentially limiting its ability to provide insights or answer questions beyond the scope of calculus and statistics.

- Dependency on available curriculum: The effectiveness of Bard heavily relies on the quality and coverage of the curriculum it is aligned with. Gaps or discrepancies in the curriculum may affect the support it offers (and the accuracy of its responses).

LLaMA

LLaMA is a large language model developed by Meta. Its developers claimed that the 13 billion parameter version of the model outperformed the much larger ChatGPT on several NLP tasks. Recently, the next-generation model

LLaMA 2 was released in partnership with Microsoft based on larger training data. Unlike other major chatbots, LLaMA is open-source software. Its relatively small size and open-source nature make it an attractive alternative to other existing chatbots. Touvron et al. (2023) showed that LLaMA is capable of outperforming Bard and ChatGPT on several NLP tasks. Similarly, Liu et al. (2023) showed that LLaMA can outperform other major chatbots in arithmetic problem-solving.

Pros

- Multimodal learning experience: LLaMA combines text-based information with visual and interactive elements, making it effective in conveying complex calculus and statistics concepts.

- Hands-on practice: LLaMA often provides interactive exercises and simulations, allowing students to actively engage with the subject matter and reinforce their understanding.

- Adaptive learning: LLaMA can adapt to the user's progress and adjust the difficulty level of the content accordingly, providing personalized learning experiences.

Cons

- Limited availability: As of the knowledge cutoff date, LLaMA is not widely accessible or integrated into various platforms, potentially limiting its reach to students.

- Resource-intensive: The integration of multimedia elements and interactive features might require robust hardware or an internet connection, which could be a barrier for some students, and in disadvantaged contexts (Shah & Calonge, 2023).

Each of these LLMs has its own advantages and limitations. Depending on the students' preferences, learning styles, and availability, they can choose the most suitable tool or combination of tools to enhance their understanding of calculus and statistics.

Methods

Seven calculus and five statistics questions were submitted to ChatGPT, GPT-4, Bard, and LLaMA 13-B via single prompts, as shown in Table 1. Each prompt was entered individually with the original question and answer choices reproduced verbatim. Each prompt was carefully designed to cover a broad range of calculus and statistical concepts. Also, the prompts varied in the level of difficulty to allow for a more in-depth analysis of the LLMs' problem-solving capabilities and to ensure a fair assessment of their mathematical skills.

Table 1. 12 prompts.

Prompts		
1	Calc	I have the below given information: "the temperature in Austin (in °F) is approximated by $T(x) = 37\sin\left[\frac{2\pi}{365}(x-101)\right] + 25$ where T(x) is the temperature on day x, with x = 1 corresponding to Jan 1 and x = 365 corresponding to Dec 31". Using this information, please estimate the temperature, to the nearest degree Fahrenheit, on day 250. Provide me with all necessary steps.
2	Calc	Calculate the limit as x goes to zero of the ratio sin(x) over x. Explain your work.
3	Stats	On average, 3 traffic accidents per month occur at a certain intersection. What is the probability that in any given month at this intersection… (a) exactly 5 accidents will occur? (b) fewer than 3 accidents will occur? (c) at least 2 accidents will occur?
4	Calc	Find the derivative of f(x)=|x| at x=0. Explain your work.
5	Calc	Explain how to find the slope of the function f(x)=x^2+2x-1 at the point (2,3).
6	Stats	Could you explain the mean square error, give me the formula to compute it, and explain the terms involved. Provide me with an example and step by step computation of the mean square error.
7	Calc	Define a new rule for calculating Riemann sums in the following way. For each subinterval, pick the halfway point between the left endpoint and the midpoint of the subinterval. Then use the selected point to calculate the height of the corresponding rectangle. Apply the new rule to find the Riemann sum for the function f(x)=x^2+5x on the interval [2, 5] using n=10 subintervals.
8	Stats	Please explain in words the following formula and give me precise examples: b1 = Σ [(xi - x) (yi - y)] / Σ [(xi - x)2]
9	Stats	In a state that did not require varicella (chickenpox) vaccination, a boarding school experienced a prolonged outbreak of varicella among its students that began in September and continued through December. To calculate the probability or risk of illness among the students, which denominator would you use, explain your choice? Number of susceptible students at the ending of the period (i.e., June). Number of susceptible students at the midpoint of the period (late October/early November). Number of susceptible students at the beginning of the period (i.e., September). Average number of susceptible students during outbreak
10	Calc	Find the constant b so that the line y=-0.5x+b meets the graph of y^2=x^3 orthogonally. Explain your steps.
11	Stats	Help me understand how the concepts of prior, likelihood, and posterior are interrelated in Bayesian Statistics, give me an example with step-by-step explanation.
12	Calc	Find the polynomial of the smallest degree that satisfies the conditions int_1^4 p(x)= 5 and int_3^6=10. Can you suggest a general rule based on this example?

Results

This section compares the features and functionalities of each of the four AI chatbot platforms, focusing on their suitability for calculus and statistics. The evaluation of the four LLMs was based on: (1) the accuracy/the correctness of the final answer to the 12 prompts, (2) verbosity and the clarity of the explanation, and (3) the presence or absence of mathematical errors. While the correctness of the answer was assessed on a binary basis, i.e., whether the answer is correct or not, the clarity of the explanation was scored on a scale of 0 to 10, with 10 being the clearer and most comprehensive answer (see Tables 2-13). The mathematical errors were classified as either major or minor based on their potential compact on the final answer.

Accuracy

In the context of this article, chatbot accuracy is the percentage of utterances that return the correct response to the prompts, as shown graphically in figures 1- 4, below.

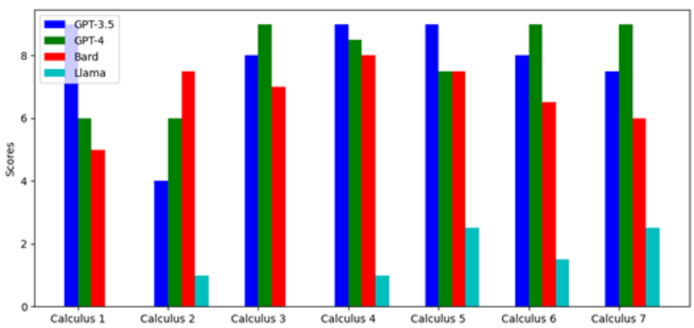


Figure 1. Accuracy scores in calculus.

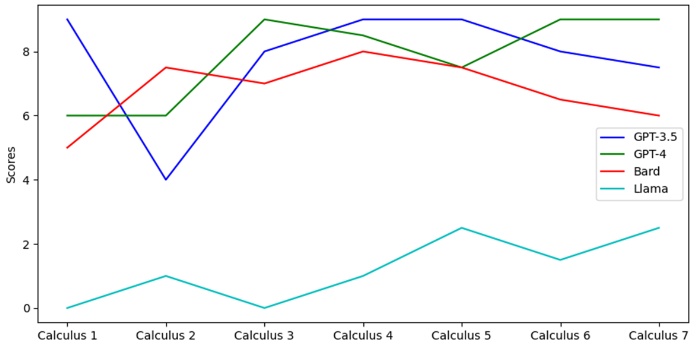


Figure 2. Calculus scores by chatbot and prompts.

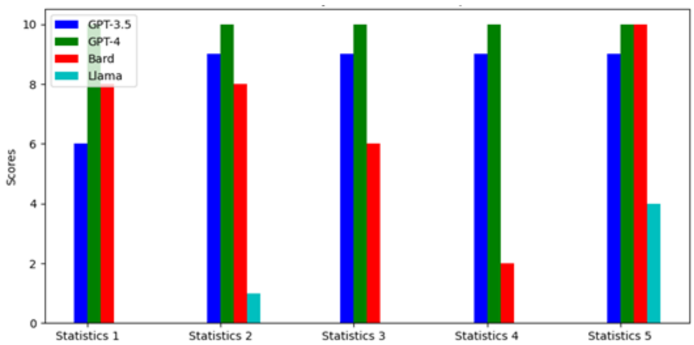


Figure 3. Accuracy scores in statistics.

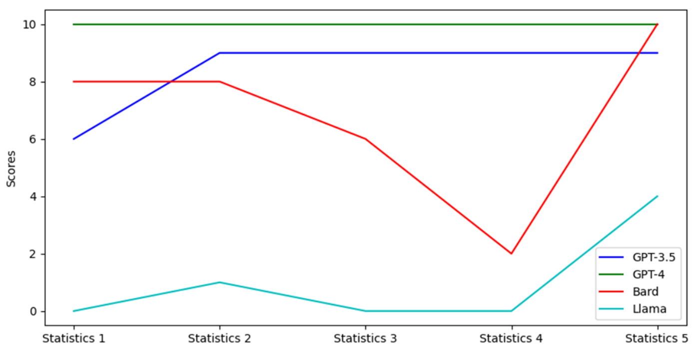


Figure 4. Statistics scores by chatbot and prompts.

Verbosity for calculus and statistics

In the context of chatbots, verbosity refers to the amount of unnecessary, irrelevant, or excessive words, phrases, or information used in the chatbot's responses (see Appendix A). A chatbot is considered verbose if it tends to provide overly detailed or convoluted answers, which can lead to a negative user experience. Zheng et al. (2023) indicated that an LLM is verbose when it "favours longer, verbose responses, even if they are not as clear, high-quality, or accurate as shorter alternatives" (Zheng et al., 2023, p. 5). Cosine similarity is a way to measure how similar two things are, e.g., two vectors or two sets of data. It calculates the cosine of the angle between the two things in a multi-dimensional space and provides a value between -1 and 1, where higher values mean greater similarity and lower values mean less similarity.

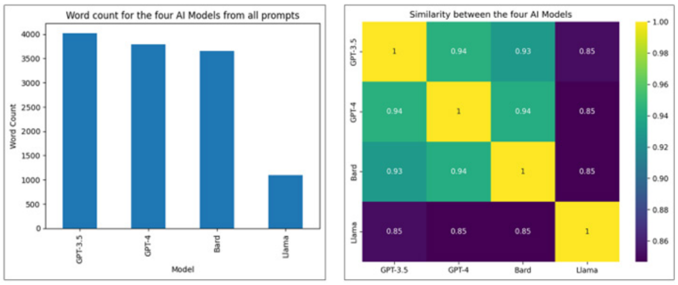


Figure 5. Verbosity (Cosine Similarity) and overlap for all 12 prompts.

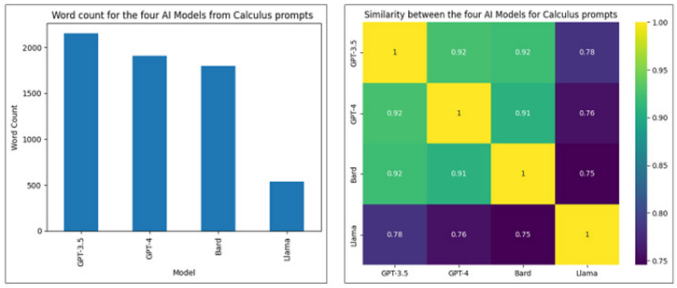


Figure 6. Verbosity (Cosine Similarity) and overlap for calculus.

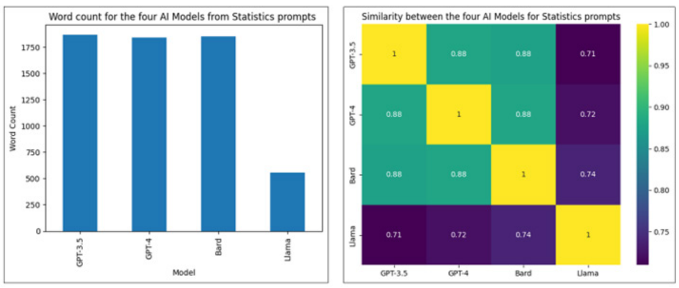


Figure 7. Verbosity (Cosine Similarity) and overlap for statistics.

Results and analysis by prompt

Prompt 1

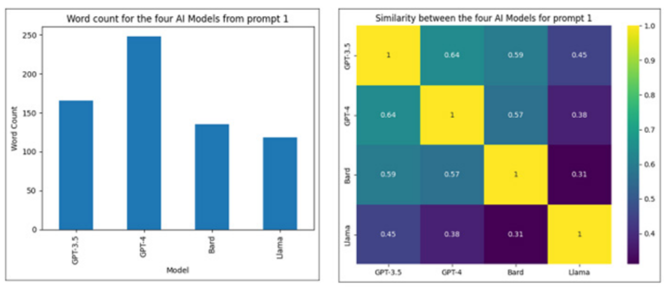


Figure 8. Verbosity (Cosine Similarity) and overlap for prompt 1.

Table 2. Answer accuracy and ratings per chatbot for prompt 1.

Chatbot	Answer accuracy	Ratings
GPT 3.5	GPT 3.5 provided an answer that is almost correct (rounding) It provided all correct steps from plugging the x value into the formula correctly, performs all necessary calculations, and comes up with an appropriate result.	▪ Final answer correct: Almost ▪ Explanation clear (0-10): 9 ▪ Mathematical mistakes: minor (rounding)
GPT-4	GPT-4 provided an incorrect answer, even though it started with a well explanation of the function, incorrectly evaluates the sine function resulting in a negative temperature.	▪ Final answer correct: No ▪ Explanation clear (0-10): 6 ▪ Mathematical mistakes: major
BARD	Bard provided accurate steps but not enough. The provided answer was not correct, there were problems with computing the angle.	▪ Final answer correct: No ▪ Explanation clear (0-10): 5 ▪ Mathematical mistakes: minor
LLaMA	Llama was completely off by using trapezoidal rule, it tried to compute the integral of the function between x=1 and x=250, which is not the right approach for this problem. It was not able to continue the answer till the end.	▪ Final answer correct: No ▪ Explanation clear (0-10): 0 ▪ Mathematical mistakes: major

Prompt 2

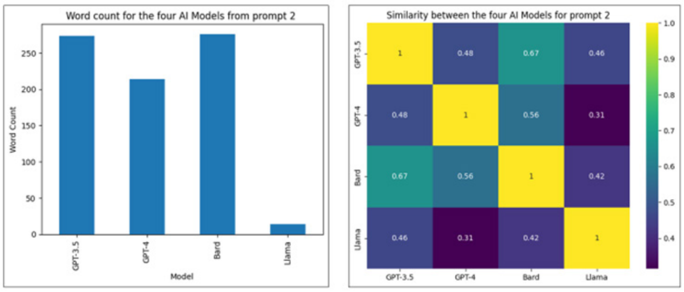


Figure 9. Verbosity (Cosine Similarity) and overlap for prompt 2.

Table 3. Answer accuracy and ratings per Chatbot for prompt 2.

Chatbot	Answer accuracy	Ratings
GPT 3.5	On the surface it appears that GPT-3.5 provides a legitimate mathematical explanation to the given question. However, a careful reading reveals several mistakes and inconsistencies in the response. The chatbot does correctly identify the squeeze theorem as a useful approach to solving the problem. But it fails to apply the theorem properly. This response provides an excellent insight into the mechanism of the chatbot which is simply trying to guess the next most likely word in the text, based on all the information that was fed into the algorithm during the training, without any logic behind it.	▪ Final answer correct: No ▪ Explanation clear (0-10): 4 ▪ Mathematical mistakes: major
GPT-4	The response provided by GPT-4 is incomplete and leaves one feeling for more information. Perhaps a follow up question to delve into more details would help in this case. The chatbot provides a short discussion of the problem without giving any technical details. Note that it makes a mistake in claiming that "This limit does not yield an indeterminate form."	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 6 ▪ Mathematical mistakes: minor
BARD	The response by Bard is arguably better than GPT. The chatbot provides 3 valid approaches to solving the problem. However, each approach either contains a mistake or incomplete. Direct substitution jumps to the conclusion using very weak logic. Squeeze theorem is applied correctly but uses prior knowledge that sin(x)<x. Taylor series is probably the best approach but there a couple of gaps in the response.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 7.5 ▪ Mathematical mistakes: minor
LLaMA	Provides a useless response.	▪ Final answer correct: No ▪ Explanation clear (0-10): 1 ▪ Mathematical mistakes: major

Prompt 3

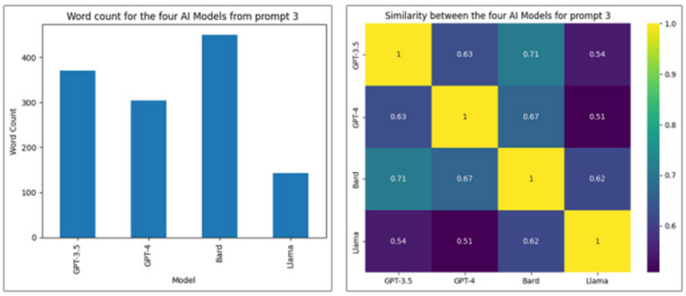


Figure 10. Verbosity (Cosine Similarity) and overlap for prompt 3.

Table 4. Answer accuracy and ratings per Chatbot for prompt 3.

Chatbot	Answer accuracy	Ratings
GPT 3.5	The explanation provided by GPT-3.5 is clear and complete, it referred to the Poisson distribution and gave detailed explanation. However, the answer was not given in full, translation to the question parts into formulas were almost correct, the part b was wrong as is translated x fewer than three to only the two cases 0 and 1.	▪ Final answer correct: No (Almost) ▪ Explanation clear (0-10): 6 ▪ Mathematical mistakes: major
GPT-4	The explanation provided by GPT-4 is clear and complete, it referred to the Poisson distribution and gave detailed explanation about what probabilities to compute, however not all probabilities were correct. Part a was wrong.	▪ Final answer correct: No (Almost) ▪ Explanation clear (0-10): 10 ▪ Mathematical mistakes: none
BARD	Bard explanation was also correct and clear and has correctly employed the Poisson distribution. However, the answers given were not correct.	▪ Final answer correct: No ▪ Explanation clear (0-10): 8 ▪ Mathematical mistakes: major
LLaMA	Llama provided an answer that is not accurate. It didn't use the Poisson distribution which is the appropriate model for these calculations. Even the provided explanation was not clear and contained major mathematical errors.	▪ Final answer correct: No ▪ Explanation clear (0-10): 0 ▪ Mathematical mistakes: major

Prompt 4

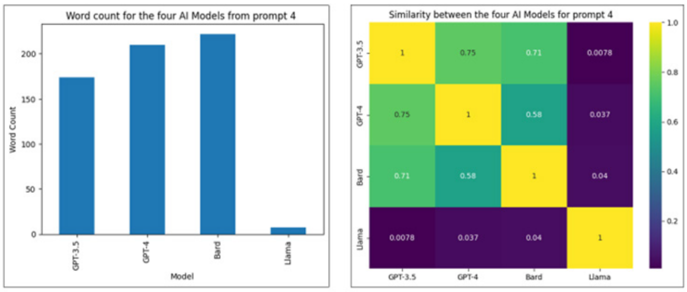


Figure 11. Verbosity (Cosine Similarity) and overlap for prompt 4.

Table 5. Answer accuracy and ratings per Chatbot for prompt 4.

Chatbot	Answer accuracy	Ratings
GPT 3.5	Overall, the chatbot provided a good response. Only a small issue with the claim about discontinuity in the end.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 8 ▪ Mathematical mistakes: minor
GPT-4	The chatbot provided a good response with some technical details which may of interest to someone who is looking for a more in-depth analysis.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 9 ▪ Mathematical mistakes: No
BARD	The response is slightly convoluted. The chatbot solves the problem using the limit definition of the derivative but goes too much into the technical details which makes it harder to follow especially given the math notation. The chatbot does make a significant mistake in the beginning to claim that the function is not continuous at x=0.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 7 ▪ Mathematical mistakes: minor
LLaMA	Most succinct response ever.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 6 Mathematical mistakes: No

Prompt 5

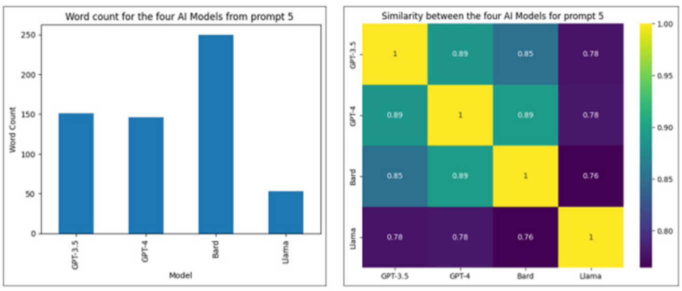


Figure 12. Verbosity (Cosine Similarity) and overlap for prompt 5.

Table 6. Answer accuracy and ratings per Chatbot for prompt 5.

Chatbot	Answer accuracy	Ratings
GPT 3.5	Overall, a good answer. I like how it breaks it down into steps so it's easier to follow. The information in step 1 could be made more concise.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 9 ▪ Mathematical mistakes: No
GPT-4	Good answer. But for a standard calculus question the response was too verbose.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 8.5 ▪ Mathematical mistakes: No
BARD	This is a standard question in calculus. The answer and explanation should be simple. While the chatbot does provide a simple explanation first, followed by a more detailed explanation, overall, it feels too verbose and hard to follow especially with all the math notation involved.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 8 ▪ Mathematical mistakes: No

LLaMA	Terrible response. 4+4-1=9?	▪ Final answer correct: No ▪ Explanation clear (0-10): 1 Mathematical mistakes: major

Prompt 6

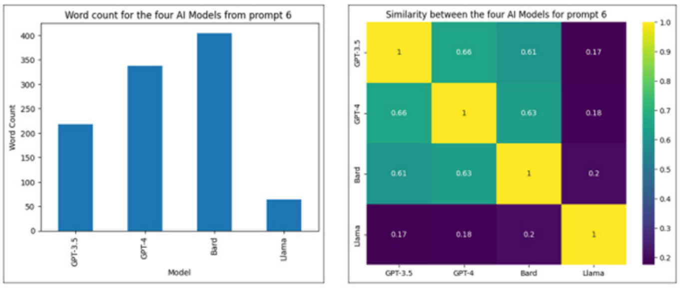


Figure 13. Verbosity (Cosine Similarity) and overlap for prompt 6.

Table 7. Answer accuracy and ratings per Chatbot for prompt 6.

Chatbot	Answer accuracy	Ratings
GPT 3.5	GPT-3.5 provided an accurate answer with a detailed explanation. Th given example was simple but answers correctly with all details.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 9 ▪ Mathematical mistakes: No
GPT-4	GPT-4 provided an even more in-depth explanation than GPT-3.5. The example was also a bit more complicated than the one provided by GOT-3.5. All calculations were correct, and all steps provided.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 10 ▪ Mathematical mistakes: No
BARD	Bard provided a clear detailed explanation and an example; however, the calculations were wrong.	▪ Final answer correct: No ▪ Explanation clear (0-10): 8 ▪ Mathematical mistakes: minor
LLaMA	Llama provided an explanation that is not clear at all and want of topic. No example was provided as requested. While the start of explanation of the formula seems accurate, the overall explanation is not as comprehensive as the other models, no mention to regression or interpretation of the MSE.	▪ Final answer correct: No ▪ Explanation clear (0-10): 1 ▪ Mathematical mistakes: major

Prompt 7

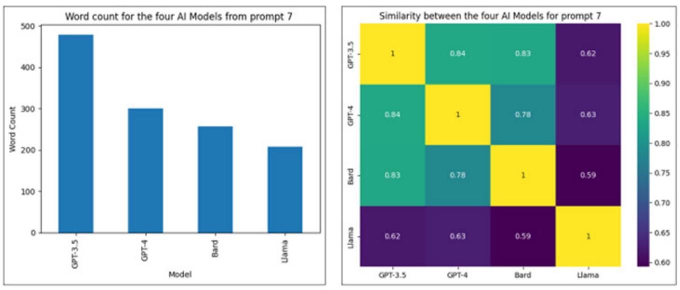


Figure 14. Verbosity (Cosine Similarity) and overlap for prompt 7.

Table 8. Answer accuracy and ratings per chatbot for prompt 7.

Chatbot	Answer accuracy	Ratings
GPT 3.5	This question is a variation of the standard Riemann sum. However, the given heuristic is not common. So, the question challenges the comprehension abilities of the chatbot. The chatbot did an excellent job of understanding the instructions and following them. It provided sufficient explanation for the reader to able to follow the process without getting confused. It correctly gave the formula for calculating the evaluation point. However, it did not follow its own formula.	▪ Final answer correct: No ▪ Explanation clear (0-10): 9 ▪ Mathematical mistakes: major
GPT-4	The chatbot did a good job understanding the instructions and providing an appropriate solution. However, the response is a bit dense and might be harder to follow for students with weaker background in math. There are 2 important flaws. First, the evaluation point is calculated incorrectly. Second, no final answer is provided. .	▪ Final answer correct: No ▪ Explanation clear (0-10): 7.5 ▪ Mathematical mistakes: major
BARD	The chatbot does a good job of understanding the instructions and provides the correct steps to calculate the answer. However, it fails to correctly calculate the evaluation point. More concrete examples of calculations would have been useful.	▪ Final answer correct: No ▪ Explanation clear (0-10): 7.5 ▪ Mathematical mistakes: major
LLaMA	Poor response.	▪ Final answer correct: No ▪ Explanation clear (0-10): 2.5 ▪ Mathematical mistakes: major

Prompt 8

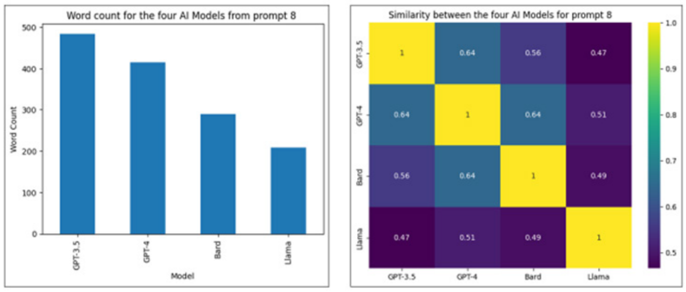


Figure 15. Verbosity (Cosine Similarity) and overlap for prompt 8.

Table 9. Answer accuracy and ratings per chatbot for prompt 8.

Chatbot	Answer accuracy	Ratings
GPT 3.5	GPT-3.5 provided accurate explanation to the question and accurate steps for the given example, however it failed to continue the steps till last step.	▪ Final answer correct: No ▪ Explanation clear (0-10): 9 ▪ Mathematical mistakes: minor
GPT-4	GPT-4 provided accurate explanation to the question and accurate computation for the given example.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 10 ▪ Mathematical mistakes: none
BARD	Bard provided an accurate explanation to the question, but not enough details about the different terms involved in the equations. The final provided answer was not correct.	▪ Final answer correct: No ▪ Explanation clear (0-10): 6 ▪ Mathematical mistakes: minor
LLaMA	Llama provided a poor response, the explanation provided does not correctly define the terms in the equation. The explanation is not clear at all as it talks about differences between pairs. A very short and simple example was provided, and steps were wrong to reach the final answer.	▪ Final answer correct: No ▪ Explanation clear (0-10): 0 ▪ Mathematical mistakes: major

Prompt 9

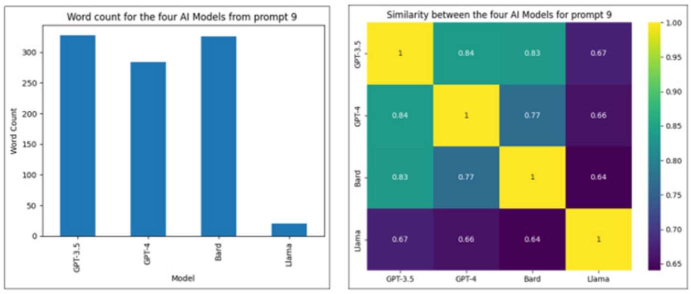
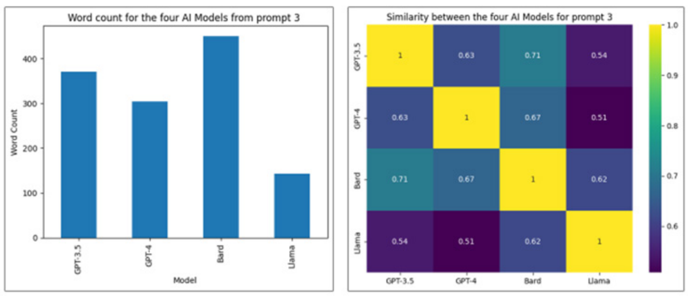


Figure 16. Verbosity (Cosine Similarity) and overlap for prompt 9.

Table 10. Answer accuracy and ratings per chatbot for prompt 9.



Prompt 10

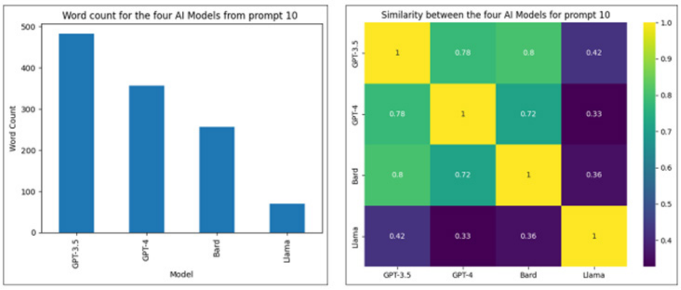


Figure 17. Verbosity (Cosine Similarity) and overlap for prompt 10.

Table 11. Answer accuracy and ratings per chatbot for prompt 10.

Chatbot	Answer accuracy	Ratings
GPT 3.5	This is a nontrivial question with a multi-step solution. Overall, it is a good attempt to solve the problem. The chatbot correctly identifies the key concepts and ingredients to solve the problem and provides a step-by-step explanation of the solution. However, GPT-3.5 fails to put it all together. It follows the correct path to solution until Step 4, after which it takes a wrong turn. While there are no major mistakes in terms of math, the chatbot pursued the wrong strategy which ultimately led to an unresolved outcome.	▪ Final answer correct: No ▪ Explanation clear (0-10): 8 ▪ Mathematical mistakes: minor
GPT-4	The chatbot provides a good step-by-step explanation of the solution. Overall, the presented solution is correct which is impressive given the level of the difficulty and the number of steps required to solve the problem. However, it makes a small mistake in a basic calculation "y = 3*(16/9) ^2 / 4 => y = 32 / 9". It is interesting to observe that while GPT-4 can tackle the problem conceptually which is the hardest part, it makes a basic calculation error especially since calculations are generally regarded as the strength of the chatbots.	▪ Final answer correct: Yes* ▪ Explanation clear (0-10): 9 ▪ Mathematical mistakes: minor *95%
BARD	The chatbot fails to recognize that the derivative is found using implicit differentiation. It also does not realize that the provided answer does not make sense.	▪ Final answer correct: No ▪ Explanation clear (0-10): 6.5 ▪ Mathematical mistakes: major
LLaMA	Nonsensical response.	▪ Final answer correct: No ▪ Explanation clear (0-10): 1.5 ▪ Mathematical mistakes: major

Prompt 11

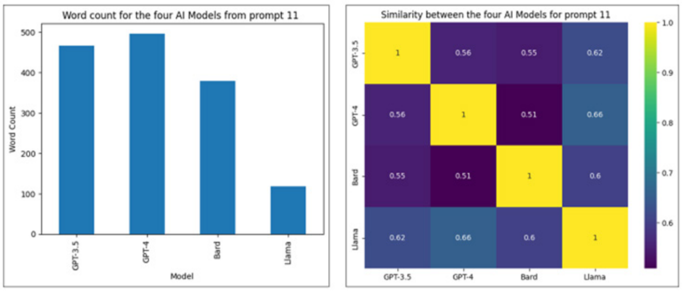


Figure 18. Verbosity (Cosine Similarity) and overlap for prompt 11.

Table 12. Answer accuracy and ratings per Chatbot for prompt 11.

Chatbot	Answer accuracy	Ratings
GPT 3.5	GPT-3.5 answer is correct and provides a clear and detailed explanation. It gave an example and explained the relationship between prior, likelihood, and posterior in Bayesian Statistics but failed in completing the answer by the end	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 9 ▪ Mathematical mistakes: none
GPT-4	GPT-4 answer is correct, it was explained in details and an example was used to provide more insights	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 10 ▪ Mathematical mistakes: none
BARD	Bard answer is correct, clearly explained with an example and enough details.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 10 ▪ Mathematical mistakes: none
LLaMA	Llama answer is correct with clear explanation, however, no example was provided, as required in the prompt.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 4 ▪ Mathematical mistakes: none

Prompt 12

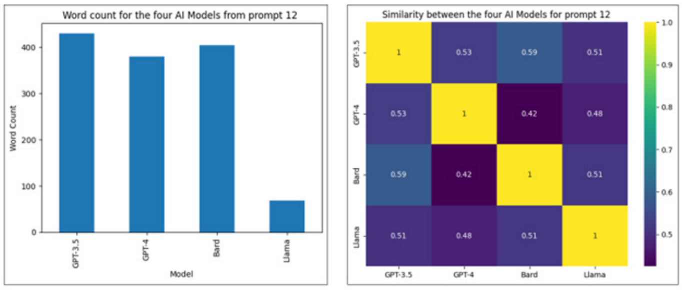


Figure 19. Verbosity (Cosine Similarity) and overlap for prompt 12.

Table 13. Answer accuracy and ratings per chatbot for prompt 12.

Chatbot	Answer accuracy	Ratings
GPT 3.5	The chatbot supplies the general strategy for solving the problem but does not execute the proposed plan. Thus, it has an incomplete understanding of the solution. While the proposed approach is far from being complete, it is presented in a clear style.	▪ Final answer correct: No ▪ Explanation clear (0-10): 7.5 ▪ Mathematical mistakes: No
GPT-4	The chatbot provides a correct and complete solution. The solution is presented clearly. However, there is a small calculation mistake "[ax^2/2 + bx] _1^4 = 5, or [(8a/2 + 4b) - (a/2 + b)]". It is puzzling that GPT-4 can solve complex problems but can stumble on a basic calculation.	▪ Final answer correct: Yes ▪ Explanation clear (0-10): 9 Mathematical mistakes: minor
BARD	The response looks legitimate on the surface, but a closer look at the details reveals multiple holes. There are serious mathematical flaws in the arguments and the final answer (deg2) is incorrect.	▪ Final answer correct: No ▪ Explanation clear (0-10): 6 ▪ Mathematical mistakes: Major
LLaMA	Incorrect solution.	▪ Final answer correct: No ▪ Explanation clear (0-10): 2.5 Mathematical mistakes: major

Discussion

Use cases in calculus and statistics

In this article, we explored potential use cases for each platform within calculus and statistics. We argue that ChatGPT and GPT-4 can be utilized in calculus and statistics to provide personalized tutoring and assistance to students. Both can generate step-by-step solutions to math problems, explain complex mathematical concepts, and offer practice exercises to reinforce learning. Students can engage in interactive conversations with ChatGPT or GPT-4 to clarify doubts, receive real-time feedback anytime, anywhere, and improve their understanding of mathematical principles.

Bard can also play a vital role in calculus and statistics. It can assist students with administrative tasks related to course registration, provide access to mathematical resources such as textbooks and study materials, and offer guidance on choosing appropriate courses for specific mathematical or statistical topics. However, it is significantly weaker than GPT-3.5 and GPT-4 in calculus and statistics. LLaMA is, unfortunately, and disappointingly not very accurate for calculus and statistics prompts.

Whilst Popenici (2023) argued that AI was facilitating the super-personalisation (p. 5) of the learning experience, Rasul et al. (2023) indicated that ChatGPT could be utilized to facilitate adaptive learning, provide personalised feedback, aid research, automate administrative services, and create innovative assessment.

Our findings indicate that chatbots can also be utilized in several ways to assist students in comprehending statistics or calculus better if they have received prior training on writing effective prompts (Eager & Brunton, 2023):

1. Concept explanation: Students, following training on prompt structuring, could engage in a conversation with a chatbot to seek explanations and clarifications on statistical or calculus concepts they find challenging. Chatbots with knowledge-tracing capabilities (Shehata et al., 2023) can provide step-by-step explanations, examples, and intuitive analogies to help students understand statistical concepts in a personalized and interactive manner.

2. Problem-solving: Students can present statistical problems or exercises to a chatbot, and it can guide them through the problem-solving process if specifically asked in the prompt. Chatbots can offer hints, ask relevant questions to trigger critical thinking, and provide guidance on the correct approach or methodology to solve the problem. It can therefore help "increase student engagement and satisfaction by relieving university staff of routine tasks and allowing them to focus on higher-order skills and mentoring" (Firat, 2023, p. 61).

3. Data analysis assistance: Students can seek help from chatbots in analyzing data sets, confirming research by Carlander-Reuterfelt et al. (2020). They can input their data, and chatbots can guide them through the appropriate statistical techniques, such as calculating measures of central tendency, conducting hypothesis tests, or creating visualizations. Chatbots can provide insights into data interpretation and explain the implications of the statistical results.

4. Real-world applications: Chatbots can showcase authentic applications of statistics or calculus to students. By discussing examples and case studies from various fields, such as social sciences, healthcare, economics, or sports, chatbots can illustrate how statistical or calculus concepts can be utilized in practical situations. Hultberg et al. (2018) argued that "making a link between often abstract concepts and pertinent examples" can help "students understand difficult ideas, thus making it easier to remember" (p. 35). This can help students grasp the relevance and significance of statistics and calculus in different domains.

5. Practice and assessment: In line with the recent extant literature in a range of disciplines, chatbots can offer interactive practice sessions and quizzes to assess students' understanding of statistical or calculus concepts. They can provide instant feedback on their answers, explain any mistakes, miscalculations, or misconceptions, and suggest further study materials or resources for improvement (Mogavi et al., 2023).

Last and not least, chatbots can serve as tireless, mobile, interactive, and personalized learning companions, offering explanations, guidance, and practice opportunities 24/7 to help students grasp statistical or calculus concepts more effectively. Its conversational nature allows for an engaging and interactive learning experience and can cater to students' individual learning styles, preferences, and needs.

Summary

In summary, the four AI chatbot platforms have a wide range of use cases in calculus and statistics, including personalized tutoring, administrative support, adaptive assessments, collaborative learning, and concept clarification. Their capabilities vary greatly in terms of responses (from very accurate to not-so-good), allowing educators and students to choose the platform that best aligns with their specific needs and goals in calculus and statistics education.

Limitations

While this study marked a crucial step in understanding the potential and limitations of LLMs in teaching calculus and statistics, it has several limitations. First, the study's focus is limited to only these two areas, which restricts the generalization of the findings to other academic disciplines. Second, the choice of the four LLMs, though considered the most well-known and used, is not exhaustive, leaving numerous other LLMs, such as Claude, Upstage, Falcon or Vicuna, unexplored. Third, the assessment of the quality of the LLMs' explanations is subjective and could differ based on individual perspectives. It is also important to bear in mind the possible bias in the chatbots' responses. Fourth, due to practical constraints, this paper could not capture the dynamic learning and evolution of the four AI models over time.

Future directions

The findings of our study indicate areas where future research on LLMs' development could focus, particularly in terms of contextual understanding and the ability to provide clear, concise, and accurate explanations of calculus and statistical prompts. We suggest training AI models using specific educational resources or textbooks commonly used in calculus and statistics, enhancing their alignment with the curriculum and their ability to provide targeted assistance. Integration with platforms such as https://www.snapxam.com/ may also improve responses' accuracy. Another suggestion for future research is to investigate the impact of using LLMs on students' performance, motivations, and self-efficacy when used along with traditional teaching methods.

Conclusion and implications

This comparative analysis provides valuable insights into the features and applications of AI chatbot platforms— ChatGPT, GPT-4, Bard, and LLaMA 13-B —in the context of calculus and statistics. Each platform offers unique functionalities that can empower students (Hutson & Plate, 2023), enhance learning, authentic assessment (Ifelebuegu, 2023), problem-solving, and engagement in these disciplines. Wu and Yu (2023) indicated that chatbots may help improve students' learning outcomes.

Overall, chatbots have the potential to transform the way in which higher education is delivered in the classroom and online. They offer a range of benefits, including personalized and on-demand learning support, the ability to scale educational services, and improved efficiency in educational delivery. However, there are also some drawbacks and challenges that need to be considered, including the potential for academic dishonesty, plagiarism (Chaka, 2023) and cheating, privacy issues, bias, and the risk of technical issues. The findings reported here shed new light on the use of AI and LLMs in teaching and learning. Students can use this information to select an LLM that best suits their needs and complements their learning style. By carefully considering the pros and cons of using chatbots in higher education, educators can make informed decisions about whether and how to incorporate this technology into their teaching practices. Despite its limitations, the findings from this study make several contributions to the current literature and lay the groundwork for future research into the use of chatbots to improve learning and teaching in a range of academic disciplines.

Data availability statement: The datasets used/analyzed during the current study are available from the corresponding author upon reasonable request.

References

Abramski, K., Citraro, S., Lombardi, L., Rossetti, G., & Stella, M. (2023). Cognitive network science reveals bias in GPT-3, GPT-3.5 Turbo, and GPT-4 Mirroring math anxiety in high-school students. *Big Data and Cognitive Computing, 7*(3), 124 https://doi.org/10.3390/bdcc703012424

Carlander-Reuterfelt, D., Carrera, Á., Iglesias, C. A., Araque, Ó., Rada, J. F. S., & Muñoz, S. (2020). JAICOB: A data science chatbot. *IEEE Access, 8*, 180672-180680. https://doi.org/10.1109/ACCESS.2020.3024795.

Castelvecchi, D. (2023). How will AI change mathematics? Rise of chatbots highlights discussion. *Nature, 615*(7950), 15-16. https://doi.org/10.1038/d41586-023-00487-2

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching, 6*(2), 1-11. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.12

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers, 25*(1), 161-182. https://doi.org/10.1007/s10796-022-10291-4

Chen, L., Zaharia, M., & Zou, J. (2023). *How is ChatGPT's behavior changing over time?*. arXiv preprint arXiv:2307.09009.

Eager, B., & Brunton, R. (2023). Prompting higher education towards AI-Augmented teaching and learning practice. *Journal of University Teaching & Learning Practice, 20*(5), 02.

Ellis, A. R., & Slade, E. (2023). A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education,* 1-10. https://doi.org/10.1080/26939169.2023.2223609

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching, 6*(1), 57-63. https://doi.org/10.37074/jalt.2023.6.1.22

Hultberg, P., Calonge, D. S., & Lee, A. E. S. (2018). Promoting long-lasting learning through instructional design. *Journal of the Scholarship of Teaching and Learning, 18*(3), 26-43. https://doi.org/10.14434/josotl.v18i3.23179

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237-257. https://doi.org/10.1111/jcal.12610

Hutson, J., & Plate, D. (2023). *Human-AI collaboration for smart education: Reframing applied learning to support metacognition*. IntechOpen.

Hwang, G. J., & Chang, C. Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments,* 1-14. https://doi.org/10.1080/10494820.2021.1952615

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2), 1-8. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.2

Kabir, S., Udo-Imeh, D. N., Kou, B., & Zhang, T. (2023). *Who answers it better? An in-depth analysis of ChatGPT and Stack overflow answers to software engineering questions*. arXiv preprint arXiv:2308.02312.

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of Artificial Intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability, 15*(16), 12451. https://doi.org/10.3390/su151612451

Kamalov, F., Sulieman, H., & Santandreu Calonge, D. (2021). Machine learning based approach to exam cheating detection. *Plos One, 16*(8), e0254340. https://doi.org/10.1371/journal.pone.0254340

Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability, 15*(7), 5614. https://doi.org/10.3390/su15075614

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies, 28*(1), 973-1018. https://doi.org/10.1007/s10639-022-11177-3

Lee, D., & Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Computers & Education, 191*, 104646. https://doi.org/10.1016/j.compedu.2022.104646

Lee, H. (2023). The rise of ChatGPT: Exploring its potential in medical education. *Anatomical Sciences Education*. https://doi.org/10.1002/ase.2270

Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching, 6*(1), 64-74. https://doi.org/10.37074/jalt.2023.6.1.32

Liu, T., & Low, B. K. H. (2023). *Goat: fine-tuned LLaMA outperforms GPT-4 on arithmetic tasks.* arXiv preprint arXiv:2305.14201.

Mogavi, R. H., Deng, C., Kim, J. J., Zhou, P., Kwon, Y. D., Metwally, A. H. S., ... & Hui, P. (2023). *Exploring user perspectives on chatgpt: Applications, perceptions, and implications for AI-integrated education.* arXiv preprint arXiv:2305.13114.

Neumann, A. T., Arndt, T., Kobis, L., Meissner, R., Martin, A., de Lange, P., Pengel, N., Klamma, R., & Wollersheim, H. W. (2021). Chatbots as a tool to scale mentoring processes: Individually supporting self-study in higher education. *Frontiers in Artificial Intelligence, 4,* 668220. https://doi.org/10.3389/frai.2021.668220

Nguyen, P., Nguyen, P., Bruneau, P., Cao, L., Wang, J., & Truong, H. (2023). *Evaluation of mathematics performance of Google Bard on the mathematics test of the Vietnamese national high school graduation examination.*

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence, 2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033

Plevris, V., Papazafeiropoulos, G., & Rios, A. J. (2023). *Chatbots put to the test in math and logic problems: A preliminary comparison and assessment of ChatGPT-3.5, ChatGPT-4, and Google Bard.* arXiv preprint arXiv:2305.18618.

Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching, 6*(2), 1-7. https://doi.org/10.37074/jalt.2023.6.2.4

Qadir, J. (2023, May). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-9). IEEE. https://doi.org/10.1109/EDUCON54358.2023.10125121.

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Rudolph, J., Tan, S., & Tan, S. (2023a). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Rudolph, J., Tan, S., & Tan, S. (2023b). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Sánchez-Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Moraño-Fernández, J. A., & Vega-Fleitas, E. (2023). ChatGPT challenges blended learning methodologies in engineering education: A case study in mathematics. *Applied Sciences, 13*(10), 6039. https://doi.org/10.3390/app13106039

Sandu, N., & Gide, E. (2019, September). Adoption of AI-Chatbots to enhance student learning experience in higher education in India. In *2019 18th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1-5). IEEE. htpps://doi.org/10.1109/ITHET46829.2019.8937382

Santandreu-Calonge, D., Medina-Aguerrebere, P., Hultberg, P., & Shah, M. A. (2023). Can ChatGPT improve communication in hospitals?. *Profesional de la información, 32*(2). https://doi.org/10.3145/epi.2023.mar.19

Shah, M., & Calonge, D. S. (2023). Refugees' experiences with online higher education: Impact and implications through the pandemic. *Journal of Applied Learning and Teaching, 6*(1), 209-221. https://doi.org/10.37074/jalt.2023.6.1.21

Shehata, S., Calonge, D. S., Purnell, P., & Thompson, M. (2023, July). Enhancing video-based learning using knowledge tracing: Personalizing students' learning experience with ORBITS. In *Proceedings of the 18th workshop on innovative use of NLP for Building Educational Applications (BEA 2023)* (pp. 100-107). https://doi.org/10.18653/v1/2023.bea-1.8

Singh Gill, S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., Fuller, S., Singh, M., Arora, P., Parlikad, A. K., Stankovski, V., Abraham, A., Ghosh, S. K., Lutfiyya, H., Kanhere, S. S., Bahsoon, R., Rana, O., Dustdar, S., Sakellariou, R., ... Buyya, R. (2023). *Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots*. arXiv e-prints, arXiv-2306. https://doi.org/10.1016/j.iotcps.2023.06.002

Sok, S., & Heng, K. (2023). *ChatGPT for education and research: A review of benefits and risks*. SSRN 4378735. http://dx.doi.org/10.2139/ssrn.4378735.

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). *Llama: Open and efficient foundation language models*. arXiv preprint arXiv:2302.13971.

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(7), em2286. https://doi.org/10.29333/ejmste/13272

Yang, S., & Evans, C. (2019, November). Opportunities and challenges in using AI chatbots in higher education. In *Proceedings of the 2019 3rd international conference on education and e-learning* (pp. 79-83). https://doi.org/10.1145/3371647.3371659

Wu, R., & Yu, Z. (2023). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13334

Zhang, S. J., Florin, S., Lee, A. N., Niknafs, E., Marginean, A., Wang, A., ... & Drori, I. (2023). *Exploring the MIT mathematics and EECS curriculum using Large Language Models*. arXiv preprint arXiv:2306.08997.

Zheng, L., Chiang, W. L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). *Judging LLM-as-a-judge with MT-Bench and Chatbot Arena*. arXiv preprint arXiv:2306.05685.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

ChatGPT and higher education assessments: More opportunities than concerns?

Kelum A.A. Gamage[A]
Shyama C. P. Dehideniya[B]
Zhiheng Xu[C]
Xiaobin Tang[D]

A Centre for Educational Development and Innovation, James Watt School of Engineering, University of Glasgow, UK

B Department of Education, University of Peradeniya, Sri Lanka

C Department of Nuclear Science and Technology, Nanjing University of Aeronautics and Astronautics, China; Key Laboratory of Nuclear Technology Application and Radiation Protection in Astronautics, Ministry of Industry and Information Technology, Nanjing, China

D Department of Nuclear Science and Technology, Nanjing University of Aeronautics and Astronautics, China; Key Laboratory of Nuclear Technology Application and Radiation Protection in Astronautics, Ministry of Industry and Information Technology, Nanjing, China

DOI: https://doi.org/10.37074/jalt.2023.6.2.32

Abstract

In recent times, higher education has seen a growing concern regarding the utilisation of artificial intelligence, especially with the emergence of ChatGPT. This technology can generate written content and respond to queries at a level that is nearly indistinguishable from a human writer. This feature has drawn substantial interest from students in higher education and has led to concern that students will use ChatGPT's capabilities to cheat on written formative and summative assessments. In this paper, we will review the usage of ChatGPT in higher education assessments and investigate why students want to cheat using artificial intelligence capabilities. It also offers a critical perspective on the challenges associated with detecting ChatGPT-generated content and its impact on academic integrity. We also consider whether artificial intelligence provides more opportunities for academics to focus on assessing higher-order thinking and strategies.

Keywords: Academic integrity; artificial intelligence; assessments; ChatGPT; higher education; learning and teaching; quality assurance.

Introduction

The timely assessment of student learning is a vital aspect of the teaching and learning process as it enables instructors to link the effectiveness of their teaching with student achievement of learning objectives. Moreover, it provides useful feedback to teachers and students about the extent to which they succeed in their teaching and learning mission. Hence, the accuracy of the assessment outcome is a crucial factor since the result would reflect the behaviour of both the teacher and the student in academia. However, recent studies have shown that the reliability of the test results is threatened as assessment cheating has become one of the major problems on many university campuses (Wang et al., 2015; Odongo et al., 2021).

Although cheating is considered an act of academic dishonesty, methods of cheating keep evolving rapidly by many means and ways, and students continue to cheat in their assessments. As far as cheating methods are concerned, almost all the methods are commonly used worldwide. In line with previous studies, academic dishonesty has several classic forms: plagiarism, reusing or resubmitting one's paper, cheating on an examination, fabricating information, collusion or illegitimate cooperation, contract cheating, impersonation, copying, and ghost-writing (Gamage et al., 2023). Different taxonomies have been used to group these dishonest behaviours, and Faucher and Caves's (2009) classification includes three categories: information exchange among students using forbidden materials and circumventing the assessment process.

With the arrival of the information age and digital technologies, new teaching and learning methods and systems have entered the education contexts while the existing ones are being improved or removed. Parallel to that, new evaluation methods and tools were also introduced to the system. However, studies have shown that new technologies inspired academic cheaters to develop new cheating methods (Wang et al., 2015; Odongo et al., 2021; Keresztury & Cser, 2013). Especially in online environments, students have come up with several techniques to cheat; for example, "students wait for answers, claim fraudulent error messages, collusion, essay plagiarism, and buying answers" (Moten et al., 2013, p. 142). According to the literature, another dimension of technological development of artificial intelligence was introduced in the late 1950s (Manning, 2020). Some researchers date 1955 as the year that John McCarthy coined the concept of artificial intelligence as "the science and engineering of making intelligent machines" (Manning, 2020, p. 1; Jantakun et al., 2021). UNESCO (2019) states that 1956 was the year when the introduction of AI happened at the Dartmouth Conference, and since then, it continued making the lives, working, and studying of mankind much more convenient. AI's ability to enable a machine to think and act like a human has gained the interest of every field and industry, including healthcare, finance, transportation, agriculture, media and communication, and entertainment. Furthermore, AI applications have also been introduced into education and are now functioning at different stages of the teaching-learning process and assessment, transforming traditional practices.

Accordingly, AI and its applications in education are advancing significantly in response to teaching, learning, and assessment from elementary to higher education levels. On the other hand, its influence on how graduates develop the competencies necessary for survival and success in the future professional world is debatable. These debates have substantially intensified with the recent introduction of ChatGPT – a conversational language model or a chatbot, another forward step of AI. Universities worldwide are concentrating more and more on ChatGPT's potential to change future higher education teaching and learning practices (Lim et al., 2023). There are many possibilities for this technological advancement to improve and change the way we interact with technology in the context of higher education. In light of the claim that there are more opportunities than problems with using ChatGPT in evaluations for higher education. Using AI-powered technology to enhance student learning, encourage innovation, and deliver personalised feedback while addressing possible ethical and dependability problems presents an opportunity to reimagine the assessment process radically. The current study reviews the usage of ChatGPT in higher education assessments and:

- Investigates why students want to cheat using artificial intelligence capabilities;
- Identifies the challenges associated with detecting AI-generated content and its impact on academic integrity; and
- Identifies the need for re-designing assessments in parallel to technological advancement.

Assessment and artificial intelligence

This section includes how learning and teaching in education connect with artificial intelligence and ChatGPT.

Artificial Intelligence

AI is defined as the ability of machines and systems to acquire and apply knowledge and to carry out intelligent behaviour (OECD, 2016; UNCTAD, 2017). However, in the recent past, it seems to get closer to the capacity of human intelligence. As a result, AI is now being used in every field, including education, to perform human-oriented work: automation, personalisation, prediction, optimisation, decision-making, robotics, natural language processing and translation, and visual and voice recognition. Moreover, it comes in several ways, with computer programs, software, and embedded control systems in equipment and robots (Jantakun et al., 2021).

There is no universally agreed definition of AI (ESCAPE, n.d.), and several definitions are found in different academic literature, varying more or less from each other depending on the context. Broadly, AI has been defined by considering four dimensions: Thinking humanly, reasoning, acting humanly, and acting rationally (UNESCO, 2019). AI has played a large role in digital transformation and is felt globally. As an emerging novel technology, the integration of AI into education (Artificial Intelligence in Education – AIEd) arose as an interdisciplinary subfield in the early 1980s (Baker, 2021). It has opened up new paths and started modifying educational tools and institutions (Kengam, 2020) with the advancements in the 1980s and 1990s. Further, new potentials for learning technologies in several areas have been identified (Baker, 2021). AI is a continually evolving field, and understanding the potential impact of these changes and advancements on future teaching and learning will bring education a long way forward.

Artificial intelligence and education (AIEd) are driven in many ways, from the classroom to the entire school administration system. Not only in teaching, AI is used to cover non-teaching aspects at the school level. For example, several independent and public schools in England use an AI tool to forecast eating disorders, drug usage, and self-harm (Kengam, 2020). Baker & Smith (2019) divided and described AIED implementations into three groups: learner-oriented AIED, instructor-oriented AIED, and institutional system-oriented AIED (Baker & Smith, 2019). In the early stages of its growth, AI in education was mostly connected with intelligent tutoring systems. Subsequently, it demonstrated greater efficiency in helping learners identify knowledge gaps and personalised support (Jantakun et al., 2021) and the gaps in teaching (Kengam, 2020). Concerning teaching from the teachers' point of view, AI reduces the teaching workload, improves information literacy, and is helpful to their professional development (Xue & Wang, 2022).

AI primarily uses advanced analytics, deep learning, and machine learning to track how quickly one person moves relative to others. Today, AIED systems have a variety of functions for recognising the learner's traits as well as a variety of ways to engage and react to learners (Baker, 2021) to simulate and predict learning processes (Rienties et al., 2020). For example, in personalised learning, students experience a unique educational approach that meets their needs and skills completely. AI analyses student performance data and makes recommendations and suggestions based on a student's strengths and weaknesses. This would help students to reach their full potential and reduce drop-out rates.

Among the novel trends for using AI-enabled technology in educational assessment (formative and summative), contexts are increasing: Automated Essay Scoring (AES) and Computerised Adaptive Tests (CAT) (Gardner et al., 2020). Moreover, AI, combined with human invigilators for educational assessment, is evolving to ensure efficient supervisory methods at the examinations. The e-Assessment Association (eAA) states that AI complements live proctors and security agents (eAA, n.d). Further, they specify that AI can be used in face recognition and behaviour assessments of test candidates to ensure their true identity. However, Popenici et al. (2023) cast some doubt on employing AI for facial recognition in exam invigilation. AI-powered facial recognition algorithms developed in China seemed to function better than the software developed in the US, as US facial recognition is poor in identifying people of colour. Such racial biases of the algorithms might lead to socially discriminatory practices, negatively affecting students with darker skin colour (Popenici et al., 2023).

Besides providing highly advanced answers to questions, chatbots help assess written responses in high-stakes selection processes such as university entry and employment tests and analyse large-scale assessment-related datasets (Gardner et al., 2020). In general, the opportunities AI presents to education are vast, particularly for tutoring, assessment, and personalisation of instruction (González-Calatayud et al., 2021). Therefore, a better understanding of the educational context and the potential of technology within education is needed to get the most out of AI in education.

ChatGPT

ChatGPT stands for "Chat Generative Pre-Trained Transformer" (CTDL, 2023), and in general, it has been identified as a large language model (LLM) and a natural language processing (NLP) tool (Taecharungroj, 2023) capable of producing replies to text-based chat inputs or prompts. In other words, it can be defined as a dialogue-based AI chatbot model (Atuhaire, 2022) or a conversational AI agent capable of understanding and generating human-like written texts (Adiguzel et al., 2023). According to Hack and Knight (2023), ChatGPT is the first self-taught text generation programme that can learn and adapt to the writing style of its users. Further, "ChatGPT is a chatbot based on artificial intelligence developed by the OpenAI consortium" (Neefe & Vogelaar, 2023, p. 1) in San Francisco "that uses a technique called transformer architecture replies [that] quickly with well-formulated responses to a given prompt or context" (CTDL, 2023) by the user. It is designed to simulate human conversation and provide relevant responses to the user's input. As acknowledged, it can be used for various applications, including customer service, personal assistants, and chatbots since the launch of ChatGPT-3.5 on November 30, 2022 (Atuhaire, 2022; Hack & Knight, 2023) has invaded the internet, particularly on social media (Atuhaire, 2022). Subsequently, it was followed by an updated version, GPT-4, in March 2023 (Ifelebuegu, 2023), and it is expected to continue to evolve further with many advanced features.

As Susnjak (2022) declares, the recent release of ChatGPT has marked a significant leap in AI competencies in natural language processing, reasoning, and providing information virtually. Further, it is capable of generating the most accurate answers to difficult questions and requires the use of higher cognitive skills. As a result, many educational institutions recognise the potential of chatbots to enhance the overall student experience (Hack & Knight, 2023). Consequently, the scholarly community has started investigating ways of integrating ChatGPT into their pedagogical approach, enhancing student involvement and educational experiences (Rasul et al., 2023).

However, it should be emphasised that ChatGPT is not a replacement for critical thinking, creativity, and human interaction. Similarly, Limna et al. (2023) viewed ChatGPT as an adjunct but not a substitute for human interaction and students' achieving their academic goals. According to Rasul et al. (2023), ChatGPT has the potential to enhance student productivity through various means, including offering valuable information and resources, guiding students in building upon their existing knowledge and experiences to create new insights, enhancing language proficiency, fostering collaboration, increasing time efficiency and effectiveness, and offering assistance and motivation. According to CTDL (2023), educators should allow students to use ChatGPT to support their learning. However, they must be warned not to rely on it mindlessly because the answers given by the AI tools are based on widely accessible material. Although they appear plausible, they may not always be accurate and factual errors can be expected. Lieberman (2023) states that some of the references it generates are outdated since it cannot produce information based on events after its most recent internet scan in 2021. Most critically, excessive dependency on these tools will hinder the development of key academic and professional skills. That applies to both students and teachers in the educational context. However, ChatGPT's ability to accomplish complex academic tasks has caused mixed feelings among educators (Baidoo-Anu & Owusu Ansah, 2023), and some think of it as a disruptive technology that poses various ethical challenges (Firat, 2023).

Although ChatGPT is the dominant chatbot, numerous other emerging chatbots and AI platforms exist. Bard is a cloud-based conversational AI platform powered by LaMBDA, which is Google's collection of conversational neural language models. Microsoft's Bing Chat, Claude, Wit.ai, Hunyuan Aid, OPT by Meta, Alexa TM by Amazon, GPT-J and GPT-NeoX by Eleuther AI and Megatron-Turing NLG by NVIDIA and Microsoft are examples of AI-powered conversational systems and chatbots (Aydin & Karaaslan, 2023; Rudolph et al., 2023b). In China, Ernie 3.0 and Ernie-VLG are among the generative AI chatbots (Rudolph et al., 2023b).

Will (2023) reveals several plus points related to ChatGPT: it can be used to plan lessons, find resources to supplement lessons, formulate summaries or reports, and grade students' work. With the aid of AI-enabled tools, teachers will have more time to dedicate to being involved in teaching or doing research (Rouhiainen, 2019). It is a powerful time saver (Will, 2023). Then again, there is a threat to the ethical contract between the teacher and the students if both use ChatGPT to formulate questions and answers. Overall, the function of ChatGPT in education appeared to be fascinating in terms of educational improvements. However, many features are of concern regarding educational assessments.

Assessing student work

Assessing and evaluating is a continuous process based on collecting information about the student's learning experience and improvement. It is the strongest evidence of what and how students learn, what they know, what they can do (Dikli, 2003), and what and how teachers teach. As Dikli (2003) and Hooda et al. (2022) elaborate, assessment feedback directs policymakers and administrators for better curriculum design, as they get a better insight into monitoring the effectiveness of existing ones. Also, timely feedback assures quality control, certification, and selection of the education system. Especially in higher education contexts, assessment and feedback are important as they improve students' grading capacity, motivation values, and academic performance, advancing learning (Dikli, 2003) and fostering learning (Hooda et al., 2022). Henceforth, assessment tasks must be fair, transparent, and authentic to recognise that students have achieved the course expectations and are eligible for appropriate recognition.

Assessment takes several forms according to the purpose of the learning environment, and the course aims and objectives often enclosed within the Standard Assessment Paradigm – SAP (Swiecki et al., 2022). Hooda et al. (2022) say it could be diagnostic, formative, summative, e-assessment, self-assessment and peer assessment. Traditionally, formative and summative assessments are practised through SAP assessing techniques like multiple-choice questions, true-false tests, essays, and short answer questions to infer student knowledge and learning (Dikli, 2003; Swiecki et al., 2022). Although traditional assessing techniques are widely used, several potential problems have been recognised: Assessments in the standard paradigm can be onerous, discrete, uniform, inauthentic, and often antiquated (Hooda et al., 2022). Despite the drawbacks of the traditional methods, some believe they are more effective, while others think alternative assessment tools are superior (Dikli, 2003). To that end, authentic, performance-based, and constructivist assessments are listed as alternative assessment tools, and online learning settings must be considered (Reeves, 2000). Simonson et al. (2000) suggest cognitive, performance, and portfolio assessments as alternative categories. As emerged from the literature, the application of AI in educational assessments can range from the designing stage to the evaluation stage of the conventional assessment process, partly or entirely, to increase the efficiency and feasibility of maintaining assessment techniques.

Methodology

Our article's critical review exclusively depends on online databases of published work related to AI and ChatGPT. The study mainly focuses on answering the specific research questions formulated at the beginning of the study. Since ChatGPT is a new phenomenon, appropriate sources discussing the topic are limited. Hence, to ensure the number and quality of articles were reasonable, the search was extended to peer-reviewed journal articles, conference papers, reports in full text, and blog posts with authorship. Regarding the year of publication, materials that appeared from 2000 were considered. For resources relevant to ChatGPT, the period was picked from late 2022 to 2023. Text titles, abstracts, and whole texts that appeared in scholarly and multi-disciplinary databases such as Google Scholar, JSTOR, Taylor and Francis Online, Elsevier, and recognised official websites of institutions were examined to select suitable sources for the study. Moreover, both empirical and theoretical studies were followed in the selection process (since the number of studies on ChatGPT is limited).

A range of key terms and phrases were used to review the sources. They include mainly "Artificial Intelligence or AI", "ChatGPT", "Assessment and ChatGPT", "Artificial Intelligence and academic cheating", "ChatGPT and Academic Cheating and Higher education", "AI or ChatGPT and detecting cheating", and "AI or Artificial intelligence or ChatGP and reasons/factors for academic cheating". Further, the reference sections of the found sources were also searched for more relevant texts. The search yielded 378 articles, and a detailed examination of the titles, abstracts, content and duplicates was done. That resulted in the removal of 314 articles which failed to meet the criteria for inclusion. As a result, 64 research articles and sources (institutional web pages and Blog posts) were eventually included in this study.

The selected resources were studied compressively for the data and organised under the following four sections, taking the research questions as the themes to arrive at conclusions:

- The usage of AI and ChatGPT in higher education assessments
- Why do students want to cheat using artificial intelligence capabilities?
- The challenges associated with detecting ChatGPT-generated content and its impact on academic integrity
- Re-designing assessments in parallel to the advancement in technology

Table 1. Inclusion and exclusion criteria.

Type of Criterion	Criteria	Inclusion	Exclusion
Type of publication	Books		*
	Blogs	*	
	Conference papers	*	
	Dissertations		*
	Institutional web pages	*	
	Journal articles Reports	*	
Access	Online	*	
	Paper		*
Publication period	2000-2023	*	
Place of publication	World wide	*	
Type of study	Empirical studies	*	
	Theoretical studies	*	
Research methods	Quantitative	*	
	Qualitative	*	
	Mixed	*	

Results

The usage of AI and ChatGPT in higher education assessments

Due to its distinctive features, ChatGPT has garnered much attention and inspired controversy ever since its release. Many scholars foresee ChatGPT to become as ubiquitous as Wikipedia and calculators (Hack, 2023), with its functions of language translation, content generation, and language modelling, generating summaries, articles, stories (Cotton et al., 2023), and many other types of texts. Furthermore, it increases student engagement and collaboration, provides a platform for asynchronous communication, and enables remote learning (Cotton et al., 2023). Even though the app is user-friendly, the content created using it is difficult to discriminate from text written by humans (Elkins & Chun, 2020), and it is questionable with respect to academic integrity. According to Sullivan et al. (2023), ChatGPT has raised both academic integrity concerns and the potential for enhanced learning in higher education.

One of the best examples of the application of AI in the early days is receiving fast feedback on the students' work in higher education, as reported in Mirchi et al.'s (2020) study on simulation-based training in medicine. A Virtual Operative Assistant was used in this study to give automatic feedback to students based on performance metrics. Another advancement of AI in assessment is vision-based AI, in which optical systems are used to grade students' work (Jimenez & Boser, 2021). AI's capacity to provide personalised feedback with quantitative and qualitative data has been proven by the StuDiAsE (Student Diagnosis, Assistance, Evaluation) System based on AI (Samarakou et al., 2016). Grading assignments and providing feedback to students in real time makes learning more efficient and personalised (Cotton et al., 2023). The development of intelligent software to select questions for online exams (Janpl & Piriyasurawong, 2020) is another instance where AI interferes with students' assessments. Nowadays, AI-powered software is used to grade exams and students' assignments automatically. It reduces the workload on teachers while providing students with immediate feedback. Paper checkers provide accurate grading of student papers without wasting time. Collectively, the incorporation of AI with educational assessments results from intelligent tutoring, testing through games, and virtual reality to AI-built mini-tests afford a wide variety of techniques to evaluate the strengths and weaknesses of students and teachers together.

Regarding the usage of ChatGPT in students' assessments, a comprehensive analysis of ChatGPT and its possible effects on conventional assessments in higher education was conducted by Rudolph et al. (2023a). As it reports, the impact of ChatGPT on essay-type written assessments, ChatGPT's inability to understand what is being shared and the relevance or accuracy of the information are the highest concerns of academics. Being an AI language model, ChatGPT is capable of producing or summing up texts, developing assignments, supporting essay writing, providing the most suitable responses to questions, and writing computer codes (Sullivan et al., 2023; Cotton et al., 2023; Crawford et al., 2023). Further, it can assist academic writing by extracting key points (Aljanabi et al., 2023) and can carry out some other tasks that often appear in assignments. The use of interactive, game-based assessments in higher education (Cotton et al., 2023) is also based on ChatGPT, and it is termed "stealth assessments: evidence-based assessment that is woven directly and invisibly into the fabric of the learning or gaming environment" (Shute, 2015). While ChatGPT creates challenges for educators, it negatively impacts students: it hampers students' learning ability by producing incorrect answers to prompts. It makes new learners struggle to differentiate between accurate and inaccurate information (Wood et al., 2023).

Other than providing plenty of accessible information and the opportunity to peruse new critical information to reproduce existing knowledge, ChatGPT helps to improve grammar and writing structure (Sullivan et al., 2023), especially when it comes to a learner who learns in a second language (Hong, 2023), preferably English—confirming the fact Aljanabi et al. (2023) mention that "there is no doubt that ChatGPT can be used to maintain the quality of academic work by using feedback on grammar and coherence". Although ChatGPT has advantages for idea generation, it is weaker in literature synthesis (Dowling & Lucey, 2023). Other than that, improving the students' desire for learning, establishing the basics of knowledge, and developing a deep understanding of the subject are also considered possible with ChatGPT (Hardman, 2023a; Crawford et al., 2023).

Overall, at the very basic level, students can use ChatGPT-generated responses as the starting point of the answer or use it as a guide to build up a well-structured, grammatically correct completed answer enhanced with their knowledge and ideas. Moreover, despite the students' identification of ChatGPT as a powerful text generator, it is vital to note the limitations: inability to provide accurate citations (Cooper, 2023), use technical terms appropriately, or develop evidence-based arguments can result in a superficial overview of a topic, which may compromise the overall quality of the assignment (Hack, 2023). In this background, students' skill development and the accuracy of the assessment results have become questionable and challenging in an environment where students are familiar with and using novel technological advancements such as ChatGPT. Thus, close monitoring of students' work with appropriate guidance should be there to reap the best of ChatGPT in education.

Why do students want to cheat using artificial intelligence capabilities?

ChatGPT is making a revolutionary advancement in conversational AI and has quickly established its position in academia. With many user-friendly advanced features, it provides credible service to students and academics. Although limited experimental evidence has been reported for how ChatGPT is perceived by students (Strzelecki, 2023) and why students use AI for academic cheating purposes, the most typical answer can be due to its progressive functions as a natural language processing model.

Despite the efforts taken by educational institutions to restrict plagiarism, it is still a highly concerning issue in academia. Students always get the assistance of the internet and other virtually available sources to complete their assessments, exams, and other academic work. Ease of accessing information is one of the closest reasons for students' tendency to use AI in their work. Agreeing with the same fact (King, 2023) mentions that online platforms are abundant with effortlessly reachable information, and students can easily copy and paste from the sources. Further, the students can get the output from the AI tool quickly and are accurate to the expected level. The efficiency of AI in terms of accuracy and quickness to give the final product is another factor that makes students look for assistance. Hence, students have identified AI as a time saver and let them obtain good grades (Haun, 2022). Particularly, some students struggle to cope with their coursework, assignments, essays, and exams within a short time. Hence, inadequate preparation for the evaluation and poor time management skills also tempt students to misbehave (Dehouche, 2021).

As Crawford et al. (2023) elaborate, ChatGPT does not predict right or wrong, but using the given prompt directly generates the output, saving students time. Research evidence from Dehouche (2021) and, most recently, CTDL's (2023) findings confirm that increased pressure created on students and competition with peers are other motives for using ChatGPT or AI output in their work. Further, CTDL (2023) has found that the increasingly competitive academic environment has made students feel that they need to score high to secure their professional positions. Another possible reason to use AI-powered tools in assessments may be the difficulty level of the assessment. Recent study findings of Strzelecki (2023) suggest that students are more likely to adopt functional technologies like ChatGPT when they have high levels of "performance expectancy". Utilising ChatGPT enhances the likelihood of completing significant academic activities, speeds up the completion of assignments and projects, and boosts productivity since students see ChatGPT as beneficial to their academic endeavours.

Generally, cheating is one of the focal concerns at all levels of academia. Advancements in AI have made it easier than ever before to cheat. Students are involved in cheating purposely or by chance due to several internal and external reasons: individual, situational, and institutional factors, specifically including competition, stress, poor time and resource management, poor academic background, and to achieve good grades (Sullivan et al., 2023). On the other hand, ChatGPT's abilities as a powerful tool for producing responses and engaging in conversations also motivate students to cheat. According to the perceptions of the university community, banning ChatGPT is too hard as many students are already using it, and blocking it on the university network will prompt students to use a VPN (Sullivan et al., 2023). Hence, rather than generalising the perception that ChatGPT is a means of academic cheating, it would be more valuable if it could be viewed as a means for improving teaching and learning. Moreover, it requires further research on students' motives to use ChatGPT in their academic work.

The challenges associated with detecting ChatGPT-generated content and its impact on academic integrity

The transformation of traditional academic practices into hybrid or digital platforms raised great concern about the potential for academic misconduct. This has been heightened with the introduction of ChatGPT, as it potentially offers many applications for higher education activities (Cotton et al., 2023). As Sullivan et al. (2023) highlight, maintaining academic integrity is a significant challenge when using ChatGPT for academic work like assessments, dissertations, and papers. As a result, academic integrity concerns are more frequently discussed than opportunities to integrate with academic work.

Plagiarism is one of the common issues attached to written texts. For example, students could use the essay-writing systems to cheat on their assignments by submitting someone else's essay (Dehouche, 2021) since AI essay-writing systems produce essays grounded on a set of limits or prompts (Cotton et al., 2023). Remarkably, that raises thoughtful questions about the authenticity of student work, especially at the stage of students' grading (Hack & Knight, 2023). CTDL (2023) verifies the claim even more and identifies ChatGPT as a motivation for plagiarism and a threat to academic integrity. On the other hand, it could ultimately devalue the earned qualification as the evaluators do not see the real skills and abilities of the student through the written work. Another challenge that comes hand in hand with ChatGPT-generated text is its ability to generate high-quality written work. Susnjak (2022) produces experimental evidence of AI capabilities, and the study results show that ChatGPT can perform high-order thinking tasks to produce text identical to human writing. This feature could be used for academic dishonesty in online examinations or assessments.

Availability and accessibility to the technological facilities are not equally distributed even among the students in the same academic group. This gives students who utilise these tools an unfair edge since they can produce higher-quality work than the rest, leading to an unfair evaluation process (Bagshaw, 2022). Similarly, students may refine the AI-generated answer several times to make it better before the submission (Cotton et al., 2023), and then again, the marker grades forged answers, which results in an inaccurate evaluation of response quality (Limna et al., 2023). As the academic staff cannot accurately judge the student's understanding of the subject, that may impact re-designing the course work.

Concerning the research community, as Bianchi (2023) states, researchers and students may submit material created using Large Language Models (LLM) as their own or may employ LLMs carelessly and generate incorrect results. As CTDL (2023) notes, ChatGPT sometimes writes believable but inaccurate or illogical responses and fixing it is not easy due to the nature of the application structure. As a consequence, deceptive knowledge sharing may occur. When the graduates work on their own in real situations as a part of their professional work, they will probably end up with a failure. That will break the public's trust towards the academic qualifications, academics, and the institution.

There is no question that submitting the raw or refined output from ChatGPT constitutes academic dishonesty, and spotting such actions is one of the main concerns in the academic community. Even though there are tools to identify plagiarism, recognising AI-generated content is still an unsolved problem (Bianchi, 2023; Hadadgar & Maunder, 2023). According to Hadadgar & Maunder (2023), detection techniques explicitly developed to recognise ChatGPT-generated content should have a high possibility of being successful. As systems like ChatGPT grow more precise and accurate with each iteration, it may become difficult to identify created content. Experimental results of Khalil and Er (2023) to determine whether plagiarism detection tools could detect essays written using ChatGPT show that out of the 50 essays tested, 40 had a similarity score of 20% or less, demonstrating a high degree of originality. In addition to that, Chaka (2023) studied the accuracy of five AI content detective tools: GPTZero, OpenAI Text Classifier, Writer.com's AI Content Detector, Copyleaks AI Content Detector, and Giant Language Model Test Room, in recognising content generated by ChatGPT, YouChat, and Chatsonic. Chaka (2023) shows that the ability to precisely and persuasively distinguish machine-generated texts from AI-generated literature in different contexts appears to be a limitation commonly shared by all five AI content detectors. However, Copyleaks AI Content Detector was the top-performing AI content detector among the five AI content detectors used for the study.

At present, finding the best method to respond to emerging AI tools is one of the main concerns of many academic institutions. In the midst of that, Bianchi (2023) argues that the unrefined output of ChatGPT is detectable on careful inspection. Giving examples, Cotton et al. (2023) mention several approaches to detect ChatGPT-assisted written texts: examining for patterns or deviations in the language or words, looking for sources and citations, checking for uniqueness and novelty, and checking for language errors like spelling and grammar. Nevertheless, Hassoulas et al.'s (2023) study revealed that experienced markers cannot consistently differentiate between student-written scripts and text generated by natural language processing tools, such as ChatGPT. Shedding light on the recent research findings, Limna et al. (2023) suggest investigating the safe and effective adoption of chatbots, particularly ChatGPT, in education rather than banning or restricting them.

Re-designing assessments in parallel to the advancement in technology

Addressing the issues created by ChatGPT regarding assessments and evaluations is not easy due to its multifunctioning features if educators neglect to rethink their assessment strategies. Mills et al. (2023) discuss the same issue, highlighting the need to rethink assessment as the generative AI poses a grave threat to academics since it appears to co-opt the assessment methods essential to their instruction. Further, they focus on Bali's (2023) idea of moving to a culture of 'transparent assessment' and designing an assessment that truly makes students want to learn.

Considering the assessment types, Hadadgar and Maunder (2023) state that written assessments, essays, short answer questions, completion questions, and dissertations are the most affected forms of evaluation form by ChatGPT. On the other hand, MCQs, matching questions, observations, performance records, peer/self-assessments, and portfolios have been identified as entirely resistant to the impacts of ChatGPT. In contrast, Cassidy (2023) suggests physical closed-book exams where the students answer using only pen and paper as one of the possible strategies to address the issue.

As a timely solution for the appearing issues, there lies an opportunity for academics to change the design of their assessment format. Hardmann (2023a) and Rudolph et al. (2023a, 2023b) propose overstepping the traditional methods with innovative ones, and some authors find this an opportunity to re-design assessing methods. As a key strategy, assignments can be designed in a way that students are required to reflect on their skills in critical thinking, problem-solving, communication, and collaboration (Rudolph et al., 2023a, 2023b). Increasing the chances for collaborative activities, such as engaging in group discussions, presentations, or other interactive activities) will prevent or minimise the use of ChatGPT by students (Cotton et al., 2023; Rudolph et al., 2023b. This can make it more difficult for students to use ChatGPT or other AI language models to complete their assignments and can promote critical thinking and independent learning. In addition, asking students to provide feedback or a personal elaboration about their assignment completion and a list of references may also help control the use of AI tools. However, studies have recognised three main limitations of GPT-3.5: the inability to answer semantic, factual, and ethical questions (Illingworth, 2023). Taking advantage of this, academics can prepare the questions or assignments accordingly, and students can be asked to justify their answers.

Another approach to be used is to design assessments with open-ended questions where students should develop and defend arguments on their own (Cotton et al., 2023). Furthermore, creating questions that involve contextual and real-world problems would probably limit the influence of AI tools on the answer. Hack (2023) and Rudolph et al. (2023b) point out authentic assessments as an opportunity for students to demonstrate the skills and knowledge required to work. Applying various assessment methods, like oral or live demonstrations (Susnjak, 2022; Hardman, 2023a; CTDL, 2023), and analysis of images and videos of longer texts that do not fit in a prompt (Rudolph et al., 2023a) will encourage students to produce authentic outcomes. Hack (2023) describes the range in which students use AI in their academic work and the intensity of academic dishonesty and authentic learning.

In this context, students can assemble their answers by incorporating outputs generated from multiple prompts based on the questions. Alternatively, they can submit the chatbot's output directly, with or without any modifications. Students may also employ AI to obtain and enhance an outline through their input, leading to a more authentic learning experience. This underscores the significance of human input in ensuring accuracy and comprehensiveness in academic work. Additionally, maintaining fairness in the assessment process is of paramount importance.

Rasul et al. (2023) propose an additional advanced step to authentic assessment, giving students autonomy and agency to answer the questions in their own way rather than forcing them to write the same essay or respond to the same question. In doing so, assessments are anticipated to become more attractive, comprehensive, and, in the end, authentic. However, it raises another issue about the assessing task's standards and uniformity. As another alternative, the authors suggest providing ChatGPT's responses to the question together with some marking guidelines. Then students can be asked to comment about and/or reason out the grade the automated response deserves. Continuing the discussion, Rasul et al. (2023) recommend ChatGPT-generated scenario-based tasks that involve analysing and solving problems they may encounter in their future careers. Additionally, Ifelebuegu's (2023) study emphasises the importance of re-designing tests and assessments to place a larger priority on higher-order skills. They favour authentic assessment methods such as open-ended activities, project-based assignments, collaborative assessments, and portfolio-based evaluations since they make it difficult for AI chatbots to replicate.

Hence, it is noteworthy that giving authentic assessments provides students with much space to use ChatGPT responsibly, and it is the best way to combat the threats to academic integrity. Also, it safeguards the quality of academic programmes while providing ample opportunities for the ethical and responsible use of AI.

Discussion and conclusion

As a whole, the study results encourage the integration of AI in education. They highlight the requirement for the collaborative effort of education providers and policymakers to go for innovative and diplomatic strategies to use AI-based tools in academic work. Since AI has been identified as an education enhancer and a valuable educational tool (Limna et al., 2023) by educators, policymakers and other stakeholders in the education context, it should be invited and encouraged. According to the outcomes of this study, ChatGPT can give more precise and accurate answers to questions, write abstracts, summarise text, and perform many functions related to academia. In contrast, Wood et al. (2023) argue that some question types are less likely to be correctly answered by ChatGPT, and thus, focusing on such question types could reduce the risk of cheating with ChatGPT.

Furthermore, many researchers emphasise the pedagogical integration of ChatGPT. Designing curricular and pedagogical methods that better use the advanced features of ChatGPT is one of their major concerns. In a way, ChatGPT is a time saver since it provides solutions to more complicated issues within a few seconds, which requires the involvement of higher-order intellectual skills. Moreover, due to the user-friendly features of ChatGPT, it is widely used in academic work, including the assessment process. Henceforth, it has become a thought-provoking problem regarding the students' authentic academic performances, as there is room for academic malpractice using ChatGPT. Simultaneously, it alarms about quality assurance of the qualification, employability of the graduates, and the skills required to succeed in an AI-dominated world. ChatGPT's capabilities, limitations, and impact on students' academic performances are among the prominently discussed themes related to ChatGPT. Accordingly, continuous and collaborative discussion among educational policymakers and stakeholders is required to develop policies and guidelines to ensure the ethical use of ChatGPT.

Although ChatGPT impacts academics and students equally, the current literature primarily focuses on academic teachers' and scientists' views on ChatGPT and its future, and students' perceptions as crucial stakeholders have not been clearly highlighted (Strzelecki, 2023). As Lieberman (2023) points out, some educators already consider integrating the app into education, while others are worried about how it may affect their pupils' drive to study. Bagshaw (2022) also stands on the same ground and states that some academics are alarmed by the function of ChatGPT. Conversely, the author refutes the claim by pointing out that the other side may benefit from its strength. As emerged in the previous discussions, academics are aided with generating lesson plans, test questions, quizzes, and rubrics to grade students with ChatGPT. However, clearly, it is an opportunity to promote graduate skills. On the side of the students, ChatGPT can be identified as a conversational and interactive tool in which students can readily find answers to their assignment questions and get the outline for essays. Hence, it is the responsibility and the duty of educational institutions and stakeholders to prepare students for the tech-driven future and let them experience the benefits of technological advancements. Further, more research on students' understanding and experiences of ChatGPT in their higher education, the benefits they have and most interestingly, why they use it and when it is needed.

Making students familiar with novel technologies is extremely important, but still, they should be aware of the ethics of using them. This is further supported by Hack (2023) and highlights the crucial need for students to learn how to engage with AI to get the advantage it brings while knowing its limitations and threats. Moreover, Cotton et al. (2023) suggest that conveying guidelines on using AI tools for academic purposes and informing about proper citation and acknowledgement of ChatGPT-generated text is necessary. In addition, getting a written declaration mentioning that they are responsible for the consequences of academic misbehaviour is also a reasonable measure. Despite applying sophisticated detective methods, teachers can react to unethical ChatGPT usage by punishing students and giving in-class examinations.

Additionally, re-designing assessments and evaluation methods in such a way as to limit the usage of AI and make room for employing students' intrinsic skills in the assessment completion process is critical. In that sense, designing higher-order cognitive assessments focusing on analysis, synthesis, evaluation, and creation effectively counteract ChatGPT (Ifelebuegu, 2023). Bianchi (2023) proposes to use tools to detect AI-generated content introduced by OpenAI: AI Text Classifier, GPT-2 Output Detector, and a classifier based on RoBERTa. Simply educating students about plagiarism and its consequences at any stage of coursework is a measurable act to minimise the unethical use of ChatGPT. Accordingly, as revealed through the study, the detection and prevention of unethical use of ChatGPT and the encouragement of ethical use of AI are other dimensions on which academics and policymakers focus. However, allowing students to utilise

ChatGPT and other AI tools per a pre-established set of guidelines and within specified parameters will be far more appropriate since it will provide students with the practice of using cutting-edge educational technologies ethically.

AI is still being developed and appears to act similarly to humans in many fields. However, combining AI with human capabilities such as creativity, cognitive skills, and social-emotional skills would bring the best outcome. Hence, despite ChatGPT's disadvantages, the pros must be strengthened while the cons are addressed strategically. Updating academic integrity policies, providing training on AI tools such as ChatGPT and academic integrity and encouraging research on AI tools' effect on higher education are equally important (Rudolph et al., 2023b). Scholars can research unexplored elements such as specific pedagogical approaches to maximise the benefits of ChatGPT, strategies addressing its limitations, or the impact of ChatGPT on different subject areas, student assessments and higher education productivity.

References

Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionising education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology, 15*(3), 429. https://doi.org/10.30935/cedtech/13152

Aljanabi, M., & ChatGPT. (2023). ChatGPT: Future directions and open possibilities. *Iraqi Journal for Computer Science and Mathematics*, 16–17. https://doi.org/10.58496/mjcs/2023/003

Atuhaire, R. (2022). *What is ChatGPT?* https://www.dignited.com/104384/what-is-chatgpt-and-how-does-it-work

Aydin, Ö., & Karaarslan, E. (2023). Is ChatGPT leading generative AI? What is beyond expectations? *Academic Platform Journal of Engineering and Smart Systems, 11*(3), 118-134. DOI: 10.21541/apjess.1293702

Baidoo-Anu, D. & Owusu Ansah, L. (2023). *Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning*. http://dx.doi.org/10.2139/ssrn.4337484

Bagshaw, J. (2022). *What implications does ChatGPT have for assessment?* https://wonkhe.com/blogs/what-implications-does-chatgpt-have-for-assessment

Baker, R. S. (2021). Artificial intelligence in education: Bringing it all together. *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots.* https://www.oecdilibrary.org/sites/f54ea644en/index.html?itemId=/content/component/f54ea644-en

Baker, T., & Smith, L. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges.* https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB. pdf

Bianchi, E. (2023). *ChatGPT unmasked: How AI-generated content can be detected.* https://pub.towardsai.net/chatgpt-unmasked-how-ai-generated-content-can-be-detected-259cbc79beef

Chaka, C. (2023). Detecting AI content in responses generated by ChatGPT, YouChat, and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching, 6*(2), 1-11. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.12

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 60*, 1-12. doi: 10.1080/14703297.2023.2190148

Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice, 2*0(3), 02. https://doi.org/10.53761/1.20.3.02

Dehouche, N. (2021). Plagiarism in the age of massive generative pre-trained transformers (GPT-3). *Ethics in Science and Environmental Politics, 2,* 17–23. https://doi.org/10.3354/esep00195

Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *Turkish Online Journal of Educational Technology-TOJET, 2*(3), 13-19.

Dowling, M., & Lucey, B. (2023). ChatGPT for (finance) research: The Bananarama conjecture. *Finance Research Letters, 53,* 103662. https://www.sciencedirect.com/science/article/pii/S1544612323000363

Elkins, K., & Chun, J. (2020). Can GPT-3 pass a writer's turning test? *Journal of Cultural Analytics, 5*(2), 1-16. 10.22148/001c.17

Faucher, D., & Caves, S. (2009). Academic dishonesty: Innovative cheating techniques and the detection and prevention of them. *Teaching and Learning in Nursing, 4,* 37-41.

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning & Teaching, 6*(1), 57-63. https://doi.org/10.37074/jalt.2023.6.1.22

Gamage, K. A. A., Dehideniya, S. C. P., Xu, Z., & Tang, X. (2023). Contract cheating in higher education: Impacts on academic standards and quality. *Journal of Applied Learning & Teaching, 6*(2), 1-13. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.24

Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?'. *Journal of Computer Assisted Learning, 37*(5), 1207–1216. https://doi.org/10.1111/jcal.12577

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial Intelligence for student assessment: A systematic review. *Applied Sciences, 11,* 5467. https://doi.

org/10.3390/app11125467

Hack, K. (2023). *Excited, concerned and curious': Student perspectives on learning and working in the era of AI*. Advance HE. https://www.advance-he.ac.uk/news-and-views/excited-concerned-and-curious-student-perspectives-learning-and-working-era-ai

Hack, K., & Knight, C. (2023). *Higher education in the era of AI*. AdvanceHE. https://www.advance-he.ac.uk/news-and-views/higher-education-era-ai

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology, 15*(2), 421. https://doi.org/10.30935/cedtech/13036

Hardman, P. (2023). *Veto, Bypass or Flip?* https://drphilippahardman.substack.com/p/vetoing-bypassing-and-flipping

Hassoulas, A., Powell, N., Roberts, L., Umla-Runge, K., Gray, L., & Coffey, J. M. (2023). Investigating marker accuracy in differentiating between university scripts written by students and those produced using ChatGPT. *Journal of Applied Learning & Teaching, 6*(2), 1-7. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.13

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation, 5*(1), 37-45. https://jeti.thewsu.org/index.php/cieti/article/view/103

Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, S. (2022). Artificial Intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering* .https://doi.org/10.1155/2022/5215722

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2), 1-14. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.29

Jantakun, T., Jantakun, K., & Jantakoon, T. (2021). A common framework for Artificial Intelligence in Higher Education (AAI-HE Model). *International Education Studies, 14*(11), 94-103.

Janpla, S., & Piriyasurawong, P. (2020). The development of an intelligent multilevel item bank model for the national evaluation of undergraduates. *Universal Journal of Educational Research 8*(9), 4163-4172.

Jimenez, L., & Boser, U. (2021). *Future of testing in education: Artificial Intelligence.* https://www.americanprogress.org/article/future-testing-education-artificial-intelligence/

Kengam, J. (2020). *Artificial intelligence in education.* https://www.researchgate.net/publication/347448363_ARTIFICIAL_INTELLIGENCE_IN_EDUCATION

Keresztury, B., & Cser, L. (2013). New cheating methods in the electronic teaching era. *Procedia-Social and Behavioral Sciences, 93*, 1516-1520.

Khalil, M., & Er, E. (2023). *Will ChatGPT get you caught? Rethinking of plagiarism detection*. Arxiv. https://arxiv.org/pdf/2302.04335.pdf

King, M. R.(2023). ChatGPT. A conversation on Artificial Intelligence, chatbots, and plagiarism in higher education. *Cellular and Molecular Bioengineering, 16,* 1–2 (2023). https://doi.org/10.1007/s12195-022-00754-8

Lieberman, M. (2023). What is ChatGPT and how is it used in education? *Education Week*. https://www.edweek.org/technology/what-is-chatgpt-and-how-is-it-used-in-education/2023/01

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education, 21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching, 6*(1), 64-74. https://doi.org/10.37074/jalt.2023.6.1.32

Manning, C., (2020). *Artificial Intelligence definitions. Stanford University*. https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1), 16-30. https://doi.org/10.37074/jalt.2023.6.1.34

Mirchi, N., Bissonnette, V., Yilmaz, R., Ledwos, N., Winkler-Schwartz, A., & Del Maestro, R. F. (2020). The virtual operative assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. *PloS one, 15*(2), e0229596.

Moten Jr, J., Fitterer, A., Brazier, E., Leonard, J., & Brown, A. (2013). Examining online college cyber cheating methods and prevention measures. *Electronic Journal of E-learning, 11*(2), 139-146.

Neefe, R., & Vogelaar, K. (2023). *Use of ChatGPT in education and testing.* https://husite.nl/toetsing-nieuw/wp-content/uploads/sites/299/2023/01/ChatGPT-Guide.pdf

Odongo, A. O., Agyemang, E., & Forkuour, J. B. (2021). Innovative approaches to cheating: An exploration of examination cheating techniques among tertiary students. *Education Research International, 2021,* 1-7. https://doi.org/10.1155/2021/6639429

Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning & Teaching, 6*(2), 1-7. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.4

Popenici, S., Rudolph, J., Tan, S., & Tan, S. (2023). A critical perspective on generative AI and learning futures. An interview with Stefan Popenici. *Journal of Applied Learning and Teaching, 6*(2), 1-21. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.5

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56. https://doi.org/10.37074/jalt.2023.6.1.29

Reeves, T. C. (2000). Alternative assessment approaches for online learning environments in higher education. *Journal of Educational Computing Research, 23*(1), 101–111. https://doi.org/10.2190/GYMQ-78FA-WMTX-J06C

Rienties, B., Køhler Simonsen, H., & Herodotou, C. (2020, July). Defining the boundaries between artificial intelligence in education, computer-supported collaborative learning, educational data mining, and learning analytics: A need for coherence. In *Frontiers in education* (Vol. 5, p. 128). Frontiers Media SA.

Rouhiainen, L. (2019). *How AI and data could personalise higher education*. Business Education. Harvard Business Review. http://hbr.org/2019/10/how-ai-and-data-could-personalize-higher-education

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Samarakou, M., Fylladitakis, E. D., Karolidis, D., Fruh, W. G., Hatziapostolou, A., Athinaios, S., & Grigoriadou, M. (2016). Evaluation of an intelligent open learning system for engineering education. *Knowledge Management & E-Learning: An International Journal, 8,* 496-513.

Shute, V. (2015). *Stealth assessment in video games.* https://research.acer.edu.au/cgi/viewcontent.cgi?article=1264&context=research_conference

Simonson, M., Zvacek, S. M., & Smaldino, S. (2019). *Teaching and learning at a distance: Foundations of distance education* (7th edition). Prentice Hall.

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching, 6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Susnjak, T. (2022). *ChatGPT: The end of online exam integrity?* arXiv preprint arXiv:2212.09292. http://aixpaper.com/view/chatgpt_the_end_of_online_exam_integrity

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence, 3,* 100075. https://doi.org/10.1016/j.caeai.2022.100075

Taecharungroj, V. (2023). “What can ChatGPT do?” Analysing early reactions to the innovative AI chatbot on Twitter. *Big Data and Cognitive Computing, 7*(1), 35. https://doi.org/10.3390/bdcc7010035

The e-Assessment Assosiation. (n.d). *How AI is transforming the education and assessment industry*. https://www.e-assessment.com/news/how-ai-is-transforming-the-education-and-assessmentindustry/#:~:text=AI%20can%20transform%20candidates’%20learning,to%20impact%20the%20job%20market.

The Economic and Social Commission for Asia and the Pacific (ESCAP). (n.d). *Artificial Intelligence in Asia and the Pacific.* https://www.unescap.org/sites/default/files/ESCAP_Artificial_Intelligence.pdf

The Centre for Learning, Teaching, and Development (CTDL). (2023). *ChatGPT for Learning and Teaching*. University of the Witwatersrand, Johannesburg.

UNESCO. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development.* Education Sector, United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000366994/PDF/366994eng.pdf.multi

Wang, J., Tong, Y., Ling, M., Zhang, A., Hao, L., & Li, X. (2015). Analysis on test cheating and its solutions based on extenics and information technology. *Procedia Computer Science, 55,* 1009-1014.

Will, M. (2023). “With ChatGPT, teachers can plan lessons, write emails, and more. What’s the catch?”. *Education Week, 42*(20), 7.

OECD (2016). *OECD Science, Technology and Innovation outlook 2016.* Paris. https://read.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-innovation-outlook-2016_sti_in_outlook-2016-en#page1

UNCTAD. (2017), *Information economy report 2017.* http://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=1872.

Wood, D. A., Achhpilia, M. P., Adams, M. T., Aghazadeh, S., Akinyele, K., Akpan, M., ... & Kuruppu, C. (2023). The ChatGPT Artificial Intelligence chatbot: How well does it answer accounting assessment questions?. *Issues in Accounting Education,* 1-28.

Xue, Y., & Wang, Y. (2022). Artificial Intelligence for education and teaching. *Hindawi, Wireless Communications and Mobile Computing,* 4750018. https://doi.org/10.1155/2022/4750018

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Mentoring international postgraduate students and early career researchers through transnational telecollaboration: a supervisor's autoethnography

Julian Chen[A]

[A] School of Education, Curtin University, Perth, Western Australia

DOI: https://doi.org/10.37074/jalt.2023.6.2.23

Abstract

The high-calibre education in Australia has attracted overseas students to pursue a Higher Degree by Research (HDR) every year. While HDR training is crucial for all HDR students, international students are relatively vulnerable due to the challenging demands of academic writing and research, coupled with the cultural and language barriers different from their own. To worsen the situation, the global pandemic forced face-to-face supervision into remote supervision mode, thus exacerbating students' social-emotional learning state even further. This article transports readers to a telecollaborative project that I initiated amid the pandemic through an autoethnographic approach. Propelled by the urgent need to better support supervisees beyond boundaries, I enacted a transnational telecollaboration to mentor international HDR students to position themselves as emerging researchers. Informed by participatory action research (PAR), I guided my junior colleagues (early career researchers (ECRs)) to conduct HDR needs analysis, hold HDR training webinars, build a virtual community via Facebook, and shadow HDR students throughout their reflection journaling via Google Docs. This viable supervision model broke down the power structure by creating an ecologically balanced framework, thus promoting collaboration rather than isolation.

Keywords: Autoethnography; higher degree by research (HDR); online mentoring; pandemic; social-emotional learning; telecollaboration.

Background story

Higher Degree by Research (HDR) supervision and early career researcher (ECR) mentoring has always held a special place in my heart. As a transnational postgraduate student in Canada (M.Ed., 2001-2002) and the US (Ph.D., 2007-2012), I have been there, done that, and can fully relate to my supervisees during their HDR journey Down Under (in Australia). According to the Australian Bureau of Statistics, overseas students yielded approximately $37.4 billion to the Australian economy before the pandemic, a majority of which came from higher education ($25.4 billion) (Ferguson & Spinks, 2021). The Organization for Economic Co-operation and Development (OECD, 2021) also noted that overseas student enrolments represented at least 28% of tertiary students in Australian higher education. If we zoom in on the recent HDR enrolments, 48% of the international students enrolled in Master's degrees, followed by 32% in Ph.D. programs (Lee, 2019). Similar to other higher education providers in the Global North, the fact that Australian universities rely heavily on international students cannot go unnoticed due to its impact on Australia's workforce, education, and society.

Even though international HDR students contribute to national economies and revitalise cultural diversity in society, their social-emotional learning is less talked about publicly and usually sidelined by how well they can perform academically (Chen et al., 2023; Prieto et al., 2022). The elephant in the room is that international students often struggle to cope with not only intellectually challenging demands in academic writing and research but also barriers raised by the new culture and language different from their own. The solitary nature of the HDR journey also exacerbates their social-emotional state and wellbeing. Some HDR students, if not all, have experienced the so-called "PhD stress", evidenced in low self-esteem, constant fear of failure, and mental exhaustion, which negatively affects their work performance and leads to depression (Hayton, 2013). In fact, one of my HDR students sadly had to withdraw from his PhD studies and return to his home country just because the PhD stress had taken its toll on his mental health. Other HDR students also experienced a similar sense of "isolation" as anecdotally shared by my colleagues.

The situation was even worse when the COVID-19 pandemic turned the whole world upside down, forcing the higher education sector into emergency remote teaching (ERT) (Hodges et al., 2020). This unprecedented shift from the face-to-face (F2F) mode to ERT, coupled with social distancing and global border closures, hit universities particularly hard. Empirical studies have also reported on the intensified emotional rollercoaster impacting the mental health of the stakeholders due to the drastic shift to and unpreparedness for ERT (see, for example, Appel & Robbins, 2022; Chen, 2022; Kozhabayeva & Boivin, 2022; McAlinden & Dobinson, 2022). For example, one of my supervisees expressed how social distancing and border lockdown inflicted agony on his mental health and research plan amid the global outbreak of COVID-19 in 2020:

> When my life becomes smaller, just spending my time from my laptop, bed, and kitchen only, it is very frustrating for me… The pandemic also makes me worried about my health…I am afraid if I get infected and die while I am here in Australia so far away from my family… COVID-19 has also changed the schedule that I had planned before (HDR student, reflection journal, 18 May, 2020).

It broke my heart to know that the sense of isolation, stress, diminished self-worth, and challenges in conducting research remotely had negatively impacted our HDR students. There is definitely more to HDR training that we (senior/junior supervisors alike) can enact than what has been exercised in our own institution. Based on my years of supervisory experience and observation, it pains me to posit that the current HDR training model is still doing them a disservice, following a 'microscopic' approach that conditions HDR students to work in silos. I can't help but wonder, 'How can we revamp our HDR supervision approaches to better understand and support international HDR students' social-emotional learning in a time of crisis like this?' More specifically, 'How can we innovate inclusive and needs-based practices to empower HDR students' agency as emerging researchers beyond boundaries?'

Taking an autoethnographic approach (Adams et al., 2017, 2021; Ellis et al., 2011), I invite readers to walk into my critical reflection on my first-hand experience and lessons learned from a telecollaborative project arising from the pandemic crisis. The project was set up to develop a more ecologically balanced HDR model that focused not solely on student academic achievements but instilled empathy into social-emotional learning by building a supportive (online) community of practice. Through digital mining, I documented how the project was implemented, illustrated by my supervision journaling, student reflection entries, ECR colleagues' reports, and a series of HDR training webinars and activities. This chapter concludes with recommendations made for a more sustainable, needs-driven framework that can benefit both international and domestic HDR students and, reciprocally, their supervisors and other stakeholders involved in HDR supervision and administration.

Autoethnograhy

Autoethnographic studies have grown exponentially over the years; more and more emerging and seasoned researchers are using autoethnography to conduct research in social science, education, and applied linguistics (Adams et al., 2021). It is encouraging to see autoethnography resonate with many qualitative researchers, as rightly explicated by Adams and his colleagues (2021):

> autoethnographers recognize and embrace the reality that the person and personal are always present in social life as well as in the processes of research and representation. Everything we say and do— the language we use; the texts, images, and embodiments we create; the values we espouse— all are guided by perspective, experience, and social position. In this way, autoethnography is a research method that allows us to explicitly bring together the personal and the political as we face and address the challenges of today in a move toward envisioning a better tomorrow (p. 1).

As a legitimate research method (Stafford, 2022), autoethnography encompasses three key elements to "use personal experience ('auto') to describe and interpret ('graphy') cultural texts, experiences, beliefs, and practices ('ethno')" (Adams et al., 2017, p. 1). To account for true autoethnographic research, the study needs to comprise all three elements rather than focus solely on one or another, such as only offering a personal narrative without considering the cultural aspect (Adams et al., 2021). To illustrate how this research project captures the true essence of autoethnography, I use my personal story (self/auto) to depict and understand my first-hand experience as a participant observer in an online community of practice (graphy), shaping and shaped by transnational HDR students and ECRs throughout the pandemic year of 2020 (ethno). By critically reflecting on my lived social experience amid COVID-19, I intend to not only portray how the struggles and anxieties had impacted all the international students, me, and my ECR colleagues but reveal the silver linings for resilient strategies and opportunities on future HDR training in a broader context (Adams et al., 2017; Ellis et al., 2011).

Evidently, the past three pandemic years (2020-2022) have spawned even more autoethnographic studies as this unprecedented crisis has pushed us to dive deep into the personal, emotional, professional, and sociopolitical challenges faced by us (Chen & Sato, 2023; McAlinden & Dobinson, 2022; Morales et al., 2022). In some ways, the pandemic also provides a third space for us to spin a yarn on the impact of the ERT phenomenon on us and how we can transform the way we teach and research in the post-COVID era. As this telecollaboration HDR project was conducted fully online, I attempted to weave the vivid digital trails into my autoethnography, drawn from my researcher journal, HDR training webinar activities, student reflection journals using Google Docs, and my colleagues' project reports. By retelling a holistic and compelling story, I hope to help readers better understand what happened in this virtual transnational project vis-à-vis social distancing and border lockdown and how the lessons gained from our first-hand experiences can refine the current HDR model or in Adams et al.'s (2021) term, "in a move toward envisioning a better tomorrow" (p. 1).

In what follows, readers will discover how I interact with my autoethnographic account through critical self-construction, tapping into my own memory, critical reflection, digital trails, pivotal moments, and events to help analyze data and bring it all together in my narrative. Evidently, my own personal experience has also embodied the data and shaped my storytelling.

Where it all started: an international HDR mentoring project amid COVID-19

> This is probably one of the most long overdue and challenging telecollaboration projects I have embarked on. So let me break it down here: In collaborating with *VG* from Aberdeen, we received the grant approval last November (2019) for this project, "Mentoring HDR students in Applied Linguistics and TESOL through interinstitutional telecollaboration: A participatory action research design". We were pumped (well, at least for me)... Given our personal and academic experiences of navigating through the PhD trajectories, we know how much this project could benefit our students, reciprocally enabling us to learn and grow as HDR supervisors. Little did we know that the COVID-19 pandemic delayed our proposed timeline and derailed our plan. It's particularly in *VG*'s case, as he had to move all the F2F units fully online overnight. It's also hard to collaborate internationally during the most troublesome time. No one knows what the future holds— let's expect the unexpected (researcher journal, 23 April 2020).

When the border closures and social distancing started to hit us all hard at the beginning of the pandemic year, I initiated a telecollaboration project in response to the impact of COVID-19 on HDR supervision between my school and a sister school in Aberdeen, Scotland. Due to the context-responsive nature of telecollaboration, I adopted participatory action research (PAR) that "embraces the concerns experienced by a group, community or organization" (Wimpenny, 2013, p. 4). PAR requires the sustained involvement of the participants and empowers them to be agents (McTaggart, 1997; Jason et al., 2004) in order to transform their particular set of circumstances (Taylor et al., 2004). Specifically, it gives voice back to the HDR students, and lessons learned from the PAR process can potentially strengthen the quality of HDR supervision, thus making a wider impact on the community (Baum et al., 2006).

As a mid-career academic, I always embrace the notion of sharing my expertise with ECRs and mentoring HDR students. In this project, my research assistant (RA), *SJ*, is a PhD candidate, and my project collaborator, *VG*, is both an ECR and a junior supervisor in his institution: "I was just beginning my career as an HDR supervisor, so am really still in the process of figuring out my supervision 'style" (*VG*, mid-project report, 30 September 2020). To ease them into the realm of online HDR supervision, I conducted a debriefing session on the co-design principles (e.g., reflective practice) whilst guiding them through a robust project implementation process. Informed by PAR, we discussed how to provide better support for the student participants remotely, such as developing a needs analysis survey to identify their urgent needs, planning HDR training workshops on topics reflecting their interests, and piloting telecollaboration platforms using Zoom (for workshop webinars), Facebook (for virtual community building), Google Docs (for supervisors' and students' reflective journals). More importantly, this telecollaboration project would allow us to establish a reciprocal mentorship model, conducted remotely but connected social-emotionally, for fostering ongoing professional development of both HDR students and ECR colleagues.

Good things don't always come easy, especially during a crisis like the pandemic. After the ethics application was approved by both institutions, we held an online information session to debrief interested HDR students on the project goal, required tasks, and duration of the study, and how their participation could benefit their networking with other HDR students and professional growth, particularly in research capabilities. They were also ensured that their confidentiality would be protected, and they could withdraw at any point of the study without coercion. Finally, they were fully aware that any video and audio recordings, as well as screenshots of online activities, were used only for the purposes of data analysis and research dissemination. Nevertheless, we only received informed consent from a handful of our HDR students. This low participation rate was due to the fact that:

> most of my HDR students are PhD candidates who have completed their thesis research and don't think this mentoring support is urgent or as much needed as their junior colleagues... while *VG*'s students have shown high interest, they are worried about the intensity and time commitment for this project besides their academic studies (researcher journal, 23 April, 2020).

My research gut feelings told me to recruit more HDR students to pre-empt the participant attrition down the track. We also needed to sort out the meeting time (considering the time difference), webinar topics, and potential guest speakers, as well as sustain virtual community building. After our Facebook group was created, only one student posted his self-introduction. I started getting a bit anxious about whether this project would fly or sink given the 'dormant progression'. So, I emailed both *VG* and *SJ* to express my concerns and also suggested that we start the first telecollaboration soon so as to stimulate the dialogue and enable students to see the merit of this project.

One thing we all learned from the pandemic is to adapt, recoup, and be resilient. In my researcher journal entry (9 May, 2020) entitled "The power of social networking!", I wrote:

> Just when I thought the ship was going to sink due to the disappointing number and stagnant interaction between students in the FB group, a silver lining just shone through! Recall that I was exasperated by the low number of participants and less active interaction among both cohorts? To mitigate the situation, I decided to send out the project invites to my other Facebook Professional Groups. Voilà— we have received a good number of expressions of interest (EOIs) from other international students around the world, so much so that I needed to broadcast that the recruitment is closed now! It's definitely a pleasant surprise for us, and the overwhelming EOIs further justify how this project resonates with international students amid COVID-19. Never say never!

My acumen brought 15 international students located in Australia, Britain, the USA, Thailand, and Vietnam to join this telecollaboration HDR project. A few of them were English for speakers of other languages (ESOL) teachers keen to grab this opportunity to better understand what is expected in HDR studies before pursuing a postgraduate degree in the future. Table 1 summarises the needs analysis results based on the online survey we sent to all the student participants. This scoping helped us tailor the workshop topics to the interests and needs of our participants before we invited the potential guest speakers who were the domain experts.

Table 1. Summary of the needs analysis survey results.

Target area	Summary of responses
Demographic info	Gender: 9 females, 6 males Native language: Arabic, Chinese, Indonesian, Korean, Polish, Thai, Turkish Age range: 20-30 years old (33.3%), 31-40 years old (53%), 41-50 years old (13.3%) HDR status: newly admitted (26.7%), pre-candidacy (26.7%), post-candidacy (33.3%), near thesis completion (13.3%)
Research areas	communicative language teaching in EFL contexts multimodality, task-based language teaching (TBLT) vocabulary learning teacher education L2 writing, second language acquisition (SLA) reading comprehension in L2
Challenges	finding gaps in literature review time management distanced from family pilot delayed by COVID-19 research planning and data analysis use of statistical tools language barrier, lack of concentration, loneliness (anxiety, insomnia, etc.)
Expectations	honing research/academic writing skills steps in publishing inspired by seasoned researchers in the field networking with other HDR students to discuss research learning something new about doing PhD research
Potential workshop topics & timetabling	Icebreaker: Where I come from and my research agenda? (Facebook) W1: How to develop academic writing skills W2: How to write a lit review W3: How to develop survey items and interview protocols W4: How to publish your first article in an international journal W5: How to design research and analyse data W6: How to transcribe and use coding software (e.g., NVivo) W7: Guest speaker series (action research, journal editors' advice, etc) W8: 3-minute thesis presentations W9: Project evaluation & FAQ

Hit the ground running

> Today is the day. We have finally kicked into gear and are about to start our first virtual session in 15 minutes! I am as excited about seeing the whole group for the first time as I am nervous about how the first webinar is going to turn out. A lot of unforeseeable factors may go against us, such as technical glitches. May the force be with us (researcher journal, 14 May 2020).

As stressed above, the project aims to empower HDR international students through empathetic mentoring whilst fostering their professional outlook and capability, resilience, and agency in the global community. We held guest speaker webinars where scholars shared their insights with our HDR students, created a Facebook group to bring everyone closer, and provided ongoing guided support in students' journal reflections. Although the workshop planning had the participants' needs and interests at heart, our first webinar was a rollercoaster ride. The first guest speaker (*ML*) was unaware of the intellectual levels of our cohort (be they PhD candidates or English language teachers), talking to them as if they were all ESL learners. After we opened the floor for discussion, it went much better as students started to ask questions about social-emotional learning strategies and even shared their struggles to find peer support or ways to hone their academic skills. When one of my own students talked about his lack of self-worth and efficacy due to his 'mediocre/poor' English and how isolated he had felt as a PhD student, especially during this social distancing time, that pulled at our heartstrings. Evidently, his story made a compelling case for conducting this project.

For us HDR supervisors, especially junior ones, it is pivotal to critically evaluate what worked and did not work in each implementation phase of the telecollaboration project. This would help us refine the HDR mentoring model that was operated in a fully online, global space, informed by diverse academic needs and experiences, and shaped by a myriad of multicultural/lingual backgrounds. Both *VG* and I reflected on the whole experience after the first session and how we could improve next time:

> *VG*: Going forward the talk should be on research and research writing… I think we should do one session without a guest speaker and focus on using breakout groups before bringing it back to a whole group discussion. I learned a lot as well.
>
> *Me*: Right on, *VG*. I think we all learn from this whole virtual mentoring experience. That is why I have been keeping a researcher journal myself to reflect on my own supervision capacity more critically. I wasn't planning to include HDR students other than ours. But I am glad I did. Look how vibrant and enriching the community has evolved now compared with the initial stage (email exchange, 14 May, 2020)!

Collaborating internationally and remotely was undeniably challenging during the pandemic, much less inviting guest speakers and scheduling webinars to accommodate the time differences. We were not even sure whether those scholars would graciously accept the invite and speak with our student participants (for free). I was (still am) extremely appreciative of their unconditional generosity and solidarity in sharing their expertise on the topics we proposed (e.g., how to publish). This will serve as a constant reminder that giving back to the community and making an impact on students' lives is as important as building one's research track record, if not greater.

Not only did the students benefit greatly from the invited scholars' knowledge sharing, but we, as HDR supervisors, were also inspired by how much we still did not know, motivating us to keep upskilling. COVID-19 may have closed some windows for us at the time, but it has also left others open for us to see through the world. Above all, it has brought us together from different continents, institutions, and corners of life whilst pushing us to think, research, and teach outside the box and connect to people that we wouldn't have previously considered. For example, I would not have reached out to and discussed research collaborations with those well-renowned applied linguists had we not all been impacted by the pandemic. Another case in point is that we invited a language pracademic (DB) to unpack the framework and practice of doing "action research" (AR) (Figure 1). In my journal (entry #11), I reflected that:

> DB's talk couldn't have been more timely and inspiring. As our project is situated in AR, his expertise in doing AR fits perfectly into the nature and scope of our research. It also resonates with all of us, especially students who are language teachers but don't know where and how to start conducting teacher research in their own setting (researcher journal, 9 July, 2020).

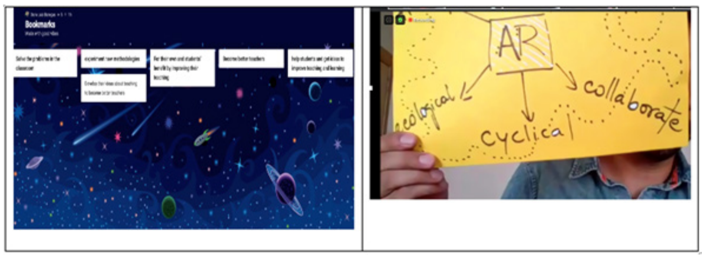


Figure 1. Guest speaker's webinar activities using Padlet for brainstorming (left) and a poster for illustration (right).

Both *VG* and I also used our own PhD journey to further motivate and engage our students. In the second webinar, I shared my story as an international postgraduate student in Canada and the US. I candidly described how I initially struggled with finding my own niche and strived to break through the hurdles through conference presentations and networking with the professional community. I illustrated how I rose to the occasion by being proactive and strategic about publishing my course assignments as an ECR and collaborating with other scholars across disciplines. This resonated so well with the participants that they posted on Facebook right after my led webinar (Figure 2). Nothing is more relatable than being vulnerable, humane, and candid about what made us who we are and how we can get stronger (together).

Student voices heard and responded to: an inclusive HDR mentoring approach

> I truly enjoy reading students' reflection journals. I am also grateful to them for sharing with me their ups and downs in life and what got them into this HDR journey… One of *VG*'s students talked about how much sacrifice he has to make and how hard it is to be away from his beautiful wife and kids. It brought back those bittersweet memories I had when doing my own PhD… it's certainly not a path for everyone, as some of them just drop out without seeing the finish line. So I totally got him (researcher journal, 16 May 2020).

Figure 2. Student Facebook posts after my led workshop (28 May 2020).

Before the first workshop, I asked students to give us some ideas on where they would like to keep their journals, such as via email, Google Docs or just on Facebook directly. Thankfully, most of them did not mind sharing their journals with each other using Google. We provided prompts for them to post their reflections in relation to workshop topics, such as "How I got into my PhD/MA/MPhil studies and embarking on this journey as an emergent researcher?" or "How COVID-19 has impacted my studies, engagement with my supervisor, peers and dissertation/HDR research?" To build a virtual community of practice and sustain member engagement, I invited *VG* to also provide mentoring support via dialoguing with the HDR students in their reflection entries (see my mentorship example using the commenting feature in Figure 3).

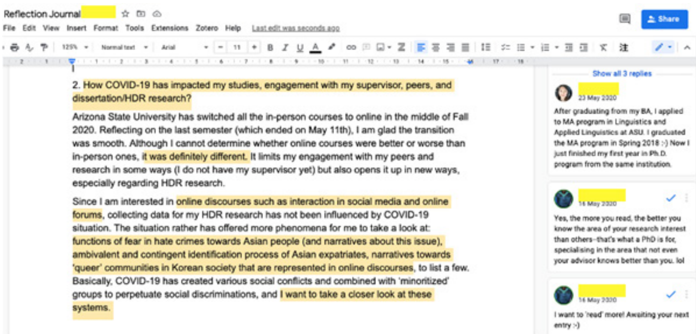


Figure 3. Mentor-mentee dialogue journaling via Google Docs.

As both a participant observer and online member of this telecollaboration community, I was able to shadow international HDR students by reading and commenting on their reflection journals. This online mentorship, augmented by two-way dialoguing, provided an organic, viable avenue for HDR supervisors to better understand and support students' social-emotional learning, which had been hampered by the pandemic crisis. Without this in place, we would not have known the deeper (sometimes darker) sides of struggles that our students had encountered, particularly at the beginning of their PhD journey. Recall *VG*'s student that I included in my researcher journal (16 May, 2020) above? He also expressed that working from home amid COVID-19 was even harder as he kept being distracted by his flatmates. He further commented that his academic writing was not up to the standards and was advised to seek academic support on campus. He used this opportunity to help him navigate the resources, and now he feels more confident about academic writing.

Is his case new to us? Probably not. It verifies the primacy of establishing this mentoring project to offer students moral and peer support by transforming social distancing into virtual community building, thus breaking down the spatiotemporal and psychological boundaries. Embarking on this telecollaboration project enabled me to help those students given the impact of solitude and self-doubt in HDR studies on their social-emotional learning. Despite the distance, I was able to provide my mentoring support remotely. I hope this project would benefit and empower them to become more socially adapted and academically proactive. Moreover, I was inspired to witness my mentoring—though physically remote but psychologically connected—had transformed the professional growth of the HDR students, and how this project championed diversity, inclusion, and social-emotional learning. I was also humbled to have met such a diverse student cohort from the globe, rather than the original bilateral telecollaboration:

> Indeed, this ever-changing time amid COVID-19 has challenged us to be more adaptable and think on our feet— turning a challenge (i.e., lower number of participants) into an opportunity (i.e., reaching out to the FB professional groups) (researcher journal, 16 May 2020).

Reflection on reciprocal mentoring telecollaboration

To assess the effectiveness of this HDR supervision model via a virtual exchange, I gathered stakeholder perspectives (HDR students, junior supervisor, ECR/RA) as part of the project evaluation. I first asked *SJ* (our RA) to conduct focus group interviews after the project ended. One of the international students is an Italian national who currently lives and teaches English in Vietnam. Before the pandemic, he was only thinking about applying for a Master's degree in the UK or Australia but didn't take action. Throughout the project, not only did he receive the peer and mentor support that led him to become an HDR student in *VG*'s institution, but he also flew with colours academically and was considering pursuing a PhD:

> Thanks to the knowledge shared by the group participants and mentors, I discovered new topics and gained new interests. I had the opportunity to get constant feedback on my first-ever research paper thanks to JC's mentorship... I went from "just" a teacher to an MSc student with a book review and research paper in the pipeline. Definitely happy to have joined and grateful for the opportunity, support, and mentorship I've received (focus group interview, 23 December 2020).

Thus far, this model has helped us better understand the social-emotional learning state of HDR students, exacerbated by social distancing during the pandemic. Besides their expectations to "learn from others and be inspired through the project" or "improve research knowledge and link with other researchers", our international students also faced different levels of hurdles, such as "language barrier, lack of concentration, loneliness, anxiety, insomnia, etc." (needs analysis response, 24 April 2020). These vivid vignettes mirror the rationale of conducting this telecollaboration project. That is, HDR supervision needs to go beyond simply the 'supervision' level. Instead, we should integrate social-emotional learning and connect to students' wellbeing, the latter of which is still 'an elephant in the room' in the current HDR practices. This also indicates the primacy of developing a more viable HDR supervision model as implemented in this project.

As a mentor for my ECR colleagues in this project, I find it also vital to evaluate how my mentoring has shaped their professional development and agency. As evidenced in the RA's mid-project report:

> JC's mentoring has greatly benefited the participants in the project and myself as an early career researcher... JC has demonstrated strong leadership [and] been instrumental in planning topics for the online seminars and for mentoring the participants through reflective journal feedback... He has a warm and approachable style when leading the project, and he always offers feedback and advice on the project's direction (*SJ*, mid-project report, 28 September 2020).

Indeed, the establishment of this online supervision model has offered both the HDR students and supervisors a viable platform for international networking, fostering professional teacher/researcher identity, polishing academic research skills, and obtaining ongoing peer/mentor support beyond geospatial boundaries. It also helps them gain new knowledge from active engagement (Rosier et al., 2015). I was also touched to know how my mentorship and guidance have inspired and empowered *VG*, my junior colleague, to refine his own HDR supervision practice:

> In the context of the COVID-19 lockdown, engagement in this group has been essential for some of my students and for myself... the international nature of the telecollaboration environment provides a broader view of student experiences and also facilitates networking opportunities for future collegial engagement in research... this is the model that not only should be carried on at my institution, but one that should be highly promoted for all majors at all universities. So, yes, I think the benefits of international telecollaboration are manifold, and I look forward to sharing and promoting this model through

> scholarship and conference presentations (*VG*, mid-project report, 30 September 2020).

This telecollaboration project has enabled me to establish a reciprocal mentorship model, conducted remotely but connected social-emotionally, for fostering ongoing professional development of both HDR students and ERC academics. These best practices, though not privy to the supervisors, are neither organically placed nor well promoted in our current HDR training. As observed in this project, international students need more guided HDR support, even more so during a crisis like the pandemic. This call for a more robust mentoring model infused with empathy is further echoed in my journal:

> Conducting this project made me realize that, yes, it's true that each School/Program has its own HDR support. However, not all the support is closely focused on individual students' needs or concerns, but more to do with showcasing the academics' own research (at least in my program). I feel that we can do better than that... HDR students might or might not have the chance to see the other side of the fence. Their journal reflections reveal that they need peer support or simply a group that can exchange ideas or listen to their concerns. I am glad that this project does just that (researcher journal, 28 May, 2020).

Final remarks

This context-responsive project has helped me rethink how we can transform conventional HDR supervision through transnational and telecollaborative exchanges beyond the localized constraints and spatiotemporal boundaries (Tran et al., 2017). The compelling case illustrated above implicates the importance of supporting international HDR students through peer support and ongoing mentoring. International telecollaboration provides an optimal channel for them to build a virtual community of practice where they can mutually support and learn from each other vis-à-vis academic challenges. This viable supervision model breaks down the power structure by creating an ecologically balanced framework in HDR training, thus promoting collaboration rather than isolation. Above all, it offers both the students and supervisors the opportunity to engage in professional dialogues by intellectually challenging each other, sharing know-how, and promoting collaboration. My ongoing reflective observations, students' constructive feedback, and junior colleagues' critical evaluations have validated this innovative HDR supervision that can benefit the impacted stakeholders.

Nevertheless, findings drawn from this unique case study should be interpreted with caution. The 'working in silos' model, which was exacerbated during the pandemic, was based on my observation and experience as a seasoned HDR supervisor in my institution. Hence, it might or might not be shared by colleagues in other institutions adopting different HDR training models. It is also not the intent of this small-scale case study to generalise findings to a bigger population but to honestly report on how this telecollaborative HDR project could make a difference in the development of professional growth, research capability, and agency of emergent researchers amid COVID-19. As such, the best practices drawn from this study might resonate with like-minded stakeholders who find the transnational, telecollaborative, and empathetic HDR mentoring model transferrable to their own settings.

References

Adams, T. E., Ellis, C., & Jones, S. H. (2017). Autoethnography. In J. Matthes (Ed.), *The international encyclopedia of communication research methods* (pp. 1-11). John Wiley & Sons.

Adams, T. E., Jones, S. H., & Ellis, C. (2021). Making sense and taking action: Creating a caring community of autoethnographers. In T. E. Adams, S. H. Jones., & C. Ellis (Eds.), *Handbook of autoethnography* (2nd ed., pp. 1-19). Routledge.

Appel, C., & Robbins, J. (2021). Language teaching in times of COVID-19: The emotional rollercoaster of lockdown. In J. Chen (Ed.), *Emergency remote teaching and beyond: Voices from world language teachers and researchers* (pp. 3-22). Springer. https://doi.org/10.1007/978-3-030-84067-9_1

Baum, F., MacDougall, C., & Smith, D. (2006). Participatory action research. *Journal of Epidemiology and Community Health, 60*, 854-857. https://doi.org/10.1136/jech.2004.028662

Chen, J. (2022). *Emergency remote teaching and beyond: Voices from world language teachers and researchers.* Springer. https://doi.org/10.1007/978-3-030-84067-9

Chen, J., Greenier, V., & Janes, S. (2023). A netnography of emergent ESOL researcher identity and development in a virtual community of practice. *International Journal of Applied Linguistics, 33*(2), 275-291. https://doi.org/10.1111/ijal.12465

Chen, J., & Sato, E. (2023). Reimagining crisis teaching through autoethnography: A case of an online Japanese course. *Innovation in Language Learning and Teaching, 17*(1), 157-167. https://doi.org/10.1080/17501229.2021.1973011

Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An overview. *Historical Social Research, 36*(4), 273-290.

Ferguson, H., & Spinks, H. (2021, April 22). *Overseas students in Australian higher education: A quick guide. Parliamentary Library Research Paper Series.* Canberra: Parliament of Australia. https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp2021/Quick_Guides/OverseasStudents

Hayton, J. (2013, February 4). *PhD stress: Don't ignore the warning signs!* The PhD Academy. https://jameshaytonphd.com/quick-tips/phd-stress-dont-ignore-the-warning-signs

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020, March 27). The difference between emergency remote teaching and online learning. *Educause Review*. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Jason, L. A., Keys, C. B., Suarez-Balcazar, Y. E., Taylor, R. R., & Davis, M. I. (2004). Participatory community research: Theories and methods in action. *American Psychological Association*. https://doi.org/10.1037/10726-000

Kozhabayeva, K., & Boivin, N. (2021). Emergency remote teaching in the Kazakhstan context: Deprofessionalization of teacher identity. In J. Chen (Ed.), *Emergency remote teaching and beyond: Voices from world language teachers and researchers* (pp. 3-22). Springer. https://doi.org/10.1007/978-3-030-84067-9_6

Lee, J. (2019, September 13). Australia should try to keep more international students who are trained in our universities. *The Conversation*. https://theconversation.com/australia-should-try-to-keep-more-international-students-who-are-trained-in-our-universities-123350

McAlinden, M., & Dobinson, T. (2021). Teacher emotion in emergency online teaching: Ecstasies and agonies. In J. Chen (Ed.), *Emergency remote teaching and beyond: Voices from world language teachers and researchers* (pp. 261-287). Springer. https://doi.org/10.1007/978-3-030-84067-9_13

McTaggart, R. (Ed.). (1997). *Participatory action research: International contexts and consequences*. Suny Press.

Morales, K., Romaguera, G., & Contreras, E. (2022). How to adapt in crisis: An autoethnographic approach to (re)building coursework. In J. Chen (Ed.), *Emergency remote teaching and beyond: Voices from world language teachers and researchers* (pp. 289-313). Springer. https://doi.org/10.1007/978-3-030-84067-9_14

OECD. (2021). *Education at a Glance 2021: OECD Indicators*. OECD Publishing. https://doi.org/10.1787/7edf2733-en

Prieto, L. P., Rodríguez-Triana, M. J., Odriozola-González, P., & Dimitriadis, Y. (2022). Single-case learning analytics to support social-emotional learning: The case of doctoral education. In *Social and emotional learning and complex skills assessment: An inclusive learning analytics perspective* (pp. 251-278). Springer International Publishing.

Rosier, K., Lohoar, S., Moore, S., & Robinson, E. (2015). *Participatory action research*. Child Family Community Australia (CFCA) Practitioner Resource. Australian Institute of Family Studies. https://aifs.gov.au/cfca/publications/participatory-action-research

Scott, H. (2019, August 9). The changing profile of Australia's international student market. *Campus Review*. https://www.campusreview.com.au/2019/08/the-changing-profile-of-australias-international-student-market/

Stafford, V. (2022). Book Review. Hernandez, K.-A. C., Chang, H., & Bilgen, W. A. (2022). Transformative autoethnography for practitioners – change processes and practices for individuals and groups. Stylus. *Journal of Applied Learning & Teaching, 5*(2), 201-204. https://doi.org/10.37074/jalt.2022.5.2.19

Taylor, R. R., Braveman, B., & Hammel, J. (2004). Developing and evaluating community-based services through participatory action research: Two case examples. *American Journal of Occupational Therapy, 58*(1), 73-82. https://doi.org/10.5014/ajot.58.1.73

Tran, L. T., Nguyen, L., & Green, W. (2017). *Exchanging the experience and outcomes of international HDR candidates: Reciprocal intercultural supervision* (Research Digest 12). International Education Association of Australia. https://www.ieaa.org.au/documents/item/1036

Wimpenny, K. (2013). Using participatory action research to support knowledge translation in practice settings. *International Journal of Practice-based Learning in Health and Social Care, 1*(1), 3-14. https://doi.org/10.11120/pblh.2013.00004

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

The critique of AI as a foundation for judicious use in higher education

Stefan Popenici[A]

[A] *Academic Lead - Quality Initiatives, Charles Darwin University, Australia*

Keywords

Artificial intelligence (AI);
bias;
discrimination;
education;
higher education;
learning;
surveillance;
teaching.

Correspondence

stefanpopenici@gmail.com [A]

Article Info

Received 25 June 2023
Received in revised form 4 July 2023
Accepted 4 July 2023
Available online 5 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.4

Abstract

The integration of Artificial Intelligence (AI) into educational settings, especially after the launch of ChatGPT into the public space, created new challenges and massive disruption for schools and universities across the world. This paper aims to state and look beyond the hype on AI, marketing and myths that are obscuring some of the most significant challenges and analyse potential risks associated with the adoption of AI in education. It also aims to find practical ways of using AI for the benefit of students, teachers and institutions of education. The analysis is focused on the key ethical implications of AI, the impact on teachers, students, and the future of learning, as well as long-term societal implications.

Ideological roots in defining human and artificial intelligence

The end of 2022, and especially the first months of 2023, will remain in history as the time when artificial intelligence (AI) invaded and colonised public imaginations across the world. The prominence of generative AI solutions such as ChatGPT, created and released for the public by OpenAI, made evident in an extraordinarily short period of time that the impact of AI on everyone's life will be exponentially more significant than it already was (Sullivan et al., 2023; Rudolph et al., 2023b). This makes it even more important for education at all levels, and especially for higher education, to stop and seriously interrogate what artificial intelligence is and what stands behind this marketing formula that came into the world in 1956 at the Dartmouth Summer Research Project (McCarthy et al., 1955). Defining artificial intelligence is notoriously difficult, especially if we consider that AI is now attached to almost any technology that can be sold and take advantage of the popularity of the term. In a paper published in 2017, I defined AI "as computing systems that are able to engage in human-like processes such as learning, adapting, synthesising, self-correction and use of data for complex processing tasks" (Popenici & Kerr, 2017, p. 2). In other words, AI describes technological solutions that are able to complete tasks that are usually associated with human abilities. And here is the point where we have a source of confusion: while it is true that AI can complete tasks that are usually requiring human abilities, it does not also mean that AI is intelligent or able to think. The second problem related to a definition of AI is that here we have a narrow understanding of what intelligence means based on a narrow and ideological description of human intelligence. We can understand this better if we simply stop to think about what stands behind the words that create the marketing label that is artificial intelligence. Records show that John McCarthy was clearly aware of the fact that artificial intelligence is, most of all, a marketing concept, a magnet for research funding, as well as a political move to distance himself from cybernetics and Norbert Wiener. When he won in 1971, the A. M. Turing Award, McCarthy mentioned in his speech to the Association for Computing Machinery (ACM) that it is ironic that his most widely recognised contribution "turned out to be in the field of marketing, specifically in choosing a brand name for the field." (Katz, 2020, p. 23).

The first part of the formula is the word "artificial", which is an unusual choice. If we try to find a synonym for artificial, we realise that this is not a positive term as it essentially opposes what is natural. Synonyms for artificial have, in general, a negative connotation, with replacements such as "fake", "fictitious", "false", "simulated", "bogus", "made-up", "forged", "not genuine", etc. This should be the first clue that we have a certain view on what kind of intelligence we are considering here. It is not a natural or human intelligence.

The second part of the formula is much more complicated because the history of the term is rooted in some dark and toxic ideological positions. The contemporary understanding of what human intelligence is is shaped by a cousin of Charles Darwin, Sir Francis Galton, who was convinced that everything could be measured. How passionate he was about measurements becomes clear when we consider that he believed possible to build an "attraction gauge" (how much a person is infatuated with another). He explored how we can scientifically measure boredom, and he authored in 1872 the paper "Statistical inquiries into the efficacy of prayer". More consequential is that Francis Galton was interested in human intelligence, and he translated his passion for measuring everything to this field, setting the foundations of a certain way of looking at human intelligence, which was later developed in the discipline of psychometrics. To understand better how Galton's perspective on human intelligence influenced our current understanding of what human intelligence is, it is important to look at his other significant contribution to posterity: he also coined in 1883 the term eugenics, the reprehensible and toxic theory of rankings of human abilities based on racial and hereditary factors, which is commonly associated only with Nazi Germany and the Holocaust. This is just the effect of these ideas, but Galton founded the theory of "racial hygiene" and eugenics, which stands as the most influential pillar for the new studies and ideas of intelligence. The real impact of this reprehensible theory erupted a few decades later when Nazis adopted eugenics as an ideological foundation. These theoretical roots stand responsible for the unprecedented tragedy of the Holocaust. Historical evidence shows that Hitler was particularly interested in Galton's ideas about intelligence, heredity and racial hygiene. Otto Wagener, the head of the Nazi Party's Economic Policy Office from 1931 to 1933, specifically noted Hitler's interest in Galton's theory of eugenics and its applications in the United States. In his notes, Wagener quoted Hitler saying that "it is possible to a large extent to prevent unhealthy and severely handicapped beings from coming into the world. I have studied with great interest the laws of several American states concerning prevention of reproduction by people whose progeny would, in all probability, be of no value or be injurious to the racial stock" (Kühl, 1994, p. 37).

What is important to realise is the fact that these toxic ideas were not adopted only by an influential thinker that set some foundations for studies in this field at the end of the 19th century. The reality is that these theories remain influential across decades until the present. Most influential scholars in the field of intelligence pay tribute, most often with explicit enthusiasm, to a hereditary and discriminatory perspective on what human intelligence is and how it can be identified and measured. The most influential figures that shaped our current understanding of intelligence, from Galton and Karl Pearson, William James and John Dewey, to Lewis M. Terman, the famous professor of psychology at Stanford University with immense studies on human intelligence, stand close to a eugenic, hierarchical view of intelligence that is presented as determined by social status and hereditary mental abilities. All these thinkers and theories stand close to the disastrous and wrong assumptions set by Galton in his attempt to scientifically justify racism, social injustice and discrimination. In fact, Lewis Terman, who is undoubtedly one of the most influential scholars in the modern studies of intelligence, was an active member of eugenic societies in the United States and actively advocated for the forced sterilisation of those labelled as inferior in American society (Leslie, 2000). There is an obvious impact of these reprehensible ideas inspired by eugenics, with events that will forever mark the history of humanity with their power to fuel extreme violence, perversion of thinking, and abysmal inhumanity. A more discrete and insidious effect of these theories is that they restricted collective imaginations and scientific research to a narrow view of what human intelligence is, how it manifests, and how and if it can be measured or nurtured.

It is important to note that the development of new technologies became the most fertile ground for eugenics and racial theories of intelligence. William Shockley is considered the "father of Silicon Valley", a winner of the Nobel Prize in Physics in 1956, and the one who "brought silicon to Silicon Valley" (Isaak et al., 2016, p. 167). William Shockley is also a known racist, white supremacist, and strident supporter of eugenics. He was invited in 1965 to

> deliver an address at the first annual Nobel Conference, a conference on Genetics and the Future of Man that was held in the United States but was authorised by the Nobel Foundation. At that conference, Shockley revealed his racist ideology. He claimed that social policies were allowing genetic defectives to proliferate... To Shockley's mind, only by systematic reduction of the African American population by sterilization and other methods of birth control could we improve our society. This would lead to survival of the fittest, and the fittest were the original European settlers into America. Racial discrimination was not prejudice, he claimed, but was justified based on statistics (Sussman, 2014, p. 236).

The adherence to this toxic view of the world and discrimination based on old misjudgements and prejudices is also closely associated with the emergence of artificial intelligence as a formula and theory. John McCarthy, who coined the term "artificial intelligence" in 1955 (McCarthy et al., 1955), openly expressed his beliefs on the hierarchical structure of intelligence based on gender, underlining in an essay titled "Technology and the position of women" his beliefs. He states there that it would be a mistake to assume that women are not inferior to men, writing that:

> The very highest level of potential in science and mathematics, which only one in a million men can attain, the fraction of women who can attain it may be biologically smaller... At present there are social movements and people with institutional power who regard there being fewer women than men at some level of some occupation as an injustice that must be corrected by quotas. This is a mistake and will not succeed because of differences in ability and motivations between males and females (McCarthy, 2006).

In June 2023, France24 published a documentary based on investigative journalism on Clearview AI, a company specialised in facial recognition that is scraping astonishing amounts of data on every person who had a picture taken and uploaded, with or without consent, on the Internet. In this documentary, Jessica Le Masurier and Romeo Langlois (2023) unveil not only the staggering breach of privacy and serious implications for citizens across the world but a surprisingly obvious link between this powerful company and white supremacists, fascists, and anti-democratic forces in the US.

We must consider in education that AI is placed at the convergence of two dangerous temptations, which both shape its development, influence, and dynamics The first is the temptation of technology and its efficiencies to lead towards a certain view of the world, which is easily adopted by totalitarian, amoral tendencies. This connection was summarised by Herbert Marcuse in his analysis of Nazi dictatorship as "a striking example of the ways in which a highly rationalised and mechanised economy with the utmost efficiency in production can also operate in the interest of totalitarian oppression" (Marcuse & Kellner, 1998, p. 416). From a purely technological perspective, Nazi Germany was the most advanced at that time, creating the first man-made object in space (the infamous V2 rocket), making technological advancements that opened the space exploration program a few years after their defeat. Technological excellence was not making that regime less evil but worse and more destructive. This historical fact invites a serious consideration of the necessity to associate technological progress with ethical considerations and to maintain a critical perspective on technological development.

The second temptation for artificial intelligence is much more straightforward and evident. There is a documented tendency of AI to immensely enhance surveillance and inequality, bias, and discrimination and widen power imbalances. A disconnect from ethical considerations is dangerous for civil societies, democratic processes and educational aims for higher education.

Two dangerous myths about AI in education

One common misconception affecting the perception of AI systems and how they impact education is that data is an objective construct, atemporal and value-neutral, shaped only by exact and cold evidence and accurate representations of reality. In fact, the perception is that technology itself is an objective medium. Hence, AI is a technological solution that is operating based on factual, unbiased and clinical processes. If we think about how technology actually operates, we realise that there is not one point in the history of humanity when technology is no directly related to specific cultures and values, beliefs and biases, religious beliefs or gender stances. A study on gender bias in technology starts from these basic facts, noting that

> first, and foremost, (there) is the notion that technology often shapes culture and its meanings. The second is that we have become so used to technology in our daily lives that we fail to see its implications. We argue that our familiarity with technology means that traditional methods of analysis will be unable to unveil the ideology that perpetuates gender bias as a mode of domination. Thus, a critical analysis of technology and society is required for technology to reach its emancipatory potential (Kilbourne & Weeks, 1997, p. 244).

Indeed, any informed and responsible use of technology, especially revolutionary solutions such as generative AI, requires a critical analysis that cannot start from the naive perception that we can have in this space, a vacuum of values and specific choices. Any technological solution and adoption involve a certain ideological choice and influence, consciously or not. If we accept the obvious fact that values and particular perspectives influence the development and applications of technologies, we have to consider a series of troubling facts, such as the vast disproportionate influence of men on the development and programming of AI. Currently, only a small percentage of women work now in AI: "The percentage of women working in AI today is approximately 30%" (WEF, 2023, p. 7). In Silicon Valley, the percentage is even smaller, just above 10%. To use just one example, we can consider that one research project conducted at the University of Cambridge has found that using AI to complete literature searches provides results with a bias favouring white, Western and male authors (Jordan & Tsai, 2022). This means that other voices and perspectives in research and the advancement of knowledge become de facto invisible. Implications for our common future are significant.

The second common belief is somehow linked to the first position that leads to errors in the use and adoption of new technologies in education and revolves around the idea that algorithms do not discriminate, as they are "*just maths*". In her book, "Weapons of math destruction: how big data increases inequality and threatens democracy", Cathy O'Neil provides a convincing and well-justified counterargument to this erroneous position, noting that mathematics cannot offer alone protection against bias, misuse, and manipulations. The book documents that:

> The math-powered applications powering the data economy were based on choices made by fallible human beings. Some of these choices were no doubt made with the best intentions. Nevertheless, many of these models encoded human prejudice, misunderstanding, and bias into the software systems that increasingly managed our lives. Like gods, these mathematical models were opaque, their workings invisible to all but the highest priests in their domain: mathematicians and computer scientists. Their verdicts, even when wrong or harmful, were beyond dispute or appeal. And they tended to punish the poor and the oppressed in our society, while making the rich richer (O'Neil, 2016, p. 10).

There is consistent research and books that are providing examples of AI algorithms that discriminate, grotesquely amplify injustice and inequality, targeting and victimising the most vulnerable and exposing us all to unseen mechanisms of decision where we have no transparency and possibility of recourse. It is worth mentioning here the book "Automating inequality: How high-tech tools profile, police, and punish the poor", by Virginia Eubanks (2018). In "Algorithms of oppression. How search engines reinforce racism", Safiya Umoja Noble starts an excellent expert analysis by reminding us that "Part of the challenge of understanding algorithmic oppression is to understand that mathematical formulations to drive automated decisions are made by human beings", and documents that "algorithmic oppression is not just a glitch in the system but, rather, is fundamental to the operating system of the web". (Noble, 2018). This point is extremely important for any informed user of AI, especially in education, and the fact that discrimination and racial biases are part of the internal design rather than a simple "glitch". In the book published in 2023, titled "More than a glitch: Confronting race, gender, and ability bias in tech", Meredith Broussard (2023) substantiates the point that value neutrality in tech is a myth and, as it expertly proves that bias and discrimination are not a simple error but a matter of design, brings new arguments to focus our efforts on holding algorithms transparent and accountable.

It is impossible to make here a comprehensive selection of some of the most relevant research, books, studies or even journals reflecting the fact that algorithmic decision-making is inherently dangerous and toxic without constant and alert human supervision and interrogation. What stands relevant is that the blind trust in and adoption of new tech by educators, which was ubiquitous for the last decades in schools and universities across the world, becomes even more dangerous in the era of AI. The challenge ahead for education is to become *users* of AI for the benefit of our students and institutions rather than simple *subjects* of AI, providers of data and digital serfs controlled by an almighty Big Tech.

AI and the aims of education

A research paper submitted for preprint by a group of researchers from Princeton University, the University of Pennsylvania and New York University identifies professions that are most likely to be lost or degraded by the impact of AI. Researchers have found here that the vast majority of those most exposed the AI disruption are teachers in schools and higher education (Felten et al., 2023). This confirms what became obvious in the first months of 2023 after ChatGPT and other large language models and generative AI solutions became popular with the general public. The impact on education, students, teachers, schools and universities was not close to the main concerns of developers of AI or tech startups. Moreover, there are sufficient reasons to claim that the aims of educators stand very far from their motivations. For example, in a 2019 interview published by Forbes, Sam Altman, the CEO of OpenAI, the company that developed ChatGPT, makes the significant observation that "AI will probably most likely lead to the end of the world, but in the meantime, there'll be great companies" (Martin, 2019). Considering that OpenAI managed to secure immense funding and profits in a very short time, we can safely assume that "great companies" describes profitable companies here, and here is the key: Big Tech is driven by the aims of profits and power, control and financial gain. Institutions of education and teachers have very different aims: the advancement of knowledge and to nurture educated, responsible, and active citizens that are able to live a balanced life and bring a positive contribution to their societies. As noted in the book "Artificial Intelligence and Learning Futures", institutions of higher education

> were created to serve the common good and, with their concentration of academics working together for research and education, can advance knowledge and serve society with wise and innovative solutions for our critical challenges. Higher education is the space where new ideas can organically emerge, when the campus ethos is defined by intellectual effervescence, and moral engagement for a civil and advanced society. The general aim of universities is to disseminate knowledge and nurture more informed, ethical and educated citizens, able to bring a positive contribution for a civil society (Popenici, 2022, p. 3).

Big Tech and new tech startups are established to secure profits and control, at least on the market. It is significant that OpenAI is an example of a startup that was established with the aim to "serve humanity", "unconstrained by a need to generate financial return", and just a few years later became entirely opaque in the design and use of their algorithms while securing billions of dollars in new funding. There is not even a point of convergence in the aims of those who currently control and build AI and the users of AI in education, a field that is undoubtedly going to suffer major changes in the near future as a result of its rapid development and adoption in teaching, learning and assessment. This is an important reason for universities and educators to stay alert and interrogate intentions and applications of AI, as well as keeping strict control on the AI inherent tendencies to discriminate and amplify biases.

The use of AI is also directly linked to a set of risks related to users' privacy. The popularity of generative AI in 2023 obscured that all details, prompts and use of AI involve two clear dynamics. First, all information and data the user provides are used to train and develop the AI models. While it is absolutely acceptable for a teacher to contribute freely to the development of models that increase the profits of private companies, some students may not be aware that their work and ideas are used by a third party in ways that are not always clear. The second dynamic is much more significant: all data provided to an AI solution is potentially filed, used, aggregated, and connected to a user's identity. Especially at a time when banks use data aggregated from the Internet to decide a credit score, insurance companies decide premiums based on information sold by data brokers and all our lives are influenced by data collected on individuals with and without their knowledge, teachers have a duty of care to protect the privacy of students and their future.

As we briefly detailed in previous paragraphs, algorithmic discrimination is a tangible reality which significant effects on large groups of people and is especially damaging for the marginalised and most vulnerable in society. This is directly associated with AI and its functions and is raising specific challenges in education as it inevitably requires data and information from the users. Any breaches of privacy and disclosure of sensitive data will have long-term impacts, which are impossible to evaluate due to the opaque nature of AI models. Even the most ardent and interested supporters of generative AI raise the alarm about the potential risks for privacy and data confidentiality: in June 2023, Google (Alphabet) warned its own staff to avoid sharing personal or professional information on AI chatbots, including on its own AI solution, Bard (Dastin & Tong, 2023).

There is also the real risk that learning is further pushed to the margins of the process in the current hype surrounding the potential of AI to improve education, assessment and teaching. This is summarised in an analysis of the adoption and use of ChatGPT in higher education, at a moment when it is tempting to use artificial intelligence to assess the originality of assignments: "A first AI circumvents a second AI and is assessed by a third AI. All that the humans do is press a couple of keys, and nobody learns anything" (Rudolph et al., 2023a, p. 354). AI presents obvious advantages in automating assessments, further personalising teaching, providing individualised assistance or replacing university administration, but it can also further alienate and dehumanise learning, to the point of technological potemkinisms as those described by Rudolph et al. (2023a). There is no evidence that universities across the world use the ChatGPT moment to radically change their governance and ideological models, and structurally change teaching and assessments to nurture key skills for the current challenges, students' creative and critical thinking abilities (Rudolph & Tan, 2022), wisdom and social responsibility.

The long-term fixation of education on personalisation also requires the collection and aggregation of student data, and AI brings now new challenges to a project that was brilliantly analysed by Audrey Watters (2023) in her book, "Teaching machines: The history of personalized learning". The rapid adoption of AI solutions by educators, administrators and students opens new possibilities for hyper-personalisation and data collection: "AI is bringing the promise of 'super-charging' personalisation of education, using data and complex algorithms to predict what is the most suitable content, teaching method, educational intervention, and pace of instruction for every student" (Popenici, 2022, p. 107). We cannot properly evaluate how super-personalisation and vast data collection will impact the future of our students, on their credit ratings, mortgages, medical services and so on, but we have the duty of care to protect them from exploitative and potentially damaging practices. The task of helping our students become informed and able users of various AI platforms becomes more important now.

Conclusion

The impact of generative AI is most visible in the area of learning, teaching, and especially academic integrity. The rapid adoption and unprecedented number of users in a very short time came as a "shock in education, like the COVID-19 pandemic" (Mills et al., 2023, p. 16). Noam Chomsky, who is most probably the most reputable professor of linguistics and cognitive science, can offer some guidance on the potential impact of ChatGPT, the most popular AI program for generative language. Chomsky succinctly defined the role of ChatGPT in education as "High-Tech Plagiarism" and "a way of avoiding learning" by students, noting in an essay written with Ian Roberts, a professor of linguistics at the University of Cambridge and Jeffrey Watumull, a philosopher and the director of artificial intelligence, that

> ChatGPT exhibits something like the banality of evil: plagiarism and apathy and obviation... ChatGPT and its brethren are constitutionally unable to balance creativity with constraint. They either overgenerate (producing both truths and falsehoods, endorsing ethical and unethical decisions alike) or undergenerate (exhibiting noncommitment to any decisions and indifference to consequences). Given the amorality, faux science and linguistic incompetence of these systems, we can only laugh or cry at their popularity (Chomsky et al., 2023).

It is deceiving to say, dangerous to believe, that artificial intelligence is... intelligent. There is no creativity, no critical thinking, no depth or wisdom in what generative AI gives users after a prompt: it is just plausible text with good syntax and grammar, and this is all that it is. Intelligence, as a human trait, is a term that describes a very different set of skills and abilities, much more complex and harder to separate, label, measure and manipulate than any computing system associated with the marketing label of AI. AI is already tentatively used to replace teachers in higher education, and publications such as The Independent in the UK are spreading the hype with titles such as "Harvard's new computer science teacher is a chatbot" (Cuthbertson, 2023). Harvard University presented this as "an evolution to tradition", that "can support their learning at a pace and in a style that works best for them individually" (Hamid & Schisgall, 2023). Universities will further integrate AI in their courses and will use AI bots to replace teachers, not because it will help students develop skills that will be relevant and help them in the era of AI, such as independent and critical thinking, superior abilities to master knowledge with genuine creativity and meaningful contributions. The AI replacement of teachers is a process well aligned with the ideological models adopted by universities for the last decades: marketisation, maximisation of profits that can be secured by culling teaching employees and viewing learning as an assembly line process where information is delivered as a product and commodity, serving models narrowly suited for employment and the job market. This is a risk that is not yet on the agenda of politicians or decision-makers, and it is not part of the agenda of university administrators. The risk will not disappear and most probably will accelerate the current crisis of learning and teaching, and the crisis impacting enrolments and the public trust in higher education.

It is evident at this point that AI is an integral part of education, as it has been for a long time – more discreetly – in many other areas of our lives. Banning or ignoring generative AI in education is an unrealistic, ignorant, and dangerous option, which was unfortunately adopted by many educators, schools and universities when it became clear how widely used ChatGPT is. It is vital for educators to understand what AI is and what it is not, what is just hype and marketing, and make the difference between the real potential for beneficial use or selling points and propaganda. It is also important to identify real expertise or just a desire to join the hype or a simple lack of knowledge. This is a hard task, as the vast resources allocated to promote selling points of companies with vested interests in this field are building passionate defences of AI, usually associated with religious fervour. However, if universities and educators want to remain relevant in the future and have a real chance to reach the aims of education, it is important to consider the ethical and intellectual implications of AI, some not even mentioned in the current paper. This will be a field open for new and extraordinarily significant future research.

References

Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability bias in tech.* MIT Press.

Chomsky, N., Roberts, I. & Watumull, J. (2023, March 8). The false promise of ChatGPT. *New York Times*, https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Cuthbertson, A. (2023, June 24). Harvard's new computer science teacher is a chatbot. *The Independent,* https://www.independent.co.uk/tech/harvard-chatbot-teacher-computer-science-b2363114.html

Dastin, J., & Tong, A. (2023, June 16). Google, one of AI's biggest backers, warns own staff about chatbots. *Reuters*, https://www.reuters.com/technology/google-one-ais-biggest-backers-warns-own-staff-about-chatbots-2023-06-15/

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

Felten, E., Raj, M., & Seamans, R. (2023). *How will language modelers like ChatGPT affect occupations and industries?*. arXiv preprint arXiv:2303.01157.

Hamid, R. D., & Schisgall, E. J. (2023, June 21). CS50 will integrate Artificial Intelligence into course instruction. The Harvard Crimson, https://www.thecrimson.com/article/2023/6/21/cs50-artificial-intelligence/

Isaak, R., Isaak, A., & Zybura, J. (2016). Replicating Silicon Valley: talent and techno-management in a culture of serendipity. In *Entrepreneurship and talent management from a global perspective* (pp. 149-187). Edward Elgar Publishing.

Jordan, K. & Tsai, S. P. (2022). 'Sort by relevance': Exploring assumptions about algorithm-mediated academic literature searches. *Society for Research into Higher Education (SRHE) Research Report*.

Katz, Y. (2020). *Artificial whiteness: Politics and ideology in artificial intelligence*. Columbia University Press.

Kilbourne, W., & Weeks, S. (1997). A socio-economic perspective on gender bias in technology. *The Journal of Socio-Economics, 26*(3), 243-260. https://doi.org/https://doi.org/10.1016/S1053-5357(97)90018-4

Kühl, S. (1994). *The Nazi connection: Eugenics, American racism, and German national socialism*. Oxford University Press.

Le Masurier, J., Langlois, R. (2023). Your face is ours: The dangers of facial recognition software. *France24,* https://

www.france24.com/en/tv-shows/reporters/20230602-your-face-is-ours-the-dangers-of-facial-recognition-software

Leslie, M. (2000, July/August). The vexing legacy of Lewis Terman. *Stanford Magazine*, https://stanfordmag.org/contents/the-vexing-legacy-of-lewis-terman

Marcuse, H., & Kellner, D. (1998). *Collected papers of Herbert Marcuse*. Routledge.

Martin, N. (2019, June 27). 13 best quotes about the future of Artificial Intelligence. *Forbes*, https://www.forbes.com/sites/nicolemartin1/2019/06/27/13-greatest-quotes-about-the-future-of-artificial-intelligence

McCarthy, J. (2006). Technology and the position of women. *Stanford Computer Science Department: Stanford University,* http://jmc.stanford.edu/commentary/future/women.html

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955, August 31). *A proposal for the Dartmouth Summer research project on Artificial Intelligence,* http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1), 16-30. https://doi.org/10.37074/jalt.2023.6.1.34

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism.* New York University Press.

O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy.* Penguin.

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12*(1), 22. https://doi.org/10.1186/s41039-017-0062-8

Popenici, S. (2022). *Artificial Intelligence and learning futures: Critical narratives of technology and imagination in higher education.* Routledge.

Rudolph, J., & Tan, S. (2022). The war in Ukraine as an opportunity to teach critical thinking. *Journal of Applied Learning and Teaching, 5*(1), 165-173, https://doi.org/10.37074/jalt.2022.5.1.24

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364.389. https://doi.org/10.37074/jalt.2023.6.1.23

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching, 6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Sussman, R. W. (2014). *The myth of race: The troubling persistence of an unscientific idea.* Harvard University Press

Watters, A. (2023). *Teaching machines: The history of personalized learning.* MIT Press.

WEF (2023). Global gender gap report 2023. *World Economic Forum,* http://reports.weforum.org/globalgender-gap-report-2023

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Rethinking online assessment strategies: Authenticity versus AI chatbot intervention

Augustine Osamor Ifelebuegu[A]

[A] Professor, Institute of Education and Lifelong Learning, Victoria University Kampala, Uganda

Keywords

Academic integrity;
AI;
chatbots;
ChatGPT;
conversational agents;
generative AI;
online assessment.

Correspondence

augustine.ifelebuegu@vu.ac.ug [A]

Article Info

Received 23 June 2023
Received in revised form 29 June 2023
Accepted 29 June 2023
Available online 3 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.2

Abstract

As artificial intelligence (AI) and chatbot technologies like ChatGPT continue to evolve, educators grapple with the risks and benefits these advances bring to online assessment. The democratisation of AI-based technologies, while offering personalised learning experiences, threatens online assessment legitimacy and academic integrity. This paper critically examines the intersection of AI chatbots and online assessments, in the context of their impact on the design of authentic online assessments. The widespread usage of AI chatbots has caused serious problems for the validity of online tests because of the possibility of student abuse. This underlines the need for 'authentic assessments' that concentrate on higher-order cognitive skills, problem-solving, creative thinking, and collaborative talents and calls for a reevaluation of conventional assessment methods. These types of assessments not only align with the evolving pedagogical needs of the 21st century but also present tasks that are significantly challenging for AI chatbots to replicate, thereby preserving their integrity. Conversely, the paper also explores how AI can facilitate the assessment process by automating certain tasks, providing personalised learning experiences, and supporting collaborative assessments. The era of AI chatbots presents an opportunity to rethink and enhance online assessments, making them more authentic, meaningful, and resistant to AI-assisted malpractice.

Introduction

In the realm of education technology, few innovations have been as intriguing and controversial as artificial intelligence (AI) and its application in chatbots. To contextualise the ensuing discussion, it is important to define what AI chatbots are. They are essentially programmed entities capable of interacting with humans in natural language and performing tasks that ordinarily require human intelligence (Daniel, 2020). In November 2022, OpenAI, a California-based company released the ChatGPT-3.5 language model which was followed by an updated version (GPT4) in March 2023. These Generative Pre-trained Transformers are able to perform language related tasks including answering questions, generating texts and many more (Rasul, et al. 2023). Since the release by OpenAI, several other chatbots have hit the markets and more continue to emerge on a weekly basis. These chatbots and AI tools are permeating educational landscapes, offering personalised learning experiences and immediate feedback (Zhao et al., 2021). However, they concurrently pose challenges to the authenticity of learning outcomes, as they can be misused to automate or assist in traditional assessments (Daniel, 2020, Rudolph et al., 2023a, Sullivan et al., 2023).

With the increasing prevalence of remote learning and online education, concerns over cheating in assessments have also risen. As an important component of online learning, assessments are crucial in assessing students' progress and understanding. However, online assessments can be vulnerable to various types of academic misconduct such as plagiarism, use of unauthorised aids, and repeated attempts. AI chatbots have precipitated significant shifts in pedagogical landscapes, offering unparalleled interactive capabilities (Abdelghani et al., 2022; MacNeil et al., 2022; Daniel, 2020). However, they also introduce new threats to the integrity of online assessments by providing means to automate or assist assignment completion (Daniel, 2020). Before the rise of AI chatbots, online assessment was already linked to greater instances of academic dishonesty (Clarke et al., 2023) as well as increased threats to academic integrity (St-Onge et al., 2022). The Covid-19 pandemic also produced a distinct set of circumstances that correlates with an increased number of academic dishonesty cases (Perkins, 2019; Henderson et al., 2022; Lancaster & Cotarlan, 2021) and the perceptions of academic dishonesty among students or academic staff (Amzalag et al., 2021; Reedy et al., 2021; Walsh et al., 2021). The rise of the use of AI chatbots in exam malpractice is expected to see a significant increase during online assessments, if educators do not rethink their assessment strategies. Hence, there is an urgent need for an overhaul of traditional assessment strategies. The thoughts expressed in this paper are based on a critical review of the existing and emerging body of literature.

AI chatbots: A double-edged sword

AI chatbots offer the promise of personalised, adaptive learning and immediate feedback (Zhao et al., 2021). Yet, they also pose challenges to the authenticity of learning outcomes, as they can be misused to bypass conventional online assessments (Daniel, 2020), creating both unethical and inequalities issues (Rasul et al., 2023). This duality raises the question of how to benefit from the potential of AI without compromising assessment integrity.

The promising edge: Advantages of AI chatbots

AI chatbots in education can contribute significantly to personalised learning experiences, providing adaptive instruction that adjusts to individual learners' needs (Daniel, 2020). By doing so, they can cater to a range of learning styles and paces, offering a more inclusive and accessible learning experience (Cheng & Chau, 2016). The immediacy of feedback that AI chatbots can provide is another advantage, allowing students to assess their understanding and adjust their learning strategies promptly (Zhao et al., 2021).

Moreover, AI chatbots can engage students in dialogic learning, simulating conversational interactions that promote active learning (Wegerif, 2006, Rudolph et al., 2023b). For instance, chatbots can ask probing questions to challenge students' understanding, fostering critical thinking and deep learning. They can also assist in formative assessments, providing immediate feedback on students' progress and guiding them towards improvement (Siemens et al., 2015). According to Rasul et al., (2023), the utilisation of chatbot technology has demonstrated beneficial impacts on various aspects of the learning process, including the enhancement of explicit reasoning capabilities, improvements in learning outcomes and knowledge retention, as well as a heightened interest and engagement in the learning process.

AI technologies can be utilised to automate and enhance various aspects of assessment design, delivery, and grading. For instance, AI can automate the generation of diverse, complex questions that assess higher-order cognitive skills, thereby reducing the manual workload for educators (Bridgeman et al., 2023; Gierl & Lai, 2013). Also, AI can be used to personalise assessments based on individual students' needs and progress, thus facilitating differentiated instruction and personalised learning (Vandewaetere et al., 2011; Stahl, 2023). Chatbot technology can also be seamlessly incorporated into assessment activities. For instance, students can critically analyse and refine text or essays generated by chatbots, thereby stimulating their existing conceptual frameworks and fostering critical thinking skills (Dennick, 2016). This method equips students with vital skills for interacting with systems like ChatGPT in future professional environments. AI can also assist in the grading of certain types of assessments. Automated essay scoring systems, for instance, can provide quick, objective grading and feedback on student essays (Shermis & Burstein, 2013). Similarly, AI systems can assist in the evaluation of complex tasks like coding assignments (Piech et al., 2015). These technologies can free up educators' time, allowing them to focus on other, more critical aspects of the teaching and learning process.

Notably, AI can play a role in supporting collaborative assessments as well. AI-based analytics can track and analyze individual contributions to group tasks, making it easier to evaluate each student's performance (Ferguson, 2012). Furthermore, AI can monitor and guide online discussions, ensuring that all students participate equally and promoting critical thinking and effective collaboration (Chan et al., 2023). However, the use of AI in facilitating authentic assessments is not without challenges. Concerns include the potential for AI to make errors, the difficulty of programming AI to appreciate nuances in human responses, and the risk of over-reliance on technology. Moreover, the implementation of AI requires significant investments in technology and training, potentially exacerbating the digital divide and increasing inequality in education (Reich & Ruipérez-Valiente, 2019).

The perilous edge: Risks of AI chatbots

Despite the promising capabilities of chatbots powered by artificial intelligence, their misuse poses significant challenges to academic integrity. The same technology that facilitates personalised learning can be applied to tasks designed to assess a student's comprehension. An AI chatbot can be readily programmed to respond to multiple-choice questions, complete fill-in-the-blank tasks, and even generate brief written responses, thereby undermining the authenticity of assessments (Daniel, 2020). Moreover, students may become overly reliant on AI chatbots to answer their queries or solve their problems, impeding the development of critical thinking and problem-solving skills. Students may forsake the deep learning process in favour of AI-generated answers (Rasul et al., 2023; Pellegrini & Quellmalz, 2010), which could result in a rudimentary understanding of course content. According to Seo et al. (2021), if ChatGPT and other AI models are used for rapid and superficial learning, they may hinder the development of graduate-level skills such as critical thinking and problem-solving.

The difficulties posed by AI chatbots are not restricted to students alone. Educators may also become excessively reliant on AI for tasks such as grading, disregarding subtleties in student responses or missing opportunities to provide valuable feedback (Brusso et al., 2012). Concerns exist regarding factual bias and information falsification by these chatbots (Dwivedi et al., 2023; Firat, 2023). Inadequate data set training, for instance, can result in skewed AI models and outputs that reinforce learners' preconceived notions rather than assisting them in acquiring accurate knowledge.

Recalibrating assessment design: strategies and recommendation

With their capacity for personalised instruction and instantaneous feedback, AI chatbots can transform the educational experiences of students. The difficulty resides in maximising the potential of AI chatbots while mitigating the dangers they pose. This necessitates a reconsideration of assessment design and an emphasis on cultivating higher-order cognitive skills that are resistant to AI manipulation (Pellegrino & Quellmalz, 2010). Assessment designs should therefore engage students with specific tasks that require critical thinking which cannot be easily replicated by Large Language Models like ChatGPT (Rasul et al., 2023; Crawford et al., 2023; Iordanou et al., 2019).

Educators must also endeavour to maintain a human element in their instruction and evaluation, recognising that AI, despite its power, cannot replace human insight and sensitivity (Brusso et al., 2012). In addition to reevaluating pedagogical strategies, regulations and guidelines are required for the ethical use of AI in education. Institutions should educate educators and students about the advantages and disadvantages of AI and devise guidelines to prevent its misuse. To navigate the challenges presented by AI chatbots, evaluations must go beyond simple recall and comprehension tests (Pellegrino & Quellmalz, 2010). Assessments should target higher-order cognitive skills to ensure a realistic measurement of a student's comprehension and reduce the likelihood of AI-assisted responses.

Therefore, to safeguard the authenticity of online assessments in this AI era, a fundamental shift from traditional assessment paradigms is needed. The assessment design should shift away from evaluating students' end outputs, which have a high potential of being repeated by AI chatbots to assessing the students' learning process. Here, we examine some strategies to construct robust, Chatbot-resistant assessments:

Higher-order cognitive skills assessments

The design of our assessments has a significant impact on how students learn and interact with course materials. Traditional assessments frequently emphasise lower-order cognitive abilities, such as recall and comprehension, which are becoming increasingly susceptible to AI chatbot intervention (Pellegrino & Quellmalz, 2010). In a world where knowledge is readily available at our fingertips, or, more precisely, at the command of an AI assistant, these types of assessment tasks are swiftly becoming obsolete.

Higher-order cognitive skills assessments are an effective way to counteract this issue. They focus on skills like analysis, synthesis, evaluation, and creation – skills central to Bloom's revised taxonomy of educational objectives (Anderson & Krathwohl, 2001). Unlike lower-order skills, higher-order skills require a deep understanding of course content, creative and critical thinking, and complex problem-solving abilities (Anderson & Krathwohl, 2001). These skills, thus, are beyond the current capabilities of AI chatbots, reducing the risk of AI-assisted responses in online assessments.

Assessments designed to evaluate higher-order cognitive skills require students to actively engage with the learning material, encouraging deep learning (Marton & Säljö, 1976). For instance, students may be tasked to critique a theoretical perspective, design an experiment to test a hypothesis, or synthesize information from multiple sources to propose a solution to a real-world problem. Such tasks are complex, context-dependent, and often yield multiple viable solutions, rendering them resistant to current AI technology.

Notably, the advantages of evaluating higher-order cognitive abilities go beyond resistance to AI intervention. These assessments are more in line with the ultimate purpose of education, which is to prepare students for a world that is complex and swiftly changing (Reimers & Chung, 2018).

In today's knowledge-based economy, the capacity to analyse complex problems, generate novel solutions, and perpetually learn and adapt is more crucial than ever before (Autor et al., 2003). By emphasising higher-order cognitive skills, assessments not only maintain their authenticity in the age of AI chatbots, but also better prepare students for the challenges of the twenty-first century.

Contextual, problem-based assessments

Contextual, problem-based assessments are a practicable approach to authentic assessment design, especially in light of the growing use of AI chatbots (Gulikers et al., 2004). These assessments require the application of learned concepts to real-world scenarios, a task that requires a unique combination of knowledge, creativity, and critical thinking. Problem-based assessments anchor learning within a context, making it pertinent and meaningful for the learner (Hmelo-Silver, 2004). These duties require students to bridge the distance between theoretical knowledge and practical application, which necessitates a comprehensive and nuanced understanding of the subject. Due to their complexity and context-dependence, these evaluations are resistant to AI interventions, as they require a level of creativity and contextual reasoning that exceeds the capabilities of current AI.

While problem-based assessments offer substantial benefits, they also pose notable challenges. The design of these assessments is significantly more complex than traditional assessment types, requiring careful alignment of problems with intended learning outcomes (Boud & Feletti, 1997). Furthermore, grading can be challenging due to the open-ended nature of responses and the diversity of valid solutions

To mitigate these challenges, educators might consider using rubrics that specify criteria for different levels of performance, allowing for a more objective and structured evaluation of students' work (Jonsson & Svingby, 2007). Further, the use of AI technology could be explored to aid grading by identifying patterns of effective problem-solving or detecting elements of critical thinking within student responses (Siemens et al., 2015). Problem-based assessments, thus, serve as a critical tool in preserving the integrity of online assessments in the era of AI chatbots. Despite their challenges, their value in fostering deeper learning and inherent resistance to AI intervention make them a compelling choice for assessment design in a digital education landscape that is becoming increasingly pervasive.

Portfolio-based assessments

Portfolio-based assessments provide a comprehensive view of a student's learning journey, as they capture progress over time and demonstrate the student's capability across a variety of tasks and contexts (Paulson et al., 1991). Their personalised and longitudinal nature inherently adds complexity to the assessment process, making them more resistant to AI-assisted cheating (Barrett, 2007). However, portfolio-based assessments also present challenges that must be evaluated critically.

A portfolio is a purposeful collection of a student's work that showcases their efforts, progress, and achievements in one or more areas. It might contain a variety of work products, such as essays, projects, self-reflections, peer feedback, and evidence of skill application. In the era of AI chatbots, portfolios offer a unique advantage: they are highly individualised, grounded in the student's personal learning experience, and often involve complex tasks that require higher-order cognitive skills. This makes it difficult, if not impossible, for an AI chatbot to convincingly replicate or assist in creating.

Moreover, portfolios can also provide a multifaceted perspective of student learning, capturing not just what students know, but how they think and how their understanding evolves over time (Barrett, 2007). This is particularly important in a world where the ability to learn, adapt, and apply knowledge in diverse contexts is more valued than the mere acquisition of static knowledge (Dochy, 2001). However, there are several challenges associated with portfolio-based assessments. First, the evaluation of portfolios can be complex and time-consuming, as it requires a holistic review of diverse work products and often involves subjective judgements (Herman & Winters, 1994). Second, developing a meaningful portfolio requires a significant investment of time and effort from students, which may not be feasible in all educational contexts (Snadden & Thomas, 1998). Despite these challenges, portfolio-based assessments offer a robust means of preserving the authenticity of online assessments in the face of AI chatbots. They align with a comprehensive view of student learning, where the focus is not only on what students know, but also on how they think, learn, and apply knowledge.

Collaborative assessments

Collaborative assessments can serve as an effective approach to maintain the integrity of online assessments in the age of AI. They emphasise the social nature of learning, fostering an environment where students construct knowledge through dialogue and mutual engagement (Vygotsky, 1978). Despite their unique potential, collaborative assessments also introduce distinct challenges that need critical evaluation.

Collaborative assessments refer to those where students work together to complete a task or solve a problem. This type of assessment, underpinned by Vygotsky's theory of social constructivism, fosters a rich learning environment where students share ideas, challenge one another's reasoning, and construct knowledge collectively (Vygotsky, 1978). In the context of AI chatbots, collaborative assessments offer an additional layer of complexity. The collaborative process involves negotiation of ideas, empathy, conflict resolution, and mutual engagement – areas where AI chatbots are currently limited (Wooldridge, 2018). Collaborative assessments can manifest in various forms, such as group projects, peer assessment, and collaborative problem-solving tasks. The emphasis is on process as much as product, rewarding students for their collective effort, negotiation skills, and ability to reach consensus (Gillies

& Boyle, 2010). Collaborative assessments align well with 21st-century skills such as teamwork, communication, and intercultural competence, which are vital in our increasingly interconnected and diverse world (Trilling & Fadel, 2009).

However, implementing collaborative assessments comes with its challenges. Accurately assessing individual contributions to a group task can be difficult, potentially leading to 'free-rider' problems where some students benefit from others' efforts (Piech et al., 2013). Furthermore, collaboration may be hindered by issues such as unequal participation, groupthink, and conflicts (Davies, 2009). Finally, the logistics of coordinating group work can be challenging, particularly in large classes or in cases where students are geographically distributed.

Despite these challenges, the benefits of collaborative assessments make them an important consideration for authentic assessment design in the current era of technological advancement. When implemented thoughtfully, they offer a compelling solution to promote deep learning, develop critical 21st-century skills, and uphold the integrity of online assessments.

Implications and future considerations

AI's continual evolution mandates a dynamic approach to authentic assessment design. It falls to educators, institutions, and AI developers to build a balanced ecosystem where technology aids learning, rather than sabotage it. Institutions need to provide training and resources to help educators adapt their assessment designs to this evolving context. Simultaneously, AI developers need to consider educational needs and ethics when designing AI chatbots for educational use (Zhao et al., 2021). As we explore the future of online assessments in the era of AI chatbots, there are several key implications and considerations that educators, administrators, and policymakers must keep in mind:

Pedagogical shift: As technology continues to develop, so must our understanding of learning and evaluation. Authentic assessments that value higher-order cognitive skills, problem-solving, creativity, and collaborative abilities are increasingly necessary to replace rote learning and recall-based assessments (Binkley et al., 2012). This transition will necessitate modifications to the curriculum, instructional methods, and evaluation criteria.

Embracing technology: Educators should not view artificial intelligence chatbots as a threat to the integrity of online assessments, but instead consider how these technologies can be leveraged to improve learning and assessment. AI can automate repetitive tasks, provide personalised learning experiences, and aid in evaluating complex tasks, for instance. However, it is crucial to maintain a balanced approach that employs AI as an instrument to augment human judgement rather than supplant it.

Digital literacy: The extensive use of AI and other digital technologies in education necessitates an increased emphasis on digital literacy. Students must be educated on the ethical use of technology, including the improper use of artificial intelligence chatbots to deceive (Park & Park, 2016). In addition, instructors need training and support to utilise AI tools effectively and comprehend their limitations.

Equity considerations: Concerns about digital divide and equity are raised by the implementation of AI and other advanced technologies in education. Reich and Ruipérez-Valiente (2019) state that not all students have access to the necessary technology, reliable internet, or the abilities to use these tools effectively. Therefore, efforts should be made to ensure that technological integration does not exacerbate existing inequalities.

Data privacy and security: As AI technologies often involve the collection and analysis of substantial amounts of data, concerns about data privacy and security arise. Schools and educational institutions must ensure compliance with applicable data protection laws and employ best practices to safeguard the privacy and security of student information.

Research and evaluation: As novel assessment approaches are created and implemented, ongoing research and evaluation are indispensable. This will enable educators to comprehend the efficacy of various instructional strategies, make well-informed decisions, and continuously improve their practices.

Cultural Context: Cultural contexts have a significant impact on the development and administration of educational assessments. For instance, in Singapore, a nation renowned for its high-stakes, exam-based assessments, this method is profoundly rooted in their Confucian heritage, which values academic achievement and effort. Education is revered as a means of social mobility and success, which explains the prevalence of a rigorous, exam-centered system. This system places a strong emphasis on objective assessments to evaluate students' subject knowledge and comprehension. In spite of the fact that this may foster a competitive academic environment and high global rankings, critics assert that it may hinder creativity and holistic development. Consequently, cultural contexts must be taken into account when designing assessments, as they have a direct impact on the educational values, practices, and expectations of a society.

Conclusion

The advent of AI chatbots has introduced a unique challenge to the integrity of online assessments, leading educators to reevaluate traditional assessment methods. As we navigate this landscape, it is clear that assessments must evolve to maintain their authenticity and effectiveness in promoting meaningful learning. This exploration has underscored the importance of reshaping assessments to value higher-order cognitive skills, problem-solving, creativity, and collaborative abilities. Authentic assessments such as open-ended tasks, project-based assignments, collaborative assessments, and portfolio-based assessments not only align with these values but also pose a significant challenge for AI chatbots to replicate or assist in, thereby preserving their integrity. AI may also aid assessment rather than just being a danger. It can automate repetitive processes, personalise learning, evaluate complicated tasks, and facilitate collaborative evaluations. AI in education should enhance human judgement, not replace it. New issues arise from assessment design and AI application. Digital literacy, educator training, equity, and data privacy and security must be prioritised. Research and evaluation are essential as we alter online exams. This will assist instructors in making educated judgements by continuously refining practices. AI chatbots are not a danger but a chance to restructure our evaluations to make them more real, relevant, and robust. With careful design, thorough analysis, and ongoing evaluation, we can guarantee that our assessments support deep learning and integrity in the digital age.

References

Abdelghani, R., Wang, Y. H., Yuan, X., Wang, T., Sauzéon, H., & Oudeyer, P. Y. (2022). *GPT-3-driven pedagogical agents for training children's curious question-asking skills*. ArXiv.

Amzalag, M., Shapira, N., & Dolev, N. (2021). Two sides of the coin: Lack of academic integrity in exams during the Corona pandemic. Students' and lecturers' perceptions. *Journal of Academic Ethics, 20*, 243-263. https://doi.org/10.1007/s10805-021-09413-5

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics, 118*(4), 1279-1333.

Barrett, H. C. (2007). Researching electronic portfolios and learner engagement: The REFLECT initiative. *Journal of Adolescent & Adult Literacy, 50*(6), 436-449.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Springer.

Boud, D., & Feletti, G. (1997). *The challenge of problem-based learning*. Kogan Page.

Bridgeman, A. J., & Liu, D. (2023, January 23). *How can I update assessments to deal with ChatGPT and other generative AI?* The University of Sydney. https://educational-innovation.

Brusso, R. C., Orvis, K. A., Bauer, K. N., & Tekleab, A. G. (2012). Interaction among self-efficacy, goal orientation, and unrealistic goal setting in predicting personal goals and performance. *Learning and Individual Differences, 22*(6), 747–752.

Chan, C. K. Y., & Tsi, L. H. (2023). *The AI revolution in education: Will AI replace or assist teachers in higher education?*. arXiv preprint, arXiv:2305.01185.

Cheng, G., & Chau, J. (2016). Exploring the relationships between learning styles, online participation, learning achievement and course satisfaction: An empirical study of a blended learning course. *British Journal of Educational Technology, 47*(2), 257-278.

Clarke, O., Chan, W. Y. D., Bukuru, S., Logan, J., & Wong, R. (2023). Assessing knowledge of and attitudes towards plagiarism and ability to recognize plagiaristic writing among university students in Rwanda. *Higher Education, 85*(2), 247-263, https://doi.org/10.1007/s10734-022-00830-y

Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice, 20*(3), 2. https://doi.org/10.53761/1.20.3.02

Daniel, J. (2020). Education and the AI revolution: From chatbots to holograms. *International Review of Research in Open and Distance Learning, 21*(2), 215–231.

Davies, W. M. (2009). Groupwork as a form of assessment: Common problems and recommended solutions. *Higher Education, 58*(4), 563-584.

Dennick, R. (2016). Constructivism: Reflections on twenty-five years teaching the constructivist approach in medical education. *International Journal of Medical Education, 7,* 200. https://doi.org/10.5116/ijme.5763.de11

Dochy, F. (2001). A new assessment era: Different needs, new challenges. *Research Dialogue in Learning and Instruction, 2*(1), 11-20.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M. Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D... & Wright, R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5-6), 304-317.

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching, 6*(1), 57-63. https://doi.org/10.37074/jalt.2023.6.1.22

Gierl, M. J., & Lai, H. (2013). The role of item models in automatic item generation. *International Journal of Testing, 13*(4), 273–298.

Gillies, R. M., & Boyle, M. (2010). Teachers' reflections on cooperative learning: Issues of implementation. *Teaching and Teacher Education, 26*(4), 933-940.

Gulikers, J., Bastiaens, T., & Kirschner, P. (2004). A five-dimensional framework for authentic assessment.

Educational Technology Research and Development, 52(3), 67-86.

Henderson, M., Chung, J., Awdry, R., Mundy, M., Bryant, M., Ashford, C., & Ryan, K. (2022). Factors associated with online examination cheating. *Assessment & Evaluation in Higher Education, 0*(0), 1–15. https://doi.org/10.1080/02602938.2022.2144802

Herman, J. L., & Winters, L. (1994). Portfolio research: A slim collection. *Educational Leadership, 52*(2), 48-55.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn?. *Educational Psychology Review, 16*(3), 235-266.

Iordanou, K., Kuhn, D., Matos, F., Shi, Y., & Hemberger, L. (2019). Learning by arguing. *Learning and Instruction, 63*, 101207. https://doi.org/10.1016/j.learninstruc.2019.05.004

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144.

Lancaster, T., & Cotarlan, C. (2021). Contract cheating by STEM students through a file sharing website: A Covid-19 pandemic perspective. *International Journal for Educational Integrity, 17*(1), Article 1. https://doi.org/10.1007/s40979-021-00070-0

MacNeil, S., Tran, A., Mogil, D., Bernstein, S., Ross, E., & Huang, Z. (2022, August). Generating diverse code explanations using the gpt-3 large language model. In *Proceedings of the 2022 ACM conference on international computing education research-volume 2* (pp. 37-39).

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British Journal of Educational Psychology, 46*(1), 4-11.

Park, C. N., & Park, S. M. (2016). The conceptual model on smart city competitiveness: A focus on human and social capital. *Journal of Open Innovation: Technology, Market, and Complexity, 2*(1), 1-15.

Paulson, F. L., Paulson, P. R., & Meyer, C. (1991). What makes a portfolio a portfolio? *Educational Leadership, 48*(5), 60-63.

Pellegrino, J. W., & Quellmalz, E. S. (2010). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education, 43*(2), 119–134.

Perkins, M., Basar Gezgin, U., & Gordon, R. (2019). Plagiarism in higher education: Classification, causes and controls. *Pan-Pacific Management Science, 2*, 3–21. https://doi.org/10.13140/RG.2.2.20694.11841

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). *Tuned models of peer assessment in MOOCs. arXiv preprint,* arXiv:1307.2579.

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2015). Modeling how students learn to program. *Proceedings of the 36th international conference on software engineering, 37*(1), 147-156.

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56, https://doi.org/10.37074/jalt.2023.6.1.29

Reedy, A., Pfitzner, D., Rook, L., & Ellis, L. (2021). Responding to the COVID-19 emergency: Student and academic staff perceptions of academic integrity in the transition to online exams at three Australian universities. *International Journal for Educational Integrity, 17*(1), Article 1. https://doi.org/10.1007/s40979-021-00075-9

Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science, 363*(6423), 130-131.

Reimers, F., & Chung, C. (2018). Preparing students for a rapidly changing world. In *Teaching and learning for the twenty-first century* (pp. 57-74). Harvard Education Press.

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18*(54), 1-23. https://doi.org/10.1186/s41239-021-00292-9

Shermis, M. D., & Burstein, J. C. (Eds.). (2013). *Handbook of automated essay evaluation: Current applications and new directions.* Routledge.

Siemens, G., Gašević, D., & Dawson, S. (2015). *Preparing for the digital university: A review of the history and current state of distance, blended, and online learning.* LINK Research Lab, University of Texas Arlington.

Snadden, D., & Thomas, M. (1998). The use of portfolio learning in medical education. *Medical Teacher, 20*(3), 192-199.

Stahl, B. C., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., ... & Wright, D. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review*, 1-33. https://doi.org/10.1007/s10462-023-10420-8

St-Onge, C., Ouellet, K., Lakhal, S., Dubé, T., & Marceau, M. (2022). COVID-19 as the tipping point for integrating e-assessment in higher education practices. *British Journal of Educational Technology, 53*(2), 349-366.

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching, 6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times.* Jossey-Bass.

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior, 27*(1), 118-130.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Walsh, L. L., Lichti, D. A., Zambrano-Varghese, C. M., Borgaonkar, A. D., Sodhi, J. S., Moon, S., ... & Callis-Duehl, K. L. (2021). Why and how science students in the United States think their peers cheat more frequently online: Perspectives during the COVID-19 pandemic. *International Journal for Educational Integrity, 17*, 1-18.

Wegerif, R. (2006). Dialogic education: What is it?. *Educational Futures, 1*(1), 61-70.

Wooldridge, M. (2018). *The road to conscious machines: The story of AI.* Penguin UK.

Zhao, Y., Ni, L., & Li, D. (2021). Unpacking AI education for K–12: Comprehensive, equitable, and effective. *Educational Technology Research and Development, 69*(3), 953-976.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Neo-neoliberalist capitalism, intensification by stealth and campus real estate in the modern university in Aotearoa/New Zealand

Martin Benedict Andrew[A]

[A] *Associate Professor of Work-Based Learning, Otago Polytechnic, Dunedin, Aotearoa/New Zealand*

Keywords

Aotearoa/New Zealand;
campus;
change management;
neoliberalism;
reform;
responsibilisation;
workload.

Correspondence

martin.andrew@op.ac.nz [A]

Article Info

Received 9 July 2023
Received in revised form 31 July 2023
Accepted 31 July 2023
Available online 2 August 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.16

Abstract

This article critiques media commentary on reforms in higher education (HE) and the polytechnic or vocational education and training (VET) in Aotearoa/New Zealand during 2023. It used as its epicentre Gaston's (2023) opinion piece, particularly on the current round of redundancies, rationalisations and restructurings in the sector in our 'post-COVID-19' world. Her suggestion that the reforms as we see them today lead me to explore the origins of neoliberalist ideology in tertiary education in the country, seeking a definition to guide how Aotearoa/New Zealand, in particular, has experienced and manifested neoliberalist policy. The ideology's emphasis on work intensification and responsibilisation at both the level of the individual and the organization lies at the heart of the rationale. Since 2023 has been a year of mysterious budget holes for many universities and for the emergent super-Polytechnic Te Pūkenga/New Zealand Institute of Skills and Technology, I investigate how the media and public domain materials report the financial state of our tertiary providers. Generating a case study of the University of Otago's property portfolio, which grew by two billion dollars between 2012 and 2022, I tap into recent critiques of the university as a real estate portfolio and campus as literally 'flat land' that has the potential to become a domain less in the name of education than neoliberalist capitalism.

Intensification by stealth

In 2023, tertiary education (TE), including vocational educational training (VET), was constantly getting bad press in Aotearoa/New Zealand (hereafter Aotearoa). Critique after critique name-checked the usual suspects: falling student enrolments, the obsolescence of languages and arts, strike action and unionism, the present government's under-investment, the previous government's under-investment and of course the lingering impact of the world's favourite scapegoat, COVID-19. The VET sector suffered the triple whammy of RoVE (Reform of Vocational Education), the circus of the creation of Te Pūkenga (aka New Zealand Institute of Skills and Technology or 16 Polytechnics centralised in one entity) and COVID and: the entire sector lost its international student cash cow. But Aotearoa, with its 'number 8 wire' 'can-do' cultural mindset, coped (Kiwis will tell you that the number 8 wire is a national symbol of — or, more accurately, a metonymy for — innovative ingenuity [Te Ara, 2010]). The current narrative is that, while higher and vocational education under COVID-19 may have led to some negative educational experiences both for those in higher education and those considering entering it (that is, school-age senior students), the pandemic positively impacted pedagogical innovation, social adaptation and human resilience. To close this opening paragraph, I cite a resonant and typical Aotearoa academic's experience of the COVID 'era':

> I was mostly plugging in daily to an increased workload of responding to anxious students and colleagues. Literally, we (and others globally) transferred our face-to-face teaching, supervision, and service responsibilities onto a digital platform and plugged into Zoom, Google-docs, emails, Facebook messenger, and other platforms... The equity divide became more apparent. The privilege of still having a job, although with increasing workloads, plus the precarity of the job situation, wreaked havoc on our sensibilities— and from New Zealand, the strangeness of it all was illuminated daily on Facebook and via Zoom meetings with international colleagues, family, and friends (Fitzpatrick, 2021).

Amidst all of the 'workload' (intensification is the main theme of the passage), she found time for joy in work. The online intensity was just that: intensification of workload. The facts that e-education is not for everyone (LaPointe & Reisetter, 2008) and that *kanohi ki te kanohi* (face-to-face) learning supports *kaupapa Māori* (Māori research approaches) (O'Carroll, 2013) have long been known in the literature. Education in person, in community, is good.

Yet we also learned that digitization is not just a mode now; it is a product: Massey University's 'Digital Plus' policy may cut back teaching of disciplines in-person at satellite campuses (Tertiary Education Union, 2023a), but is dangerous in prefiguring the fear that synchronous humans can now be replaced by (as opposed to supplemented by) asynchronous versions of themselves. I spoke at a conference with a learner who discovered her online professor had been dead since 2019. We are squarely back in the space of precarity Standing (2011) had defined (Blackham, 2020; Spina et al., 2023), where increasing numbers of educators are Temporary Tutors or Teaching Fellows, "undervalued, overused, and stigmatized" (Solomon & Du Plessis, 2023). It may be reductionist to say this, but even our current disruption by Artificial Intelligence (AI) has massification and commercialism at its core and may yet harbour sinister potential, possibly in the form of *superintelligence*, as Popenici (2023) details. Popenici et al. (2023) say: "Artifical intelligence is a marketing concept... It's not a real thing" (p. 2). HE stakeholders, including academics, need to exercise the critical thinking they sell to their learners a little more deeply themselves. Popenici et al. (2023) voice a topos of the subgenre of the ruined university: "Universities are under attack by neoliberalism and the obsession to make education a business and reduce all to profits and markets" (p. 2). This article concludes by exploring an underexamined facet of university business: real estate.

It was not, then, such a surprise to see a return to the old neoliberalist ideology: *let's get back to campus and put those bums back on seats like it's 1984*; in other words, neo-neoliberalism, or return to normal by default. But look closely now: the catalyst of COVID, Saado-Filho (2021), discerned, changes "the transformation of the *crises in neoliberalism into a crisis of neoliberalism*" (p. 186). In Aotearoa, what we have witnesseed to date is a return to the old normal, not any new normal, and as Martin Parker, for one, would maintain, this is a worst-case scenario (Parker et al., 2021): "the #buildbackbetter ideas are having a very marginal effect. And substantially, we're just moving back to a slightly lower carbon version of the same system" (p. 61).

A key factor in the post-COVID-19 decline in morale in HE and VET in Aotearoa was the Labour government's "COVID-19 budget" of 2020, where investment in education and research was "not merely in action, but in aspiration" (Gaston, 2023). Despite a 2018 budget to elevate research funding to 2% of the gross domestic product (GDP), only 1.47% was ever achieved, indicating an obvious blackspot of under-investment. University leaders watching the money-flow carefully can be forgiven for thinking the pot still had something in it. Those who have lived through decades of higher education reform in Aotearoa, however, are aware, as Gaston (2023) wrote online: "the crisis in Aotearoa's university and wider research sector did not happen overnight. While funding shortfalls and sweeping redundancies are now making headlines, the underlying problems have been evident for years". We will go back there shortly, but first we will look at some numbers to illustrate the human impact of under-investment.

The numbers of mooted 2022-23 redundancies we know about include Te Pūkenga (400 roles in a VET mega-merger (Schwanecke, 2023); update; to 950 in July (Simmonds, 2023); Massey University (70 non-academic roles as part of a 'funding black hole' (Gerritson, 2022); Auckland University of Technology (AUT) (170 academics); Victoria University/ Te Herenga Waka (up to 275 roles or sufficient to cover a \$33m deficit blamed on government underfunding); and Otago University ("several hundred" or sufficient to cover a \$60m oversight). The story of how AUT had to postpone its redundancy round due to failure to follow process in 2022 is now moot; neoliberalism always overcomes legal technicality.

Otago University's drop is amusing, given the reported statistics that international enrolments increased by 495 and domestic enrolments were down by 670 (*Radio New Zealand*, 2023b) and given the reference to "the expanding student roll" in the preface to the university's real estate master plan (Otago University, 2023a). We know the numbers are approximate and that there will always be other historic factors like the reported 700 job losses nationally in 2021 due to the international student COVID-19 shutdown. But does the loss of 170 students really correspond to a reported shortfall of $60 million? More recent reporting said, "we don't have a roll drop any more" (*Otago Daily Times,* 2023), 2023). Where did the money really go?

The universities put out an SOS to the government to raise the debt ceiling (Tertiary Education Union, 2023b). On June 26, 2023, the government announced $120 million to increase tuition subsidies at degree level and above and also help research capability, not applicable until 2024 and not a bail-out, but perhaps an election bribe (the election will be on 14 October 2023): it the most significant increase in 20 years (*Radio New Zealand,* 2023a). The money came not from the above underspend but from an appropriation from *Vote Tertiary Education,* a scheme/report comprising the sums from the 2022 budget cycle (Treasury, 2022). It can, therefore, be best seen as not a bonus but a 'correction' (Tertiary Education Union, 2023b). A simple but biting distinction was made between institutions that managed their money and those that did not (*Radio New Zealand,* 2023a).

The entire research-practice nexus was put under erasure in the HE and VET sector. Simultaneous with the announcement impacting HE research, the role of research in the VET sector, if there is still even to be one, is under investigation as part of the merger of 16 regional Polytechnics into a mega-entity, Te Pūkenga. The mechanism for research funding, the Performance-Based Research Fund (PBRF), with a national budget currently capped at $315m, would be scrutinised too, even though it was last reviewed only in 2020 with its more "capacious" definition of research for "vibrant, diverse research cultures" (PBRF Review Board, 2020, p. 8). Apparently, post-COVID-19, it is not "fit for purpose", but what, 25 years after its inception, is that purpose? A neo-neoliberal reform where co-governance meets austerity? And what was it ever? A neoliberal reform of responsibilisation. Capacity in research in VET was explicitly protected (PBRF Board, 2020), but will this protection remain? It's postponed until 2026, so anything may happen.

The cost of the Te Pūkenga merger is reported to exceed $420 million (Simmonds, 2023). The announcement of the $120 million (for HE) was preceded by an announced $420 million (a very coincidental number) over four years for the Polytechnic sector as a post-COVID measure and singled out as beneficiaries were Māori, Pacific Island people and women. These monies were for those supporting the workforce, not the providers. The scheme pays employers $500 per month for the first two years of an apprenticeship (*One News*, 2023). It remains to be seen how this might be implemented during a time of turmoil for the sector but would seem to be cut from the same ideological cloth as the COVID-19 subsidies: an idea curating business monies over those of individual students or workers. The prime ministerial rhetoric was unsurprising: "Apprenticeship Boost is a great example of how Government can get alongside and support business to invest in their next generation of talent. It's a win-win" (*One News,* 2023). The message for the class of 2023/4 is clear: plumbing good; German and Music (to name the two usually-most-hard-hit disciplines) bad.

Since the Orwellian year of 1984, or thereabouts, the story retracts to an era of neoliberalist capitalism, under which the intensification of academic labour gained a permanent mandate. With this intensification, every year, workload models were tweaked upwards, forcing ever more hours into spreadsheets and definitions of full-time hours were continuously revised to as to reduce hours for preparation, marking, research, scholarship and service. Equations squashing reduced research allotments into Excels meshed with those calculating EFTS, "Equivalent Full Time Student," a measure cast as being to help students settle their workloads but, in fact, weaponised in the battles over staff workloads. The university's process of ongoing intensification to protect the proverbial but mysterious 'bottom line' resulted in tertiary educators literally working themselves to death (Fleming, 2021), while a bloated middle-management siphon off the bucks (Simmonds, 2023).

The *reductio ad absurdum* of intensification has been the post-2020 appearance of job advertisements for 'Teaching Fellows'. These are usually hot-off-the-PhD early career educators, either bright-eyed at getting on the ladder or more worldly permanent members of the precariat who accept that research will not be such an inscribed expectation in the neo-neoliberalist or post-COVID era. Often overheard in corridor conversations is a reference to the fact that The Education Act (1989) mandates that all degree-level educators must be research-active (Gaston, 2023). In fact, in times of change, this nexus is the holy grail, the one fixed point in any submission process.

However, what happens when you do not keep your eye on the ball? The updated 2020 Education and Training Act, entered under stealth when all eyes were on COVID-19, simply says that the "characteristics" of universities are that courses are taught by those involved in "advancing knowledge"; in short, this 2020 update paved the way for the generation of Teaching Fellows, their salaries approximately two-thirds those of traditional academic. The idea of the research-active academic will, I fear, increasingly become a relic. Now, there already, is a massive saving to the bottom line; but under neo-neoliberalist post-COVID capitalism, more cuts must be made. Not cuts from universities' massive property portfolios or Vice Chancellor-type packages, but from staff redundancies, and when the expensive academics are out, we can, at least until the digital lecturers are ready, replace them with more Teaching Fellows. Before we hop to this dystopian future, we need to return once more to the past to identify the problems Gaston (2023) says have been evident for years. I continue this discussion with an attempt to define the 80s neoliberalism in which today's neo-neoliberal university leaders were indoctrinated.

What is neoliberalist capitalism?

Defining neoliberalism in 2023 is notoriously difficult owing to strong reactions, pro or con and due to its different manifestations and mutations and the unavoidably emotive positionality of authors. Giroux (2008, p. 9) cast neoliberalism as "a broad-based rhetorical and cultural movement designed to *obliterate* public concerns and *liquidate* the welfare state". McPhail and McNeil (2021) see neoliberalism as best viewed as a confluence of such theories as monetarism and supply-side economics. My emphasis here is on hegemonic neoliberal capitalism (McMaster, 2013). Without the capitalist element, writers like Kelly (2020) might get away with such a statement as "work intensification does not necessarily stem from neoliberalism" (C2). Using sweeping or fit-for-purpose definitions runs the risk Hofmeyr (2021) warned, of being "vacuous and unenlightening" (p. 591). Kelly (2020) does open the possibility that neoliberal is experienced differently by students and staff, about whom the bulk of the literature relates. Everyone will take a side here, but Hofmeyer's well-considered 2021 definition suffices:

> Neoliberalism may be understood as the globalized and globalizing political programme that espouses economic liberalism or 'laissez faire economics' as the only means of promoting economic development and securing political liberty (p. 591).

Extending this definition, we can say that 'neoliberalist capitalism', usually shortened to 'neoliberalism' and synonymous with New Public Management (NPM), denotes the free-market-driven government policies promoting the deregulation and privatisation characterising Thatcherism and Reaganomics. Whitty et al. (1998) wrote that, in the UK, public-sector institutions operated in the manner of private-sector ones. Under neoliberal ideology, private (individual/family) decision-making replaced community-level political or professional judgments but was underpinned by corporate globalism. The individual is free from social constraints or responsibilities to the community (McMaster, 2013). Simultaneously, in HE, collective resistance was forestalled (Davies & Petersen, 2005). The trend reverberated through other countries where the ideology spread like a virus to nations such as Aotearoa. Simultaneously it spread to Howard's investor state in Australia (Redden, 2017) and, in Aotearoa, both 'Rogernomics' (after Roger Douglas, Finance Minister, 1984-1988 and progenitor of the Commerce Act, 1986) and 'Ruthanasia' (after Ruth Richardson, Finance Minister, 1990-1993) in Aotearoa.

This fight against the welfare state was known as the New Zealand experiment (McMaster, 2013; McPhail & McNeil, 2021). However the ideology impacted, it manifested itself everywhere as free market thinking and investor ideology. With it came the ethos of individual responsibilisation that blames individuals for their ill health, unemployment or fiscal choices and normalises such repellent stereotypes as the dole-bludger, the domestic purposes benefit parasite and the lazy Māori. This 80s/90s rhetoric of responsiblisation is echoed by 2023's economic *autonomy for universities catch-call*, which has led to 2023's division of the good and naughty universities, cast like biblical wise and foolish virgins. Heckled by distraught students at Otago University who thought universities were 'spiralling down', Prime Minister Chris Hipkins said: "the universities make their own decisions about how they manage their finances, so it's not something we can intervene on as a government" (Brunton, 2023). By law, "universities are autonomous, publicly funded institutions. Along with institutes of technology, polytechnics, and wānanga, universities are Tertiary Education Institutes (TEIs) under the Crown Entities Act 2004" (Te pōkai tara/Universities New Zealand, 2023). Despite officialdom proffering public funding, there is an ongoing hegemonic neoliberalist battle over the word 'autonomy'. If you have a funding shortfall, it's your own fault.

Responsibilisation is identified as a driver to develop Otago University's new approach to campus planning: *the Learning Landscape*: "governments will continue to pressure universities to seek more diverse sources of income that reduce the reliance on government funding and increase the commercialisation of the education sector" (Otago University, 2023a, p. 20). The fact that COVID-19 *interacted* with structural deficiencies or systematic inequalities differently in different locales is white noise. The 'shady spaces' of Kiwi fair-go 'cultural resistors' that may have appeared to have lessened the impact of full-blown neoliberalism on Aotearoa (McMaster, 2013) are much darker spaces under neo-neoliberalism. This is because it lives under the ever-watchful eye of emergent populism informed by an extreme right, not a new right, making the hegemonic base look very inconsistent and shaky indeed. Totally obscured from this vision is where the money really went, but we suspect it has to do with free market thinking and investor ideology.

In Aotearoa/New Zealand, both the investor state and the Commerce Act leave a legacy of social inequality (Bertram, 2020a, 2020b; Redden, 2017). This places what Bertram (2020b) calls an "iron cage" around policy, impacting education and its marketized efficiency agenda which treats consumers' well-being and potential learners' access to education as "irrelevant", evoking Noam Chomsky's ironic 1999 catch-call of *profit over people*. In Australia, Andrew et al. (2020) dissect the "straitjacket" of neoliberalism knee-capping public response to COVID-19. Neoliberalism has increased the frequency and impacts of "the intensification of inequality, the complexity of financial markets, the rise of digital monopolies and, above all, the twin crises of climate change and biodiversity loss" (Andrew et al., 2020, p. 767). The prison and straitjacket metaphors recall Giroux's (2007) critique, *The university in chains*. Of neoliberalist logic in HE, he writes (2009, p. 673):

> Within this discourse everything is for sale, and what is not is relegated to relative invisibility. The traditional academic imperative to "publish or perish" is now supplemented with the neoliberal mantra "privatize or perish" as everyone in the university is transformed into an entrepreneur, customer, or client, and every relationship is ultimately judged in bottom-line, cost-effective terms.

Hofmeyer (2021) stresses that in critiques of neoliberalism, "analyses need to be distinct, local studies that are context- and time-specific" (p. 591). In an Aotearoa context, Bertram (2020a) sees the regulatory failure of neoliberal logic as repeatedly inevitable in crises: he namechecks the leaky homes 'scandal', the Pike River 'disaster,' the 'outcry' over electricity company profits, the collapse of finance companies and ongoing workplace accidents and deaths as moments of crisis. The year 2023 has already brought an extended cost of living 'crisis' impacted by a bank excess profit 'crisis' and a housing crisis caused by a shortfall of affordable accommodation. In each case, Bertram (2020a, 2020b) writes, there were (and are) 'victims' of corporate power, their voices again dismissed as white noise and ultimately irrelevant. This is an enaction of the dystopic manta, *Profit over people* (Chomsky, 1999).

In higher education contexts, there were impacts beyond just labour intensification in the form of "overloading of responsibilities" (Ball, 2003, 2012; Shore, 2010, p. 20). The scene was set for the toxic university (Smyth, 2017). There were rules for educators to publish increasingly in quality (e.g. Scopus-indexed or high impact factor) journals, 'precariousness' for even tenured academic workers (Standing, 2011; Solomon & Du Plessis, 2022), inequity of access to the education market for learners, and health crises for both staff and students (Tregear et al., 2022), "suicidal thoughts, depression and incapacitating anxiety" (Fleming, 2021, p. 25), impossible workloads and stress-related illnesses (Shore, 2010). Management functions envelop audits for 'accountability' and assurance regulations for 'quality' (Craig et al., 2014). Ball (2003, p. 224) declared: "Performance has no room for caring". The Faustian cost for regulators of the efficiency agenda would be their soul (Berkovich, 2018; Minina, 2018)—if they had one. As Ball (2003) allegorised, many educators were forced to give theirs up, Faust-style, in the game of performativity and annual evaluation. In Aotearoa, however, the issue is less speculation as to the presence of souls in neo-neoliberals than the obvious absence of government- or organisation-level leaderly strategic vision (Gaston, 2023) and the stymying of insider voices who lived and breathed higher education under the increasingly dominant populist anti-elite agenda.

Where has all the money gone?

As more and more universities, internationally and locally, restructure, shed disciplines and staff, you can hear people ask, like a Pete Seeger lyric, where has all the money gone, long time passing? In Aotearoa/New Zealand, we can start with the fact that the estimated salaries of Vice Chancellors (Auckland c. $768,000; Otago c. $656,000) exceed those of the Prime Minister (c. $470,000). We might wonder at the statistic that Otago University has a reported ratio of 1.9 professional/general staff to every teaching staff member (Elder, 2017). We might wonder about legal costs as those in contested and reconfigured positions defend their spaces. We can also stop and wonder at Victoria University's spending half a million to rebrand to Victoria University of Wellington/Te Herenga Waka (Te, 2020): "In an Official Information Act response, the university said it expected expenses to be 'more than recouped' through the gains it made as a result of the brand refresh". In a paywalled article on the 2023 redundancy round, the Vice Chancellor proxy demonstrated tight-lippedness on the topic of finances: "The balance of extra revenue opportunities is confidential as it is commercially sensitive business information" (Littlewood, 2023a) while the 107 redundancies to date, seen as a $10 million saving, were enumerated.

With accusations of tone deafness ringing, on March 23, 2023, Otago University set out to do the same thing to a more complex response. They were changing the ingoa Māori (Māori 'name') from Te Whare Wānanga o Otāgo/ The University of Otago, to Ōtākou Whakaihu Waka/A Place of Many Firsts (McNeilly, 2023). Gifted by the local people (Ngāti Whātua ki Ōrākei), the metaphor is grounded in *whakapapa* ('heritage') and corrects the provincial *ingoa* ('name') 'Otago' to 'Ōtākou'. The ingoa embraces the university's place as a *kainga* ('home') for the mana *whenua* ('people of the land') and *manuhiri* ('guests'), who collaborate there on educational journeys (de Silva, 2023). De Silva (2023) cites the mana whenua co-design team about the word *Whakaihu*: "it references Otago University as the *motu's* ('nation's) first university and its recognised international standard of academic excellence. Yet it also references the students as champions of their *whānau* ('families') and communities". The decision is compatible with both governmental Statements of National Education and Learning Priorities (NELP) and the Tertiary Education Strategy (TES) to establish priorities for education that will ensure the success and wellbeing of all learners (Ministry of Education, 2023). The rationale is compelling, but there were now four issues: the timing, the cost, the fact that rebranding is something corporations do — and the honouring of the 1840 national treaty, Te Tiriti o Waitangi as per their strengthened Māori Strategic Framework, launched as part of Otago University's *Vision 2040* (Otago University, 2023b).

The timing and cost came at a bad time: Otago University announced a 60m budget 'hole' and a drop in roll cited by the acting Vice Chancellor as the main suspect (McNeilly, 2023). Those still in trauma remembered two years prior when 150 staff were axed in the previous restructure, so the reminder of the corporate nature of the rebrand stung nerves. Populist rhetoric around co-governance/ partnership, seen by the university in *Vision 2040* as the foundation for meaningful and sustainable relationships, put it on a par with critical race theory in the States, a rough approximation, but it serves. The fact that King Charles's coronation had just happened made the republican debate resurface, and as of mid-2023, co-governance had already become a hot *kumara* (Māori sweet potato and another national symbol) with a looming election.

The problem was not merely the expenditure (reportedly $670,000, consultation fee of $126,000 included, but still a figure that barely covers the stationery), but also the 'wokeness' of it all. Later in 2023, the absentee but-on-extended-sick-leave new (17 months) Vice Chancellor David Murdoch resigned, triggering another expensive international spin of the VC Merry-Go-Round. Te Whare Wānanga o Otāgo's long-planned *Vision 2040* (Otago University, 2023b) launch, the expensive footprints of consultants more vividly on display with the budget crisis, was drowned out by the white noise. Where, indeed, does all the money go? The trail on both universities' rebranding went cold until Otago announced its rebranding was full-steam ahead. Three-quarters of submissions were in favour (Ellis, 2023). More good news came to deflect the eye: the Tertiary Education Commission gave the university strong scores for "student qualification completion (74%), course completion (88%), first-year retention (83%) and progression (93%)" (Littlewood, 2023b). But rebranding (and the spin that inevitably follows) is just one topical site where neoliberalism meets culture wars. Nowhere is this the question of where the money goes more mysterious than in universities' acquisition of and investment in real estate — seemingly over quality education. Are our universities real estate portfolios with education on the side?

Here's a case study. At Otago University, a four-person Property Management team acts as 'landlord', managing the University's ostensible $1.6 billion property inventory. In 2012, Otago University's assets eclipsed those of all universities nationally and were reported as including $855 million in buildings, $205.9 million in land, $19.4 million in cash/cash equivalents and $9.4 million in rare books, manuscripts, and artworks (Elder, 2021). With expert consultants (cost certainly in an annual general report), the university enacts a 25-year master plan which includes "acquisition of adjacent properties where possible" (p. 109) within an extensive 8-precinct zone. Here, in fluent 'corporatese', is the vision: "we create, develop and maintain the University's campuses in a way that inspires and supports excellence in all who experience them". Ostensible goals of sustainability and pedestrianisation are pleasing on paper but hardly gel with a culture where owning and parking a car has increasingly become a student's individual right and choice since around 1984. In addition to bisecting arterials and s-bend state highways, the university's current major bugbears are the historic Kelsey Yaralla Kindergarten, around which plans need currently to be diverted, and the Cumberland Court Motel, "and it is recommended that no development *be countenanced on this site that would preclude* [the university's] option" (italics to emphasise the formality; p. 123). "Significant costs" will be incurred by correcting the S-bend highways (p. 117) in this 25-year plan, so a Plan B is in place: a new pedestrian mall with a pedestrian bridge over the biggest arterial.

There are master plans for Christchurch and Wellington, too, largely for medical schools, again predicated on growth in student numbers: "For the purpose of Master Plan scenarios, the University has forecast a potential increase of 212 postgraduate EFTS (coursework) and 140 postgraduate EFTS (research) by 2030" (p. 151). The Master Plans identify numerous desired sites. In any statement of financial position released, the category 'property, plant and equipment' towers above any other category as a "non-current asset". The university's interests in developments around the campus were reported as "underpinning" the commercial property market in the area (Elder, 2012). A nest egg of $2 billion was described, "mostly made up of property, plant and equipment" (Elder, 2017). Nowhere in the recent public domain is information on the estimated cost of the acquisitions and corrections decipherable by non-accountants, nor of the costs already incurred in more

Figure 1. The University of Otago. Photo by the author (30/07/23).

recent acquisitions and in ongoing consultancies. The annual general reports are extensive and dense, and there is no better place to bury data than in plain sight, and the devil is always, usually literally, in the detail.

The potential property portfolio of the university is speculative locally and nationally but is part of an internationally normal practice. This practice is seen as impacting the availability of affordable accommodation, a palpable trend in Dunedin, the city that houses the University of Otago. In *The University in Ruins*, Readings (2011) drew a potent analogy between the ruined city and the university: Neither longing for the past nor present consumerism bring our disparate disciplines' values back together. With reference to the University of Chicago, Haar (2011) demonstrated the parallel nature of developments in pedagogical transformation and a location for the larger purpose of the academic community.; the interrelationship of knowledge and urbanism. *Campus*, the Latin for 'field', becomes in itself an ideal "univerCity", a planned city in microcosm (Turner, 1987).

The case study of Otago demonstrates a desire to extend its campus utopia of historical and bespoke architecture and precincts into the city itself. While universities both in Aotearoa and the States are non-profits, Grabar (2018) wrote online, "you would be hard-pressed to find a [university] that has *not* thrown itself into commercial property development". In the notion of the campus, we witness a blurring of the not-for-profit and the profit-driven. Indeed, economic development becomes the academy's main focus to the point that educational institutions no longer exist for the public good, for equity. Since a sustainability agenda is writ large, as in our case study, we cannot say that neoliberalism is "unsustainable in view of the imperative of protecting the known forms of life on Earth" (Saad-Filho, 2021, p. 181). Baldwin (2021) writes about how universities are annexing and plundering our cities' real estate. Mission creep is at work: their core business is not education but profit-driving real estate, the funding arms hidden from public scrutiny. If the education arm loses money, don't cross-subsidise; cut it off. Economic development is the modern university's focus (Grabar, 2018).

Anti-conclusion

Where has all the money gone? And where is evidence of universities' ability to self-critique? It would be premature to suggest Aotearoa/New Zealand universities would sooner lay off staff, including educators, than reduce their empire-building ambitions. It would be merely a hunch to suggest that monies invested in real estate are more economically productive in that form to stakeholders than they would be as educator salaries. It would also be amiss to posit any link between housing shortage and university property buy-ups, though the languaging cited above about Cumberland Court near Otago University is a compelling example of 'hands off that property, the university has first option'. Any suggestion that universities should deprioritise image building as rebranding during staffing crises is merely something a reader might infer. I certainly have not suggested that there may be substantial legal costs in the current landscape of layoffs, redundancies and sudden departures. And any insinuation that one Vice Chancellor is worth six or seven teaching professors is simply beyond the scope of this paper. Whatever you may infer from Otago University's 2b increase in property portfolio between 2012 and 2023 based purely on published figures is your own business. Neo-neoliberalism not only brings in work intensification by stealth; it also blurs the university's function as a fair employer/education provider/real estate portfolio holder. The scholarship might say the HE sector has lost its heart as well as its soul, sucked by neo-neoliberals. How do we, in Fitzpatrick's words (2021, p. 796), "'kick-start' our heart beat again"?

References

Andrew, J., Baker, M., Guthrie, J., & Martin-Sardesai, A. (2020). Australia's COVID-19 public budgeting response: the straitjacket of neoliberalism, *Journal of Public Budgeting, Accounting and Financial Management, 32*(5), 759-770.

Baldwin, D. L. (2021). *In the shadow of the ivory tower: How universities are plundering our cities.* Bold Type Books.

Ball, S. J. (2003). The teacher's soul and the terrors of performativity, *Journal of Education Policy, 18*(2), 215-228.

Ball, S. J. (2012). Performativity, commodification and commitment: An I-spy guide to the neoliberal university, *British Journal of Educational Studies, 60*(1), 17-28.

Berkovich, I. (2018). Policy makers' regulatory mind-set, the regulatory governance body of the system, and the regulated soul of educators. *Comparative Education Review, 62*(2), 178–198. https://doi:10.1086/696822.

Bertram, G. (2020a). Why the Commerce Act 1986 is unfit for purpose. *Policy Quarterly, 16*(6), 80-87.

Bertram, G. (2020b, September 8). Breaking out of neoliberalism's "iron cage". *Newsroom*, https://www.newsroom.co.nz/ideasroom/breaking-out-of-neoliberalisms-iron-cage

Blackham, A. (2020). Unpacking precarious academic work in legal education. *Law Teacher, 54*, 426–442. https://doi.org/10.1080/03069400.2020.1714276

Brunton, T. (2023, June 2). Government can't stop University of Otago cutting staff. *Radio New Zealand*. https://www.rnz.co.nz/news/national/491235/government-can-t-stop-university-of-otago-cutting-staff-pm

Craig, R., Amernic, J., & Tourish, D. (2014). Perverse audit culture and the modern public university. *Financial Accountability and Management, 30,* 1–24.

Chomsky, N. (1999). *Profit over people: Neoliberalism and global order.* Seven Stories Press.

Davies, B., & Petersen, E. B. (2005). Neoliberal discourse in the academy: The forestalling of collective resistance. *Learning and Teaching in the Social Sciences*. http://ezproxy.uws.edu.au/login?url=http://dx.doi.org/10.1386/ltss.2.2.77/1

De Silva, T. (2023, March 15). The process to rebrand our oldest university. *The Spinoff*. https://thespinoff.co.nz/atea/15-03-2023/the-process-to-rebrand-our-oldest-university?fbclid=IwAR3ynPHdopiCFdbJgvTZX-MQLnWEHoDRRDm2YM35i-iwv48SfFOS_NVxPdg

Elder, V. (2012, May 29). Otago University worth $1.4 billion. *The New Zealand Herald*, https://www.nzherald.co.nz/nz/otago-university-worth-14-billion/A6MMV2PW4KK3UZBQNEJHLMQCNM/

Elder, V. (2017, June 1). Otago second richest uni. *The Otago Daily Times*. https://www.odt.co.nz/news/dunedin/otago-second-richest-uni

Ellis, F. (2023, July 12). Uni brand gains assent. *Otago Daily Times*. https://www.odt.co.nz/news/dunedin/campus/uni-rebrand-gains-assent

Gerritson, J. (2021, March 1). International student downturn costs 700 university jobs. *Radio New Zealand*. https://www.rnz.co.nz/news/national/437379/international-student-downturn-costs-700-university-jobs

Ellingham, H. (2022, November 10). Massey university looks to cut more than 70 non-academic jobs. *Radio New Zealand*. https://www.rnz.co.nz/news/national/478430/massey-university-looks-to-cut-more-than-70-non-academic-jobs

Fitzpatrick, E. (2021). #Plugging into hope. *Qualitative Enquiry, 27*(7), 790–797.

Fleming, P. (2021). *Dark academia: How universities die.* Pluto Books.

Gaston, N. (2023, June 20). Starved of funds and vision, struggling universities put NZ's entire research strategy at risk. *The Conversation*. https://www.rnz.co.nz/news/national/492345/starved-of-funds-and-vision-struggling-universities-put-nz-s-entire-research-strategy-at-risk?fbclid=IwAR2T0I8DnIEfIH9a-8OZFxahmG-Yvt8VCHwqfDE22kiq_6p9H5z03ot30HY

Giroux, H. (2007). *The university in chains: Confronting the military-industrial-academic complex*. Paradigm Books.

Giroux, H. A. (2008). *Against the terror of neoliberalism: Politics beyond the age of greed*. Paradigm.

Grabar, H. (2018, May 11). Why universities became big-time real estate developers. *Slate*. https://slate.com/business/2018/05/universities-like-harvard-and-yale-are-real-estate-titans-too.html

Haar, S. (2011). *The city as campus*. University of Minnesota Press.

Hofmeyr, A. B. (2021). Foucault's analyses of neoliberal governmentality. Past investigations and present applications. *Etica & Politica, 23*(1), 589-616.

Kelly, A. (2020). Rethinking the neoliberal university and its impact on students, *Journal of academic language and learning, 14*(2), C1-C6.

LaPointe, L., & Reisetter, M. (2008). Belonging online: Students' perceptions of the value and efficacy of an online learning community. *International Journal on E-Learning, 7*(4), 641- 55.

Littlewood, M. (2023a, July 26). Varsity restructuring may add to savings. *Otago Daily Times*. https://www.odt.co.nz/news/dunedin/campus/varsity-restructuring-may-add-savings

Littlewood, M. (2023b, July 31). Uni's top ranking "heartening to see". *Otago Daily Times*. https://www.odt.co.nz/news/dunedin/campus/uni%E2%80%99s-top-ranking-%E2%80%98hearteningsee%E2%80%99?fbclid=IwAR3M8Jq5oPkmsX1ATFTuFKkSN1QB_x1jZKFAtWl_ZN6GTPVaN6EYXvT5woQ

McMaster, C. (2013). Working the 'shady spaces': Resisting neoliberal hegemony in New Zealand education. *Policy Futures in Education, 11*(5), 523-531. http://dx.doi.org/10.2304/pfie.2013.11.5.523

McNeilly, H. (2023a, April 21). 'The hard work starts now': Otago University yet to confirm which jobs on chopping block. *Stuff*. https://www.stuff.co.nz/national/education/131833095/the-hard-work-starts-now-otago-university-yet-to-confirm-which-jobs-on-chopping-block

McNeilly, H. (2023b, April 29). Bananas and 'woke nonsense' - ex-students show their concern about University of Otago rebrand. *Stuff*. https://www.stuff.co.nz/pou-tiaki/131893183/bananas-and-woke-nonsense--exstudents-show-their-concern-about-university-of-otago-rebrand

McPhail, G. & McNeil, J. (2021). Music education and the neoliberal turn in Aotearoa New Zealand. *Action, Criticism, and Theory for Music Education 20*(3), 44–81. https://doi.org/10.22176/act20.3.44

Minina, E. (2018). Money versus the soul. Neoliberal economics in the education modernisation reform in post-soviet Russia. *Globalisation, Societies and Education 16*(4), S. 435–452. Https://DOI: 10.1080/14767724.2018.1456322.

Ministry of Education. (2023). *The Statement of National Education and Learning Priorities (NELP) and the Tertiary Education Strategy (TES)*. https://www.education.govt.nz/our-work/overall-strategies-and-policies/the-statement-of-national-education-and-learning-priorities-nelp-and-the-tertiary-education-strategy-tes/

O'Carroll, A. D. (2013). Kanohi ki te kanohi – a thing of the past? Examining The notion of "Virtual" ahikā and the implications for Kanohi ki te kanohi. *Pimatisiwin: A Journal of Aboriginal and Indigenous Community Health, 11*(3), 441-445.

One News. (2023, May 28). Election 2023: Labour will make apprenticeship boost scheme permanent. https://www.1news.co.nz/2023/05/28/election-2023-labour-will-make-apprenticeship-boost-scheme-permanent/

Otago Daily Times. (2023). Otago welcomes news of student numbers boost. https://www.odt.co.nz/news/dunedin/campus/otago-welcomes-news-student-numbers-boost?fbclid=IwAR2CQ1duGMWWzlDw18fnP3mPgv0yc5QX2P8I6NbxdYJlSbjNSF8nCZMfpqM

Otago University. (2023a). *Options for future campus development publication*. https://www.otago.ac.nz/administration/campusmasterplan/

Otago University. (2023b). *Vision 2040*. https://www.otago.ac.nz/about/official-documents/#vision2040

Parker, M., Rudolph, J., Shukaitis, S., & Tan, S. (2021). Strategic utopianism and the avoidance of dualisms: An interview with Martin Parker. *Journal of Applied Learning and Teaching, 4*(Sp. Iss. 1), 60-73.

Popenici, S. (2023). *Artificial intelligence and learning futures: Critical narratives of technology and imagination in higher education*. Routledge.

Popenici, S., Rudolph, J., Tan, S., & Tan, S. (2023). A critical perspective on generative AI and learning futures. An interview with Stefan Popenici. *Journal of Applied Learning and Teaching, 6*(2). Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.5

PBRF Review Panel. (2020). *Toward the tertiary research excellence evaluation (TREE)*. https://www.education.govt.nz/further-education/policies-and-strategies/review-of-the-performance-based-research-fund-3/

Radio New Zealand. (2023a, June 27). Government announces extra $128m for cash-strapped universities, tertiary institutions. https://www.rnz.co.nz/news/political/492705/government-announces-extra-128m-for-cash-strapped-universities-tertiary-institutions

Radio New Zealand. (2023b, April 20). University of Otago proposes cutting 'several hundred' jobs. https://www.rnz.co.nz/news/national/488349/university-of-otago-proposes-cutting-several-hundred-jobs

Redden, G. (2017). John Howard's investor state: Neoliberalism and the rise of inequality in Australia. *Critical Sociology, 45*(4-5), https://doi.org/10.1177/0896920517745117

Saad-Filho, A. (2021). Neoliberalism and the pandemic. *Notebooks: The Journal for Studies on Power 1,* 179–186. https://brill.com/view/journals/powr/1/1/article-p179_179.xml?language=en

Schwanecke, G. (2023, June 15). More than 400 roles to go under Te Pūkenga's proposed new 'unified structure'. *Stuff*. https://www.stuff.co.nz/national/education/132335796/more-than-400-roles-to-go-under-te-pkengas-proposed-new-unified-structure

Shore, C. (2010). Beyond the multiversity: Neoliberalism and the rise of the schizophrenic university. *Social Anthropology, 18*(1), 15-29.

Simmonds, P. (2023, June 11). Polytechnic mega-merger unravelling at pace. *The Southland Times*. https://www.nzherald.co.nz/business/steven-joyce-polytechnic-mega-merger-unravelling-at-pace/YXIV4QIE7NFAHOXIZBHGNV4TYY/

Smyth, J. (2017). *The toxic university: Zombie leadership, academic rock stars and neoliberal ideology.* Palgrave Macmillan.

Solomon, S., & Du Plessis. M. (2023). Experiences of precarious work within higher education institutions: A qualitative evidence synthesis. *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.960649

Spina, N., Smithers, K., Harris, J., & Mewburn, I. (2022). Back to zero? Precarious employment in academia amongst 'older' early career researchers, a life-course approach. *British Journal of Sociology of Education, 43*(4), 534-549. https://doi.org/10.1080/01425692.2022.2057925

Standing, G. (2011). *The precariat: The new dangerous class.* Bloomsbury Academic.

Te Ara. (2010). *The 'no. 8 wire' tradition.* https://teara.govt.nz/en/inventions-patents-and-trademarks/page-1

Te, M. (2020, February 10). Victoria University of Wellington spends nearly half a million dollars on brand 'refresh'. *Stuff*. https://www.stuff.co.nz/national/education/119314897/victoria-university-of-wellington-spends-nearly-half-a-million-dollars-on-brand-refresh

Te pōkai tara/Universities New Zealand. (2023). *How the sector works.* https://www.universitiesnz.ac.nz/about-university-sector/how-sector-works

Tertiary Education Union. (2023a, July 6). *Massey policy changes would diminish the university.* https://teu.ac.nz/news/massey-policy-changes-would-diminish-the-university/

Tertiary Education Union. (2023b, June 27). *Permanent solutions needed after welcome funding correction.* https://teu.ac.nz/news/permanent-solutions-needed-after-welcome-funding-correction/

Tertiary Education Union. (2023c, June 21). *Staff and students rally as Victoria announces 230 job losses.* https://teu.ac.nz/news/staff-and-students-rally-as-victoria-announces-230-job-losses/

Tregear, P., Guthrie, J., Lake, S., Lucas, A., O'Connor, J., Pellzzon, A., & Vodeb, B. (2022). 'Enough to make you sick!' Pathological characteristics of the Australian academic workplace. *Social Alternatives, 41*(1), 44-51.

Treasury. (2022). *Vote tertiary education.* https://www.treasury.govt.nz/publications/estimates/vote-tertiary-education-education-and-workforce-sector-estimates-appropriations-2022-23

Turner, P. V. (1987). *Campus: An American planning tradition.* MIT University Press.

Whitty, G., Power, S., & Halpin, D. (1998). *Devolution and choice in education: The school, the state and the market.* Open University Press.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

The (academic) road less travelled: From dropout to recovery

Harvey Gilmore[A]

[A] Assistant Professor of Taxation, University of Hartford, Connecticut, USA

Keywords

Confidence;
failure;
hope;
hopelessness;
recovery;
success;
trauma.

Correspondence

HGILMORE@hartford.edu [A]

Article Info

Received 26 May 2023
Received in revised form 23 July 2023
Accepted 23 July 2023
Available online 27 July 2023

DOI: https://doi.org/10.37074/jalt.2023.6.2.14

Abstract

In our educational pursuits, many of us have fond memories of teachers and professors who supported us, believed in us, and were very instrumental in our ultimate success. Some of us also have memories of certain *other* educators who were, to be charitable, not very kind at all. I have the indelible recollection of an entire high school that ran its day-to-day affairs like a prison, and the educators there (if you can call them "educators") were much more prison guards than actual teachers. I eventually dropped out of high school. I was burned out and demoralized in irrelevant classes to the point where I just didn't care anymore. This opinion piece is not only about my own academic redemption, but it also gives hope to other people who might have had negative educational experiences. To them, I say, you are not alone, you can recover, and your own success will be a perpetual stab in the eye (figuratively) to those who wanted to see you fail.

When I think back on all the crap I learned in high school, it's a wonder I can think at all (Simon, 1973).

Introduction

I am a college graduate and a law school graduate. I was a practicing accountant for close to ten years. Currently, I have been a college professor for twenty-five years, teaching taxation, business law, and accounting. However, prior to my being blessed to have academic and professional success and the resulting accolades therefrom, I was also a completely demoralized, hopeless high school dropout. Thus, I have to slightly disagree with the above quote as I learned absolutely nothing in high school – *nothing beneficial*, that is. Instead, high school merely shoved a bunch of irrelevant, worthless classes down my throat with a culture of *learn this… or else*!

So, how did I go from the depths of high school failure to the summit of professional success? The story of my journey is in the coming pages. Was it an easy journey? Absolutely not, but it wasn't impossible, either. While this article is about my academic redemption, more importantly, it is also a story that will provide some hope for students starting their own journey through higher education after bad experiences in high school and perhaps battling their own self-doubt.

Major reasons for dropping out of high school

Needless to say, there are many reasons why people drop out of high school. The reasons can be academic, job-related, or personal. The academic reasons can be missing too many classes or getting too many bad grades to make continued attendance a viable option (Bouchrika, 2023). Sometimes students have to choose between holding down a job or going to class (Bouchrika, 2023). Personal reasons can include "becoming pregnant, getting married, and having to take care of or support a family member. Students whose families struggle financially are more likely to drop out of school" (Bouchrika, 2023, p. 5).

In my case, admittedly, I dropped out of high school precisely because I skipped too many classes and failed too many classes. The daily routine in my doomed senior year was simply this:

Show up to class (occasionally).

Not care.

Fail one exam.

Fail another exam, and still another exam.

Fail final exam.

Fail course.

Rinse and repeat.

I like to joke that I was a straight-F student in engineering school. The problem was it wasn't exactly a joke. As I will discuss below, I was so demoralized and burned out that I blew off every class in my senior year of high school and consequently failed everything under the sun.

My experience with high school failure

The kind of school I went to

Believe it or not, before the train wreck that was my time in high school, I was actually a pretty good student in elementary and middle school. I had been on the honor roll a few times and also won a few academic awards. With that kind of track record, I figured that I should be able to at least hold my own once I got to high school.

The high school that I went to was very competitive, and one needed to pass a specialized exam to get in. It was a specialized high school and, reputedly, one of the best in New York. What I didn't realize was this school had a well-known national reputation for its STEM programs (Science, Technology, Engineering, and Math). At that time, this school had majors in civil engineering, architecture, industrial design, chemical engineering, electrical engineering, and upper-level math, and was just starting a major in computer science (that's what I remember). Well, I quickly discovered that I had very little aptitude for, and even less interest in, just about all of the classes in the major I was eventually saddled with. And this was just the beginning of the horrors to come. Suffice it to say, this school's STEM programs inspired me so much that I went on to college to become an accountant.

Angry, overbearing, bullying teachers

In the movie "The Breakfast Club" (Hughes, 1985), Principal Richard Vernon was, for me, far and away the villain of the piece. He was evil, heartless, hostile, impatient, and overbearing – everything a good teacher should never be. I reference the movie's Principal Vernon here because the majority of the teachers I had in those days had those same qualities. Most of them were unapproachable and stern, and they never came off as empathetic people you could go to for any help or guidance. The pedagogical culture was more like a prison than an actual school. For me, the (alleged) prestige of going to a supposedly top-tier high school quickly wore off.

I wonder now if Principal Vernon might have been based on at least one, if not more, of my high school teachers. The late Paul Gleason, who played Principal Vernon, perfectly captured the anger, bitterness, and vitriol of so many of my high school teachers. If he were still with us, I would personally congratulate him on his tour-de-force portrayal. The culture at my high school was one of lingering anger from the top down. I had teachers there who called me a loser, said my work product was 'crap' (Gilmore & Smith, 2014) and were firm in their conviction that I would never amount to anything. For me, the typical school day was equal parts anger, verbal abuse, and mental cruelty, along with a seemingly 1950s mentality of *learn this or else*! It was as if those teachers enjoyed themselves barking orders at students who they knew couldn't defend themselves. Adding more insult to injury, the teachers there would yell at me if a gave a wrong answer, if I said I didn't know the answer (Worrell et al., 2023) or if I didn't do a demonstration correctly. I was nothing but a verbal pin cushion for these people, who really weren't teachers, as they were really prison guard-type enforcers of the school's petty yet draconian rules.

To this day, I just can't understand how academicians expect students to succeed under conditions like that – or would want to. I am also convinced that my old school has the (academic) blood of a lot of disillusioned, demoralized students whose desire for learning was browbeaten out of them on its hands. I can tell anyone that it is easy for a student to become demoralized when all teachers seemingly do is focus on what students can't do instead of focusing on a student's potential for future success. That's what learning is supposed to be about; isn't it nurturing a student's potential for success? Nothing can ever justify that kind of negative reinforcement. My old high school was truly a place where hope went to die, and to quote the classic Steely Dan song, "and I'm never going back to my old school" (Steely Dan, 1973).

The day-to-day isolation and hopelessness

For me, there is probably no worse feeling a student can have than sitting in classes every single day, knowing that nothing good will ever come of it. That happened to me. I would get up every day at 6 am and commute ninety minutes to sit in class after class to learn absolutely nothing that would help me in real life. As I recall, classes started at about 8:30, and by 9:46, I had already lost the will to live; that's how stultifying, boring and worthless those classes were. And the school insisted that I do two hours of homework every night for these useless classes (which I never did). What for? Again, these classes were nothing but a daily exercise in hopelessness. The single biggest problem I always had was that no one would ever tell me how and why these classes would be useful in real life. And my reward for enduring the day's uselessness was I get to do it all over again tomorrow. The problem was, with this kind of pedagogy, I just couldn't enjoy learning for its own sake. I always had the lingering, inescapable dread that I always had to fight my way through one class, only to have to defend myself in the next one (Geometry 1 and 2, for example). Students just cannot receive the full benefit of a class when they just cannot see the point of it and must always have their mental guard up with almost no relief.

Being accused of cheating on a test

Ordinarily, in the big picture, failing a test might not be the worst thing to ever happen as long as a student takes the necessary corrective measures to do better the next time around. Admittedly, I failed more than a few exams in high school. Sometimes, I was sloppy and not as prepared as I should have been, and I deserved what I got. Other times, I just didn't have a clue, and I knew it. And later, as I was rocketing towards flunking out, I failed some other exams precisely because I just didn't care. For what it's worth, I was never reduced to cheating on any test; at least my F's were honest.

So, imagine my surprise when I received a letter from one of my teachers making a horrible allegation that I cheated on a test. This teacher was one of the many overbearing, egotistical, blowhard types I had the misfortune of taking more than one class with. With his letter, he included a test that I had taken earlier that week, which I failed (what a shock). His letter claimed that another student (I'll name him "Slater Hepplewhite" to protect the guilty) had test answers wholly identical to mine and that I must have copied his answers.

I was outraged to no end, and I could not wait to meet him in person to rebuke his charges. When I met him in his office, I told him point blank that he was so convinced of my guilt as judge, jury, and executioner that he missed two very important details: first, he was so firmly convinced that I cheated that he obviously never considered the reverse possibility that the other person copied my answers, and second, if I was going to copy from another student, I most assuredly would have copied from a much better student than Slater Hepplewhite. My outrage, logic, and self-defense must have convinced him that I was telling the truth because he backed off after I fought back. He did have the (rare) common decency to tell me that a student that reacted as I did most likely did not cheat. I even passed his class later that semester. Well, that was one of the few victories I had in high school, and I even had lawyering skills way back then and didn't know it.

Failure and more failure

I think it is obvious that students do well in classes when *they want to be in class* (Wood, 2023) This is especially true when students can see the relevance of certain classes to real life, as well as their future career pursuits. "Students only want to take courses that would help them later on in life" (Knight, 2016, p. 2).

> Most students only want to take courses that will prepare them for the real world. When high schools offer courses like personal finance, public speaking or independent living, students are more willing to take these classes because they know what they take away from these classes will actually help them later on in life. When the students can apply their knowledge learned in high school classes to their real life, that is a successful outcome (Knight, 2016, p. 3).

I couldn't agree more. Unfortunately for me, I didn't have that freedom of choice in high school. During my sophomore year, I had to pick a major, and I chose graphic communications, which at that time was the forerunner to what we know today as computer science. I was looking forward to taking classes in computer graphics, programming languages, and maybe even spreadsheets. Well, I never got that opportunity.

Why? The unthinkable happened: my choice of major was overruled and rejected. Instead, I was "awarded" the major of civil engineering. There was no conversation nor any right of appeal. *This* was my major, come hell or high water. (Can one imagine college being this inflexible?) I will also add here that because of my school's hostile delivery of instruction, there was nothing civil about the civil engineering major. What I will never understand is what a high school gains by forcing a student into a major that is so far removed from the student's preferences, interests, and talents. Common sense would suggest that a student has a much better chance of success in classes the student actually wants to take. One would think that a high school would have a vested interest in student success and do everything within reason to help the student be successful. In my case, as things ultimately played out, Gilmore, the involuntary civil engineering major, just didn't have a prayer. For me, that kind of academic despotism was both unconscionable and unforgivable.

Going back to my earlier point of my having little to no aptitude in STEM classes, I got saddled with classes in the strength of materials, structural design, surveying, structural shop, patternmaking, chemistry, and physics, to name just a few. So, I sat in class after class, day after day, knowing these classes would never do anything for me in day-to-day, non-engineering life and that nothing good would ever come from the experience. These courses were just pointless, irrespective of any career pursuit I might have taken. Even if I never became an accountant and never went to law school, those classes would never have helped me get a job as a house painter, a cab driver, a janitor, a cosmetologist, a grave digger, a bartender, or anything else. As to my earlier facetious comment about losing the will to live by 9:46 am in some nondescript class, well, if a serial killer had barged into the classroom and driven a railroad spike through my skull, that would have been an act of mercy for me. That's how disconnected I was. Eventually, I started skipping classes more and more, and truth be told, I had mentally checked out long before I finally stopped showing up for classes.

Again, I ask: *where is the relevance?*

> Many students feel that courses they are taking in high school are not relevant to the field of study they would eventually major in at a higher level of education. Should a student who hopes to major in journalism be forced to take classes such as chemistry, physics, and calculus in order to graduate? Should a student who hopes to major in biochemistry be told he must take three years of Spanish and four years of English in order to get his diploma? (Knight, 2016, p. 2).

And because I was forcibly pigeonholed into "my major" and taking the resulting useless classes, ennui and failure were certainly inevitable. So, here's my litany of "non-success" (yes, failures) as a civil engineering major:

> Trigonometry (twice)
>
> Freehand Drawing
>
> Chemistry (twice)
>
> Physics
>
> Structural Design (twice)
>
> Patternmaking (Wood Shop)
>
> Geometry

> Students have different minds with different interests, and it's unfair to 'universalize' the courses that every student at a school has to take. If a student does well in and has a passion for English courses but struggles in math and science courses, that student should have the freedom to take more literature so he or she can thrive and learn about a subject that he or she loves. Too many students are failing classes that they should never had to take in the first place (Knight, 2016, p. 3)

To that, I say bravo and amen! I will also add this: as a tax professor, those civil engineering classes have never helped me teach my own students how to claim a charitable contribution deduction (IRC § 170) on their income tax returns or a marital deduction (IRC § 2513) on their gift tax returns. The defense rests.

The point where my failure was complete

Sometimes, when people go through a prolonged bad stretch, they tend to feel that they have gone as low as they could possibly go. And it is at that point that the elevator of fate reveals an even lower floor. That's what happened to me. After failing exams, failing classes, and repeating some classes, I finally reached rock bottom and the point of no return.

This happened in my senior year, November 1981, to be exact. I had taken an examination in a structural design class, just one of the many classes that I had no hope of passing in that doomed senior year. When I received the graded exam, my score was a three (yes, as in one, two, three!). When I saw that score, I didn't feel anything. I wasn't mad, I wasn't embarrassed, and I wasn't morose. I just felt nothingness. I stuck the exam in my pocket, and that was the exact moment when I no longer had any hope of graduating in June. And I also knew if I was going to graduate, I would have to come back and repeat the entire senior year. Well, that was not going to happen; I had taken every indignity that I was going to take from that place. Obviously, I had no idea of what I would do next, but I was not about to extend my high school horror into a fifth year. High school had seen the last of me. As Mr. Frank Sinatra so sagaciously put it, "'scuse me while I disappear" (Sinatra, 1966).

The road to redemption

Getting a G.E.D. (General Equivalency Diploma)

My parents were honor students all throughout school, and my mom was the salutatorian of her graduating class. They both enjoyed school and for a child of theirs not to like school was the ultimate insult to them. So, how does a failed high school dropout like me tell his honor student parents that there is no way on God's green earth that he is going back to school? With great difficulty, I readily admit. But I had to honestly tell them that I wasn't going back. Even if I found another school willing to let me transfer there with my sorry record, I knew I was going to bail out at the first sign of trouble. So, transferring was not an option, either. Like I said, I was done with high school.

So, what happened next? That summer, my mom saw an advertisement for a G.E.D. preparation course at a local community college not far from our house. We all liked the idea, and I went and registered. I even remember the registration fee was $35. The class met every Monday night from 6 pm to 9 pm from September through November. The room was filled to capacity every night, and everyone in the class, of all ages, was so supportive of each other. Every week, we would go over a specific subject area that would be tested on the New York State High School Equivalency Exam at that time (writing, social studies, science, math, and reading skills).

This program even gave us two simulated exams similar to the real thing to assess our readiness for the exam. The minimum passing grade was 225, and I scored in the high 270s both times. I knew that I was locked, loaded, and ready to crush this test. When I took the exam that December, I walked out of the room knowing I knocked it out of the park. It was such a sweet feeling having regained that level of academic confidence after having so completely lost it just recently. On the actual exam, I did even better than I did on the practice exams: I was blessed to score 291 (I have that scorecard to this day). That officially drove the stake through the heart of the high school vampire (Moxey, 1972) and freed me up for my next opportunity… whatever that was going to be.

Overcoming my trauma and doubt to start college

Now that I was free of the high school albatross, I really didn't know what I wanted to do next. Frankly, I was so traumatized by high school that I didn't initially see college as a viable option. I thought that college was going to be High School Hell, Volume 2. What ultimately changed my mind was that I had several friends in college who told me it was nothing like high school and that the lifestyle was far less regimented and pressurized than high school. I sent away for some college catalogues from various schools, and I liked the major offerings and course descriptions. I even browsed in a couple of college bookstores, and I enjoyed looking through some of the textbooks, even though I had no clue what was in them. I slowly began to realize that maybe this college thing might not be so bad after all.

My parents even told me that they would pay my tuition, and if I didn't like it after the first semester, I could walk away and get a job with no questions asked. So, with a little remaining trepidation, I enrolled at Hunter College and eventually majored in accounting. I must say here that I appreciate my mom and dad not giving up on me during this time. Even though we did not always agree about my academic prowess at the time, they were never going to disown me, no matter how disappointed they were with my grades at the time. I am sure that there are stories out there of parents and children who have strained or perhaps irreconcilably fractured relationships because of academics. And this disconnect can be compounded by the student's thinking (right or wrong) that the parents only see the student as just a report card instead of a person trying to find their own way (Gale, 2022; Li, 2022; Nolan, 2020; Chng, 2023). Thankfully, many decades later, my mom (still with us as of this writing) and my dad saw me graduate from both college and law school as well as bury the high school demon.

Navigating college on my own terms

Once I made the decision to go to college, I quickly realized that if this was going to work, I was going to do this *on my own terms*. I knew instinctively that if college was going to be high school all over again, where I had no say over my own existence, I would scrap the whole thing and get myself a job pumping gas and go from there. A large part of my trouble in high school was that I somehow had to be all things to all people and live my life on everyone else's terms instead of my own. Once I started college, I saw right away that I could pick my own major, pick the times I wanted to go to class, choose my own career, and run my own race. It was a completely different world for me, and I loved it.

Because of this, I regained the confidence that had been ripped away from me in high school. There's something to be said about being able to make *your* own academic decisions without all of civilization second-guessing your every move. Therefore, I would tell any student in college (and beyond) to run their own race at their own pace. For example, if someone wanted to be a music major and both of that person's parents are already attorneys, that's fine. Remember, mom and dad have already run their academic race. So run your race, and don't let them (or anyone else) run it for you! This is *your* turn, now. Most importantly, I knew that I was going to do this for me and only me. Not for my parents, siblings, or friends. And sure as hell not for my old high school.

Sweet freedom

The great thing about college is the flexibility in scheduling. Once I had the freedom to pick the major I wanted without being overruled, I could even pick the times I wanted to go to class. Most classes have multiple sections, so depending on one's preferences or life situation, one can take classes in the morning, afternoon, evening, weekends, or online. In four and one-half years of college, I had only one 8 am class. Otherwise, my school days mostly started at 10 or 11 am, with a couple of required classes being offered at night. I had no problem with that, and it was far better than vegetating in some irrelevant, worthless class at 8:30 am, as I suffered so incessantly in high school. Anytime I received an assignment in college (and beyond), my only obligations were to do them right and submit them on time. The rest of the time, my life was my own.

So, how did I pick my accounting major? Almost by accident, actually. I originally intended to major in economics, which also required courses in calculus and statistics. Remembering how much I detested geometry and trigonometry in high school, I also understood that calculus and statistics were part of the requirements for the major (and the accounting major as well). Happily, I had done well enough in my college proficiency exams that I didn't need to take any remedial courses (Worrell et al., 2023), making me immediately eligible to take calculus and statistics.

In my third semester, the section of calculus I wanted to take was already closed, and I instead took accounting principles 1 in that same time slot. Turns out, I loved the class, and I officially declared the accounting major the following semester. By the way, unlike high school math, I also enjoyed calculus and statistics in college. I scored a B in calculus and a C in statistics (oh, well).

I can certainly understand the outrage if any college deliberately blocked a student from a preferred accounting major (for example) and then somehow compelled that student to major in something like underwater basket weaving or something just as useless to the student's individual taste. Or worse, imagine a student who is great at math and science and wants to major in engineering but is instead forced into a business administration major. Thus, instead of taking desired courses like aerodynamics, fluid mechanics, geothermal engineering, vector analysis, plane geometry, and the like, the would-be engineering student gets saddled with courses in accounting, finance, economics, management, taxation, etc. I can understand the agony of an engineering student who wants to learn about mechanical stress, torsion, probability theory, and shearing stress but is instead forced to learn about financial statements, financial ratio analysis, or preparing a Schedule M-1 reconciliation on a corporate income tax return. It wouldn't take long for that student to either go scream bloody murder to the administration or, worse, transfer elsewhere and take his tuition money with him.

Understandably, it's one thing to be removed from a college major if a student does not meet its academic requirements, but that's wholly different from being arbitrarily kept away from the major of one's choice. Otherwise, there's no way a *tuition-paying* college student would ever sit still for something so egregious as being blocked from choosing the major that student wants.

The realization that I belonged in college

After my experience in high school, one might think that I might have struggled with the impostor syndrome once I started college. In a nutshell, imposter syndrome refers to a person's internal belief that the person is not worthy of the attained position or status and that the person will eventually be exposed to be the fraud that person really is (Imposter syndrome, n.d.).

Luckily, I never had that problem. My 291 on my high school equivalency exam was good enough to get me into most colleges in New York. Hunter College accepted my application for admission on that basis. Thus, my getting accepted was not a fluke, and my getting accepted was my opportunity for a fresh start after the high school nightmare and away I went. I also began to understand how certain required classes outside my major were relevant just the same. For example, in economics, one has a better understanding of supply curves, demand curves, and marginal revenue curves with a basic understanding of calculus. While I'm no mathematician, I could at least see the connection between calculus and economics, and this understanding only added to my academic confidence. I knew beyond any doubt that college was for me. In fact, one month into my first semester, I was having lunch in the cafeteria one day, and I suddenly realized: *so, this is college, and I belong here!*

My finest hour in college

As an accounting major, one of the required courses I had to take was the introductory course in federal income tax. As a full-time student, I had never worked anyplace before, "so I was intrigued about what I would learn in an introductory tax class. Turns out I enjoyed the class, and the attorney who taught the class was phenomenal and just plain brilliant. He broke the concepts down so they made perfect sense to a novice like me" (Gilmore, 2022, p. 49).

Schedule-wise, I had taken all the other required accounting classes, and I was taking tax in my final semester. So, there was a little bit of pressure as this was the class that stood between me and graduation (I wasn't at all worried about the other fluff classes I took that semester; *tax was the big one*). So, for obvious reasons, I could not fail and lose out on graduating from college that semester. It would have been like going all the way to the Super Bowl just to lose (just ask the Cincinnati Bengals and Philadelphia Eagles the last two seasons).

Thus, I studied my tail off for that class like I never had before. I got a B+ on the midterm, so that took a little pressure off. The final exam was on a Tuesday night, so I'd have to sweat out the entire week, plus the weekend before the grades were posted. Was that wait ever excruciating! Finally, on the following Monday, I rushed to campus at about one in the afternoon, and the tax grades were posted. Come hell or high water, I had to see how I did; graduation was riding on this final throw of the dice. In those days, the department offices used to post computer printouts of the final grades, and you looked up your Social Security number to see your grade. And there it was: *God blessed me with an A!* My graduation was secured! By the Grace of God, I was actually a college graduate! The comeback was complete! As I had a great time in college overall, that moment was my finest hour as an undergraduate (Gilmore, 2022). And that started me on the road to the teaching career I enjoy today.

What helps retention in higher education?

As I mentioned earlier, students do well when they *want* to be in school. Thus, institutions of higher learning understand that and therefore have various initiatives in place to get students – and keep them. These initiatives can help students negotiate the journey from academic failure to academic success.

Showing college students that they belong in college

College gave me a fresh start after high school. When I attended my first orientation, one of my advisors told me in no uncertain terms to check any high school baggage at the door. The advisor also told me that my score on the high school equivalency exam, along with my college proficiency exam scores, left no doubt in anyone's mind that I was ready for college and up to the challenge. That meant a lot to me, and once I started, I never looked back; Hunter College (my first alma mater) believed in me, and I believed in myself.

Your admission is not an accident

Sometimes, it might take a student a while to get used to the idea that he or she is really in college, and the student might inadvertently believe that the admissions committee somehow made a mistake in letting him or her in. That is not the case at all. An admissions committee is a group of people who meticulously read through a person's college application at least twice before making the ultimate decision to accept the student (Milliman, 2022). The admissions committee is thus charged with the responsibility of bringing in the best possible freshman class for the college/university. Ordinarily, two different groups will look at your application and will decide if you happen to be a good fit for the university and, therefore, membership in that year's incoming class (Milliman, 2022).

In my case, the admissions committee at Hunter College saw my high score on my high school equivalency exam. The committee saw that I was someone who had recovered from the worst academic failure, which was completely failing out of school. And based on that recovery, the committee reasonably believed that I could successfully complete college and graduate, which I did. My high G.E.D. score was really the only thing I had going for me when I applied for college. But sometimes, it only takes one spark to start the chain reaction from college acceptance to admission to graduation. Once I got in, I never thought it was an accident, and I never pulled off any fraud to get in. Actually, because there were so many people looking at my application, I couldn't scam that many people simultaneously into thinking that I was better than I actually was. I'm neither that smart nor that devious.

Again, my admissions committee thought enough of me to accept my application, and I happily vindicated their confidence by graduating. My getting admitted to college was not a fluke, and your getting admitted was no fluke either. Accept the committee's gift of their confidence in you and go validate that confidence. If I can do it, with my track record, then you sure can! Go get 'em!

Asking what students expect to get out of the class

When students register for a class, they usually expect to get something out of the class. That expectation could be general knowledge, how the class is connected to their major, how the class has a direct impact on their lives, or anything else. To a certain extent, a student wants to know the real-life relevance of a given class. Thus it is certainly reasonable for an instructor to ask what a student wants to get out of the class, and the resulting conversation can help a student become more engaged in the class. In other words, the student will want to go to class and feel less compulsion to do so.

Showing real-world examples connected to the subject and how they relate to the student

As I mentioned earlier, practically all my high school classes had nothing to do with my life outside of school, and nobody cared enough to try to tell me why they mattered. I submit, students have a legitimate educational right to know how certain classes will impact their lives. So, why not show them? When I teach contract law, for example, I break it down so that my students can see that a contract is not always a multi-billion-dollar corporate transaction that they can't relate to. I break the concept down to something students *can* relate to. A contract is, by definition, a legally enforceable agreement (Black's Law Dictionary, 2016) for which the law gives a remedy.

I tell them, for example, when they go into the local Dunkin' Donuts and purchase a large French Vanilla coffee and pay $3, they have completed a contract. Just like that! They pay the $3 and walk out with the item just purchased. And to further show that a contract is not at all removed from everyday life, I also tell the class that when they buy coffee, they do not swear on a Bible and take an oath that they are creating a contract for the sale of goods pursuant to Article Two of the Uniform Commercial Code. They just did it. I also tell them that they create legal contracts all the time: registering for classes, paying tuition, renting an apartment, buying a car, painting a fence, buying groceries, babysitting a friend's cat, and so on.

Similarly, in financial accounting, I show my students, irrespective of major, how to keep track of their money. They can keep track of where the money comes from, where it goes, and how to record every transaction. This is certainly relevant in showing students how to balance their checkbook or how to stay on top of their finances if they own their own businesses or eventually become accountants.

Finally, in income tax, I show my students the basics of income tax preparation and the legal rules that apply to them. They truly want to know how to keep as much of their money away from the tax authorities as legally possible. I show them what deductions they can properly take to reduce their taxable income and eventual tax liability (prescription medications, co-payments, braces, psychiatric visits, charitable contributions, and state/local taxes, among others). I also show my students whom they can claim as their dependents (parents, children, siblings, etc.), which results in tax credits that also reduce their tax liability. Last but not least, I tell my students that they can at least do their own taxes and they won't have to pay H & R Block anymore, and that is another way for them to save money. Thus, my students can definitely appreciate how the subject matter applies particularly to them.

Discovering classroom humor

How having fun can lead to academic success

The best times that I had in school were in college, graduate school, and law school. Why? Because the environment was conducive to learning and academic success. Thanks to the many great professors I came to know – and became friends with, the classroom aura was almost always easygoing, light-hearted, and low-stress. Yes, I understood the obligation to do the work on a timely basis and give my best efforts, but the experience was so much fun that I actually thrived. This was light years removed from the authoritarian and dystopian culture of my old high school. It's no wonder that after my high school disaster, I found success in higher education. It was so liberating to learn at my own pace and have fun in the process. "Looking back at my time in college, graduate school, and especially law school, I know that the classes that were the most fun, and where I had the most laughs, were many of the same classes where I also did the best" (Gilmore & Smith, 2014, p. 294).

This is proof positive that learning is not a chore, and a student does not always have to choose between learning and fun. I would not have been successful in college and law school otherwise. By the way, I am not the only one who sees a direct connection between having fun and academic success:

> In response to the open-ended questions at the end of the survey, students identified 'humor' and making the material 'fun' as specific characteristics they sought in an 'entertaining' teacher. For instance, a Colorado student said: 'Teachers that are funny, friendly, make sarcastic comments, etc... are more real and it makes class more interesting which facilitates learning.' Several others said things like: 'the best way to keep students' attention is to try to make class fun or entertaining.' 'Be a dynamic lecturer' was another comment echoed by several students. 'If a teacher is entertaining, knows the material and enjoys teaching, then learning is so much easier' nicely summarizes what many students value in a good teacher (Levy, 2006, p. 82).

And at the other end of the spectrum, students can also perceive how a teacher's complete lack of humanity and affability can make learning a living hell:

> Although the importance of teacher friendliness may be obvious, many students made it clear how detrimental to learning it can be when their teachers are not friendly. Several students warned that teachers should not be 'intimidating,' 'hostile,' or 'unfriendly and aloof. Another student said: 'A stiff, cold, unapproachable personality makes someone an ineffective teacher.' Yet another said: 'Poor social skills inhibit learning in an interactive classroom.' Finally, a part-time UNLV student said that when the [professor] is unapproachable and barely human, the class is truly brutal (Levy, 2006, p. 86).

This is proof positive that learning and fun are not mutually exclusive or necessarily an either-or proposition. Having fun and not being isolated by the learning process can make the single biggest difference between passing and failing:

> It makes sense to me why I didn't feel comfortable saying my piece in the classroom for so many years. I was afraid to say the wrong thing and feel stupid. In a domination style learning environment, students' creativity and, ultimately, their potential is limited. In a holistic setting, students teach and learn interactively with the teacher. As they are teaching and learning, equality is created between them (Robinson & Kakela, 2006, p. 206).

Law schools are rooting for student success, too!

Let me start by getting the obvious out of the way: successful retention, graduation, bar passage, and success in professional life for law students are good for law school business. Why? It keeps the business open. Anyone who has ever seen the movie "The paper chase" (Bridges, 1973a) knows Contracts Professor Charles W. Kingsfield as the stereotypical Socratic professor. Kingsfield professes that his only goal is to teach the uninitiated how to think like a lawyer: "You teach yourselves the law, but I train your minds. You come in here with a skull full of mush; you leave thinking like a lawyer" (Bridges, 1973b). Being told that your brains are full of mush is not the most welcoming or confidence-inspiring thing one wants to hear on the first day of law school.

This brings us to the Socratic Method of law school teaching. This is a teaching methodology at the law school level that suggests that a student learns the law much more effectively by answering a series of rapid-fire questions asked by the professor on the spot, as opposed to being given the answer during the professor's lecture (Gilmore, 2013). That said, the biggest criticism about the Socratic method is that it puts students in the professor's crosshairs, where they can be cold-called at any given time, and the student can only answer so many questions before the professor pulls the rug out from under the student and proceeds to disassemble the student's points, one by one (Gilmore, 2013).

That said, I believe the law school teaching model has somewhat moved away from the Kingsfield model of forcing students to answer questions under unceasing pressure. In my own experience, professors have allowed me to volunteer or assign me to lead the class discussion on upcoming cases. Although I was confident in my ability once I was in law school, I do admit this kind of methodology removed the sense of foreboding that my professors were looking to take me apart as soon as I opened my mouth.

Some of you reading this might ask, ‘Well, wait a minute Professor G. If this article is the story of your academic redemption, what does law school have to do with this?’ Well, law school was the next part of my academic journey, in which I was blessed to be successful. By the time I started law school, I already had undergraduate and graduate degrees in my pocket. Therefore, I had already recovered the confidence that was so completely dismantled in high school, had success in higher education, and I was ready for my next adventure.

Going back to my earlier point that law schools want and need successful graduates in order to continue as viable going concerns, I wasn’t just being facetious. One way law schools help students to become successful students and professionals is by showing that they care about their students, not just as students, but as people. “All told, faculty who let their students know that they care about them will, in turn, earn enhanced respect from students” (Jaffe, 2023, p. 13).

In addition to caring about students as people, law schools are increasingly cognizant about students having mental health issues that could adversely affect their academic progress (Jaffe, 2023). Thus, in recent years, law schools have had mechanisms in place to help students with any issues affecting their mental health. In addition to on-campus counsellors, law schools have lawyer assistance programs that help students with different issues, mental health, depression, substance abuse, and the like (Jaffe, 2023). During my time in high school, I am happy to report that despite the daily browbeating, I was blessed not to have suffered from depression or any substance abuse issues. Just getting my self-confidence pulverized every day was bad enough.

How my learning experiences inform my teaching style today

Showing my students that I’m just like them (only slightly older)

My opening day icebreaker every semester is that I tell my students that when I was much younger, much skinnier, and when I had hair, I sat exactly where they sat (not literally, of course), trying to make sense out of the subject matter for the first time. I tell my students that the only difference between me and them is that I’m a forty years older version of them, and believe it or not, I’m just like them. That normally puts them at ease, and they know I’m someone who’s been there and am willing to help them be successful. My students truly appreciate that.

Learning and good storytelling

Part of the fun in learning and retaining concepts is in the professor being a good storyteller while teaching the day’s topic. It’s been my experience on both sides of the lectern that students retain the subject matter longer and do better when the professor can tie a wild, zany story back to the day’s topic.

In my tax classes, I tell a story about finding money in the street and how it triggers a taxable event if the finder then claims it as his own. I then point them to the specific section of the tax law that defines gross income (IRC § 61), and I point them to the actual tax case that proves found money is subject to tax (Cesarini v. United States, 1969).

In my business law classes, I show first that there must be an offer and acceptance working in tandem to create a contract (Miller, 2017). I then show that once an offer is rejected, it can never be resurrected to create a contract. Assume, for example, I ask a young lady out on a date, and she responds that she would rather stick her face in acid before going out on a date with me. That is an unmistakable rejection (Miller, 2017). Three seconds later, she says that she was only kidding and would love to go on a date with me. (Suspend your disbelief about the legality of a date; I’m only showing the necessary contractual components of offer and acceptance.) Since a contract requires both an offer and an acceptance together in tandem, once she rejected my original offer, the offer is automatically terminated and is no longer on the table, and therefore there is nothing for her to accept.

Here's a more concrete example. My friend George offers to sell me his baseball card collection for $750, and I say thanks but no thanks. Later on, I change my mind and tell George that I would love to buy his baseball card collection. Unfortunately, I would be out of luck because the offer is now terminated, and I cannot accept an offer that is no longer there.

Here’s another story I use to keep my students awake and laughing. I’m a senior in high school, and I want to play college football at the University of Miami. All the big-name Division One schools are intensely recruiting me to play for them: Georgia, Alabama, Clemson, Notre Dame, UCLA, Stanford, Michigan, and Ohio State. I reject them all because I’m holding out for the University of Miami. I’m overjoyed when I receive a recruitment offer from the University of Miami, and I sign off on the letter of intent to play for the University of Miami. However, I didn’t read the letter too carefully, and instead of playing for the University of Miami Hurricanes in Florida, I committed to play for the Miami University Red Hawks, located in Oxford, Ohio (yes, there is such a University)! Oh, no! Because this is a unilateral mistake (Miller, 2017) on my part, I am legally obligated to play in Ohio for at least my freshman year before I can try to transfer elsewhere. Because of my haste, I am now stuck with a contract due to my own rushed, negligent reading. As Charlie Brown would say, “RATS!”

I've had the experience in college and law school of taking wonderful professors who had great storytelling skills that helped me make day-to-day practical sense of the subject matter. This certainly helped me succeed and made college and law school so worthwhile for me. And this was light years removed from my high school experience of being force-fed pointless classes.

The college lifestyle is far less regimented than high school.

I am doubtless overstating the obvious when I say that college is a completely different experience than high school. One of the biggest differences is that college is far less regimented and draconian than high school. Of course, in college, one attends classes, submits assignments, takes exams, writes papers, and meets deadlines. However, because college scheduling is much more flexible than high school, a student can easily get the work done on time and still have plenty of time to enjoy other things in life.

> Because in college you are typically only taking between three and five classes at a time, you will likely have much more free time than you did in high school. Most students in high school go to class from 8 am to 3 pm and then have sports practice or other activities after school and on top of that have homework… In college, you typically only have each class two to three times a week. You will have homework, but professors usually do not assign busywork, and instead, it will be larger projects such as essays. Time management is a key part of college and if you manage your time properly you will have plenty of time to spend with your friends and pursue other activities (Korn, 2021, p. 4).

Another key difference that separates the flexibility of college from the regimentation of high school is that a student won't have someone breathing down his neck every second about any given assignment. One of the many things I hated about high school was I always had some teacher nagging me about some assignment, usually twenty minutes after I got it.

With that kind of petty administrative housekeeping, I eventually decided that I'll hand something in just to hand it in. Why should I break my neck to do an assignment that I didn't care about in a worthless class that I wasn't going to pass anyway? That does nothing for positive reinforcement and student motivation.

Whereas in college (and beyond), you, as the student, can make the call on what you are going to do and when you will do it. You also understand that you have to work within certain parameters, and that's okay. Ultimately, that is your decision and responsibility. Welcome to freedom!

> In college, you have the ability to make your own decisions on many things. For example, no one is going to ask you why you are leaving your dorm or apartment to get food in the middle of the night. Or you can head out Saturday morning to explore new places without needing an explanation as to why you aren't working on that big homework project! This freedom of being able to do whatever you want whenever you want truly makes college and the transition into adulthood so much fun (Korn, 2021, p. 2).

Another key difference between high school and higher education is the relationship between the student and educator. During my high school incarceration, in a typical eight-period class day, at least seven of my teachers were martinets and held themselves out as strict, no-nonsense drill instructors. When the learning environment is nothing but militaristic, I believe that does the student much more harm than good and eventually crushes the desire for learning. (Sparks, 2018) My time in high school reminds me of a classic Night Gallery episode called "The Academy," (Serling, 1971), where the learning environment in an unusually strict military school was nothing but discipline and endless drills. While this would be normal in military school, the episode's punch line was that in *this* school, the students are forever matriculated and never graduate. (Skelton & Benson, 1999).

In college (and law school), the majority of professors are genuine people who want their students to succeed. A professor is not a student's presumed enemy. Professors want to help their students do well and don't want to put the screws to anyone. I had professors who were not only great people but enjoyed teaching and were passionate about it. Passion is something one cannot fake. Thanks to them, not only did I become successful, but they enjoyed my success as much as I did.

> Professors truly care about what they are teaching and want to help you. For the most part, college professors are excellent at what they are teaching and have been researching the subject for many years. Professors are experts at their subject and most of them have written textbooks or numerous research papers on the topic. And because they are so passionate about the subject, this usually will show when they teach their lectures, which in most cases are engaging and entertaining to hold your attention (Korn, 2021, p. 5).

A student can be happy to know that a professor has a back story just like theirs

Sometimes when a student is going through a bad academic time, the last person the student might want to speak with is the student's own professor. Why is this? Sometimes the student might have the perception that the professor is someone from another planet, far removed from whatever problems the student might be having. To put it another way, the student believes that the professor can't relate, doesn't care, or even both. I *know* the trepidation because most of my high school teachers were unfriendly and unapproachable. So, I certainly couldn't talk to them because they just couldn't care less and surely did not want to hear anything I had to say.

That said, I must point out that professors have office hours, where students can come in for academic advice, career advice, life advice, or just get a sympathetic ear. When I occasionally mention that I'm a high school dropout, my students are both amazed and stunned. Initially, students see their professors as professional subject matter experts in the classroom. And in my case, when I tell my students that I flunked out of high school but recovered, I think that psychological barrier between me and my students disappears. At that point, they realize that I'm a (slightly) older version of them. And it really helps their classroom experience with me because if they have academic insecurities, they know that I, as their professor, have a backstory similar to theirs. And they know that *I can relate*. And my door is always open.

Having difficult conversations with compassion, not confrontation

Because I've personally experienced academic trauma and failure, I think I have a better understanding of how to converse with students who might be struggling. For example, when I sit down with a student who failed my exam, I do not place any judgment. I show the student where he went wrong and how he can improve the next time around. That is far different from a pedagogical mindset that says, "You blew the exam; you're an idiot; you're never going to get this; you don't belong in college; change your major; get a refund of your tuition," and so on. If a student's confidence has taken a hit because he's struggling with a class or didn't do well on an exam, that student might be alienated from wanting to improve if he senses the professor is apathetic at best or confrontational at worst.

So, in my very small way over the years, I've helped some students overcome a tough start to finish well and succeed in my classes. Because I've seen those students recover and go on to success, there is no doubt in my mind that students do better when they know their professor is in their corner, as my professors were for me. Finally, that success is manifested when they graduate! Over the years, I've seen my former students become Certified Public Accountants, Enrolled Agents who can practice before the Internal Revenue Service, Business Owners, and Professional Chefs, and two of them have become Attorneys. They did the heavy lifting, and I was happy to play a small role in their development. I am certainly proud of them.

Conclusion

If there is any piece of advice I can give to any academic administrator that would be any good, it would be this: one cannot run a school/classroom like a prison, treat students like inmates, shove irrelevant classes down their throats, and then be surprised that students do not want to learn or succeed under such dehumanizing conditions. That *modus operandi* is counterproductive at best or demoralizing to students at worst. It will never do any good for a student's confidence for that student to go to class after class every day expecting nothing but vitriol from the very people entrusted to facilitate their learning, knowledge, and intellectual curiosity.

I was fortunate enough to escape from that spirit-crushing environment and recover to succeed in higher education. But how many similarly situated people are out there who weren't as blessed as I was and perhaps lost an opportunity to succeed in higher education? I submit, any institution that inflicts mental cruelty and tries to call it academic instruction is committing academic malpractice and should be held legally liable for both malpractice and intentional infliction of emotional distress.

Knowing this, institutions of higher learning have resources in place to help students feel welcome, to help students find their way, sometimes to lend a sympathetic ear, and even help with mental health (and other) issues. It also helps to have professors and administrators who are committed to student success. As discussed above, it also helps students to have professors like me who can more intimately understand a student's struggle and resulting insecurity and doubt.

Because I was able to get a clean break from high school and a fresh start thereafter, I never had to speculate on "what might have been" had I been able to pursue the major I really wanted in high school. I have no doubt that the Grace of God definitely returned "beauty for my ashes" in the form of a successful academic and career trajectory (*Holy Bible*, 2015, *Isaiah* 61:3). That said, I must point out that my commentary here is not intended to be a blanket demonization of all high schools. Many people look at their high school experience with fondness. And I do not suggest that the college experience is as universally rewarding for others as it was for me. Each of us is a product of our own experiences.

It has been well over forty years since my escape from high school, and I haven't gone back since. Although I have absolutely no plans to go back to visit, I do hope that, over time, my old school has at least tried to make some lasting changes for the better. My story gives even more meaning to the old advertising slogan, *been there, done that*. Thus, the support systems mentioned above go a long way to helping students succeed far better than the "old school" explanation of "sit down and shut up" (Liberals response to dissent, 2015). In short, intentional classroom vitriol does nothing to help student success and retention. Nurturing and support are what help students succeed. I am living proof.

References

Black's Law Dictionary. (2016). Thomson Reuters.

Bouchrika, I. (2023, May 12.) *High school dropout rate is decreasing, but race, income & disability issues persist.* https://research.com/education/high-school-dropout-rate#:~:text=The%20top%20three%20school%2Drelated,Dropout%20Prevention%20Center%2C%20n.d.).

Bridges, J. (Director). (1973a), *The paper chase.* 20th Century Fox.

Bridges, J. (1973b) *The paper chase – The Socratic* *method* [Video]. YouTube. https://www.youtube.com/watch?v=kp7hBHdcv0s&list=PLLt5qtosTj5OzfTMrWg3-_DVRdUALIYm3&index=3

Cesarini v. United States, 296 F. Supp. 3 (N.D. Ohio 1969). *Justia Legal Resources*. https://law.justia.com/cases/federal/district-courts/FSupp/296/3/1982804/

Chang, C. (2023, January 25). *How to get your parents to calm down when you get a bad grade*. https://www.wikihow.com/Get-Your-Parents-to-Calm-Down-when-You-Get-a-Bad-Grade

Charlie Brown's quotes. (2023). *Famous quotes & sayings,* https://quotestats.com/topic/charlie-browns-quotes/

Gale, A. (2022, September 19). Mom dragged for saying daughter's grades are more important than her pain. *Newsweek*, https://www.newsweek.com/mom-dragged-saying-daughters-grades-are-more-important-her-pain-1747733

Gilmore, H. (2013). Misadventures of a law school misfit. *Duquesne Law Review, 51*(1), 191-216.

Gilmore, H. (2022). Why like taxes? It's not the reason you think. *Rutgers Business Law Review. 18*, 43-62.

Gilmore, H., & Smith, G. (2014). The joy of passion – Finding your calling. *WMU Cooley Journal of Practical & Clinical Law, 16*, 259-313.

Holy Bible, new living translation. (2015). Tyndale House Foundation.

Hughes, J. (Director) (1985). *The breakfast club*. Universal Pictures.

Imposter Syndrome (n.d.). *Psychology Today,* https://www.psychologytoday.com/us/basics/imposter-syndrome

Jaffe, D. (n.d.) *The key to law student well-being? We have to love our law students. The key to law student well-being?* (american.edu)

Kenton, J. (2022, November 23). *When the teacher is the bully.* https://www.greatschools.org/gk/articles/when-the-teacher-is-the-bully/

Knight, S. (2016, January 3). Why students should choose their own classes. *The Sun Chronicle*. https://www.thesunchronicle.com/opinion/columns/guest-column-why-students-should-choose-their-own-classes/article_9b794965-5d99-55c0-9890-fa66d23ce407.html

Korn, M. (2021, March 30). *10 Reasons why college is better than high school*. https://grownandflown.com/college-senior-10-reasons-college-better-than-high-school/

Levy, J. (2006). As a last resort, ask the students: What they say makes someone an effective law teacher. *Maine Law Review. 58*(1), 58-98.

Li, P. (2022, December 31). *Bad grades – How parents can help teenagers achieve*. https://www.parentingforbrain.com/how-to-discipline-a-teenager-for-bad-grades

Milliman, H. (2018, December 15). *Who is the admission committee? What do they look for?* https://blog.prepscholar.com/college-admissions-committee

Miller, R. (2017). *Business law today,* (12th ed.). Cengage.

Moxey, J. (Director). (1972). *The night stalker.* ABC Circle Films.

Nolan C. (2020, October 12). *8 tips for talking about bad grades.* https://health.clevelandclinic.org/8-tips-for-talking-about-bad-grades/

Liberals response to dissent: Shut up. (2015, November 5). *Chicago Sun-Times*, Opinion, https://chicago.suntimes.com/opinion/2015/11/5/18540394/opinion-liberals-response-to-dissent-shut-up

Robinson, C., & Kakela, P. (2006). Creating a space to learn: A classroom of fun, interaction, and trust. *College Teaching, 54*, 202-206.

Serling, R. (Writer), Corey, J. (Director). (1971, October 6). The academy (Season 2, Episode 4 [TV Series Episode]. In *Rod Serling's night gallery*. J. Laird (Executive Producer). NBC.

Simon, P. (1973). *Kodachrome. There goes Rhymin' Simon*. Columbia Records.

Sinatra, F. (1966). *Angel eyes. Sinatra at the sands*. Reprise Records.

Skelton, S., & Benson, J. (1999). *Rod Serling's night gallery: An after-hours tour.* Syracuse University Press.

Sparks, S. (2018, May 11). Students learn less when they sense teacher hostility. *Education Week*. https://www.edweek.org/leadership/students-learn-less-when-they-sense-teacher-hostility/2018/05

Steely Dan. (1973). *My old school. Countdown to ecstasy.* MCA Records.

Wood, S. (2023, January 2023). What students are saying about what motivates them to learn. *New York Times*, https://www.nytimes.com/2023/01/19/learning/what-students-are-saying-about-what-motivates-them-to-learn.html

Worrell, G., Edelman, S., Linge, M., Calder, R., & Genaous, J., (2023, February 25). *Nearly half of NYC DOE grads at CUNY need remedial classes.* N.Y. Post.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Developing and analysing an authentic technical proposal writing assignment through the lens of an authenticity framework: Implications for practice

Muhammad Rahimi[A]

A Assistant Professor of Applied Linguistics, Centre for Communication Skills (CCS), Singapore Institute of Technology (SIT), Singapore

DOI: https://doi.org/10.37074/jalt.2023.6.2.28

Abstract

This paper presents an analysis of an authentic writing assessment to externalise the knowledge, skills, and attitudes necessary for developing such assessments. Specifically, the paper examines a technical proposal writing assignment developed and used as a continual assessment in an undergraduate engineering course by employing an authentic assessment framework as the analytical lens. The findings showed that the framework can serve as a valuable guide in developing authentic writing assessments. Further, it became evident that achieving functional authenticity in all dimensions may not always be viable. Developing fully functionally authentic writing assessments/assignments for beginners is not always advisable, as completing such assignments may lead to cognitive overload. Therefore, it is recommended that both the dimensions of authenticity and students' ability levels be considered when analysing, developing, and using authentic writing assessments and assignments. Indeed, striking the right balance between authenticity and students' cognitive capacity is crucial for creating optimal learning experiences.

Keywords: Artefact analysis; authentic assessments and assignments; authenticity framework; latent and functional authenticity; technical proposal writing; writing assessment.

Introduction

IIt is widely acknowledged that assessment plays a significant role in shaping both learning and teaching processes (Preston et al., 2020). As Gibbs (1992) metaphorically stated, "the tail wags the dog." This raises the question of how to carefully design assessment practices to ensure that they effectively guide students' learning journeys. Authentic assessment (i.e., the assessment that resembles professional practice) could be the solution as it has the potential to motivate students to actively learn the essential knowledge, skills, and attitudes (i.e., competencies) required for their future professional lives (Gulikers et al., 2006). Previous research has made significant contributions to understanding authentic assessment, with Gulikers et al. (2004) presenting a five-dimensional framework. These scholars, along with other researchers (e.g., Bachman, 1990; Bachman & Palmer, 1996; Kohnen, 2013; Wargo, 2020; Weir, 2005), acknowledge that authenticity is a continuum, and the authenticity level of an assessment is determined by how closely each dimension of the assessment resembles real-world practice.

Furthermore, the authenticity of an assessment is also influenced by its implementation (Ghosh et al., 2021). For example, if real-world tasks involve seeking advice and feedback from experts, the assessment should reflect this real-world implementation to approach authenticity. Previous research has identified the dimensions of authenticity and examined the relationship between students' perception of assessment authenticity and their study approaches, development of generic skills, and academic grades (e.g., Gulikers et al., 2004, 2006). This paper aims to advance prior research by utilising the findings of previous studies to develop and analyse an authentic writing assignment. Specifically, in this paper, the author analyses a technical proposal writing assignment that was developed and implemented in a critical thinking and communication skills module for first-year undergraduate engineering students. The aim is to provide insights into developing and implementing authentic writing assignments.

Authentic assessment and construct and consequential validity

The assessment paradigm has shifted from traditional standardised tests, which focused on assessing discrete points of knowledge, to a new approach that emphasises the integration of knowledge, skills, and attitudes in performing real-life tasks in realistic contexts. The functions of assessment in this new paradigm include stimulating learning, promoting competency development, and evaluating students' performance (Gulikers et al., 2006). Therefore, assessment is no longer decontextualised and focused on discrete points; instead, it is performance-based, integrated, and contextualised (Birenbaum, 2003).

Authenticity is crucial for achieving both construct and consequential validity in assessment.

The construct validity of a competency-based assessment relies on its ability to measure the competencies required for real-life task performance in real-life situations. In essence, when the assessment requires the same competencies that target tasks do, and there is a correspondence between the assessment and target task performance situations, it is more likely to effectively assess the intended competencies (Messick, 1994). Additionally, assessment entails unintended and intended consequences (Schuwirth & Van der Vleuten, 2004). By designing and implementing authentic assessments and aligning assessment, instruction, and learning (Biggs, 1996), the consequential validity of the assessment can be enhanced, leading to more desirable outcomes. In other words, designing and implementing authentic assessment and aligning authentic assessment, instruction, and learning are essential to enhance the consequential validity of the assessment, as these practices can stimulate the development of the competencies that students will need to perform real-life tasks in their professional practices.

Previous studies (e.g., Herrington & Herrington, 1998) showed that students prefer assessment tasks that closely resemble real-life tasks and help them acquire skills applicable to their future professions. They value assessments that prepare them for their professional lives beyond the school setting. Recognising the importance of authenticity in enhancing the validity of assessment and its positive impacts on students' learning, Gulikers et al. (2004) have developed the Five-Dimensional Framework (5DF) for authentic assessment. According to Gulikers et al. (2004), authenticity in assessment is a multidimensional construct that exists on a continuum. They propose that the authenticity of an assessment is determined by the extent to which its five dimensions align with those of the target tasks performed in professional settings. These dimensions include the task, the physical context, the social context, the form of the assessment, and the criteria used for evaluation. Together, these dimensions form the framework for authentic assessment. The task dimension relates to the content being assessed, the physical context dimension pertains to the assessment environment, the social context dimension considers the interaction possibilities and constraints during the assessment, the form dimension focuses on the assessment method, and the criteria dimension addresses the valued characteristics of performance. In a subsequent study, Gulikers et al. (2006) found that students' perceptions of authentic assessment were positively and significantly correlated with their study approach, development of general skills, and grades.

Latent and functional authenticity

Scholars have distinguished latent and functional authenticity (Kohnen, 2013; Wargo, 2020), which correspond to interactional authenticity and situational authenticity in language assessment, respectively (Bachman, 1990; Bachman & Palmer, 1996). Latently authentic assessments/assignments require students to utilise metacognitive and cognitive processes similar to those necessary for real-life task completion. On the other hand, functional authenticity necessitates authenticity across all dimensions. In writing assessments/assignments, students may find functionally authentic tasks more engaging compared to latently authentic tasks, as the former involve addressing real-world concerns and influencing real-world audiences. This difference in the depth and breadth of student engagement may result in noticeable differences in the quality of the written genres produced by the students. For instance, a writing assignment completed for a module to obtain grades with the lecturer as the audience may have limited functional/situational authenticity. However, in certain circumstances, due to resource constraints and students' ability levels, a latently authentic assessment is more advisable. This is particularly relevant for beginners, as fully functional authentic assessments can lead to cognitive overload for them (Sweller et al., 1998).

The context

The technical proposal assignment analysed in this paper was designed for mandatory university-wide four-credit critical thinking and communicating (CTC) modules for first-year undergraduate students in Singapore. These modules focus on developing students' critical reflection, critical reading, writing, and presentation competencies. The Paul-Elder framework (Paul & Elder, 2019) was incorporated into the module to teach critical thinking skills explicitly. Specific materials, assignment briefs, and resources were created and curated for the CTC modules. The instructional materials and activities were designed to scaffold students' acquisition of the necessary knowledge, skills, and attitudes (competencies) required to successfully complete the assignments. By employing authentic assignments and aligning instruction and learning with them (Biggs, 1996), efforts were made to enhance the desired outcomes of the assessment and improve its consequential validity. Another important reason for using authentic assignments was the rise of generative artificial intelligence (see Ifelebuegu, 2023; Rudolph et al., 2023a, 2023b)

To approach the implementation authenticity of the assessments (Ghosh et al., 2021), students were provided with the opportunities to give and receive feedback and utilise available resources, mirroring the practices in professional settings. The technical proposal assignment was one of the four assignments included in the continual assessment. The students were required to submit a soft copy of their assignment to the designated assessment folders by the specified deadline. Students were informed that their assignments would undergo scrutiny by the Turnitin software, and severe penalties would be imposed for cases of plagiarism and collusion.

There were four assignments for the module: critical reflection, reader response, technical proposal, and oral presentation of the technical proposal. The technical proposal assignment for one of the undergraduate engineering programmes is analysed and discussed in this paper. Prior invaluable studies informed the analysis of the assignment (Bachman, 1990; Bachman & Palmer, 1996; Ghosh et al., 2021; Gulikers et al., 2004; Gulikers et al., 2006;

Kohnen, 2013; Wargo, 2020; Weir, 2005). However, the five-dimensional framework (Gulikers et al., 2004, 2006) that conceptualises authenticity as a continuum was adopted as the main analytic lens.

The technical proposal

The technical proposal assignment analysed in this study can be considered an authentic assessment as it requires students to apply the competencies necessary for writing technical proposals in a real workplace situation (see Appendix). In this paper, the authenticity of the assessment/assignment is defined based on its resemblance to the criterion situation across five dimensions (Gulikers et al., 2004, 2006). The following section discusses the authenticity of the technical proposal assignment in detail.

Task

In the framework, an authentic task is defined "as a task that resembles the criterion task with respect to the integration of knowledge, skills, and attitudes, its complexity, and its ownership" (Gulikers et al., 2004, p. 71). The technical proposal assignment involved a task that is commonly performed by engineers when writing a technical proposal. Students were required to identify a significant problem within a specific system, design, process, procedure, or protocol. They then had to review existing solutions, analysing their strengths and weaknesses. Finally, students were expected to propose a solution and develop a method to test its superior efficiency compared to the existing solutions. These activities closely mirror the process followed by engineers in real-world technical proposal writing. The students were given the freedom to choose their own topic, identify a problem within that topic, conduct a literature review, analyse existing solutions, and propose a new solution along with a method to evaluate its efficiency to give them a sense of ownership over the task. Some individuals contend that assessment authenticity is not entirely objective and is contingent upon students' perceptions of the assessment's connection to real-life situations, its relevance to their future professional success, and its value in facilitating the acquisition of transferable knowledge, skills and attitudes. However, from the author's perspective, student perception represents an individual difference dimension that varies among students and evolves over time. To bolster this facet of authenticity, educators could consider assisting students in recognising the authenticity of the task at hand and its pertinence to their academic pursuits and future careers. A significant concern related to the task dimension of the technical proposal assignment is the students' level of discipline-specific content knowledge. As first-year students, some may struggle to generate innovative solutions and employ effective methods to evaluate the efficacy of their proposed solutions. This concern is consistent with the literature, which suggests that fully functionally authentic tasks can overwhelm beginners and lead to cognitive overload (Sweller et al., 1998).

Physical context

The authenticity of an assignment or assessment, in terms of its physical context, refers to the degree to which the physical conditions, availability of relevant and irrelevant sources, materials, and time resemble those of the criterion situation (Gulikers et al., 2006). The technical proposal is typically written within an office-based physical context, where individuals have access to computers, the internet, and library resources. Students were trained on how to conduct effective library searches to locate relevant articles and received lessons on how to write different sections of the proposal. They were expected to independently distinguish between materials that were pertinent to their proposal and those that were irrelevant. In terms of time, the development and composition of proposals often happen over time, allowing for iterative refinement. In the case of the students, they were given a six-week timeframe to complete and submit their proposals. In a real-world work environment, engineers have the advantage of accessing sites and laboratories to investigate the problems they intend to solve. The students were in their first year of studies, and opportunities for internships and access to labs and sites are typically provided in other discipline-specific modules.

Social context

The social context of authenticity emphasises the similarity between the social processes involved in completing professional tasks and the corresponding assessment or assignment. If the target task is typically performed individually, then the assessment should also be conducted individually. Conversely, if the task is typically completed collaboratively, then the assessment or assignment should be designed to be team-based (Gulikers et al., 2006). The technical proposal assignment was designed as a team-based task to mirror the social processes involved in the development and writing of technical proposals in real-life contexts. In such situations, individuals often present their proposals and receive feedback from others. For the technical proposal assignment, students received peer and instructor feedback. They were then given opportunities to revise and enhance their proposals before submitting the final versions. The approach of providing students with multiple opportunities to revise and refine their work based on feedback received also aligns with the implementation authenticity proposed by Ghosh et al. (2021).

Assessment result or form

The assessment result or form refers to the outcome of the assessment or assignment, which should closely resemble a product or performance that professionals are typically asked to produce or perform (Gulikers et al., 2006). The product/performance should provide the assessor with sufficient data regarding the intended underlying competencies (Darling-Hammond & Snyder, 2000). If a single product/performance does not provide sufficient information about the relevant competencies, a series of assessments or assignments should be completed by the students (Darling-

Hammond & Snyder, 2000). Additionally, it is beneficial for students to present their work to an audience, similar to what professionals typically do, to demonstrate the authenticity of their mastery of knowledge, skills, and attitudes (Wiggins, 1989). The written text of the technical proposal closely resembled the proposals that professionals typically write. It provided ample information about the target competencies. Moreover, students were required to present their proposals to other students and respond to queries from both their peers and the lecturer to demonstrate that their mastery was genuine. However, it is important to note that in real-world contexts, the audience for technical proposals is typically supervisors or professionals who evaluate the quality of the proposal, rather than fellow students or lecturers. In this regard, the assignment can be considered latently authentic as it required students to utilise metacognitive and cognitive processes similar to those used when presenting authentic professional tasks to genuine audiences.

Criteria and standards

The criteria used should accurately reflect the underlying competencies necessary for successfully performing the target tasks in real-life situations, and the levels assigned to the criteria should correspond to the progressive development of these competencies (Darling-Hammond & Snyder, 2000). In most professional contexts, employees are typically aware of the criteria used to assess their performance. Similarly, it is essential for students to gain a deeper understanding of the marking criteria for performance outcomes in advance. The technical proposal was assessed by using a rubric with content, organisation, and language use criteria and five levels of development, namely exemplary, proficient, competent, developing, and beginning. To approach real-life situation scenarios, the lecturer unpacked the rubric for students to learn the criteria and developmental levels and strive for their best performance.

Conclusion

The users of artefacts, such as authentic assignments and assessments, typically internalise the knowledge, skills, and attitudes embedded in these artefacts. This internalisation can later be externalised and applied in the development of new artefacts (McAvina, 2016). By engaging with analyses of these artefacts, prospective and novice users can expedite the process of externalising the knowledge and skills embedded in them. This practice, in turn, can enable the users to apply the artefacts more effectively and potentially create new ones. This paper aims to externalise the knowledge, skills, and attitudes necessary for developing and implementing authentic writing assignments and assessments. Its objective is to facilitate the analysis of existing assignment and assessment tasks, their implementation, and the development of new ones. To achieve this goal, a technical proposal assessment is analysed in this paper by using the Five-Dimensional Framework (5DF) for authentic assessment (Gulikers et al., 2004) as the primary analytical lens. The analysis of the technical proposal assignment through the lens of the authenticity framework highlights the value of using this framework as a guide for developing authentic writing assignments and assessments suitable for students' ability levels.

Further, it became evident that achieving functional authenticity in all dimensions may not always be viable and that developing fully functional authentic writing assignments/assessments for beginners is not always advisable, as performing such assignments may lead to cognitive overload for them (Sweller et al., 1998). Therefore, it is recommended that educators consider both the dimensions of authenticity and students' ability levels when analysing, developing, and using authentic writing assessments and assignments. Indeed, striking an appropriate balance between authenticity and students' cognitive capacity is crucial for creating optimal learning experiences.

The paper contributes to the understanding, development, and implementation of authentic assessments and assignments in the context of writing. It explores how to use an authenticity framework as a valuable guide to design and develop authentic assignments that mirror real-world technical communication scenarios. It provides a worked example of applying the authenticity framework as a systematic approach for evaluating the authenticity of writing tasks, materials, and assessments. This paper may benefit instructors and curriculum designers by providing insights into creating meaningful and relevant instructional and assessment materials and experiences for students. Additionally, the findings and insights from the paper might be relevant and applicable to contexts beyond technical proposal writing. The applied authenticity framework and pedagogical implications can be adapted and applied to other genres and domains that can extend the reach and impact of the paper, benefiting a broader range of educators.

References

Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Bachman, L., & Palmer, A. (1996). *Language testing in practice.* Oxford University Press.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education. 32*,347-364.

Birenbaum, M. (2003). New insights into learning and teaching and the implications for assessment. In M. Segers, F. J. R. C. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards*. Kluwer.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment in teaching in context. *Teaching and Teacher Education, 16*(5-6), 523-545. https://doi.org/10.1016/S0742-051X(00)00015-9

Ghosh, S., Brooks, B., Ranmuthugala, D., & Bowles, M. (2021). Investigating the correlation between students' perception of authenticity in assessment and their academic achievement in the associated assessment tasks. *The Journal of Navigation, 74*(2), 293-310.

Gibbs, G. (1992). *Improving the quality of student learning.* Technical and Educational Services.

Gulikers, J., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development, 52*(3), 67-86.

Gulikers, J. T., Bastiaens, T. J., Kirschner, P. A., & Kester, L. (2006). Relations between student perceptions of assessment authenticity, study approaches and learning outcome. *Studies in Educational Evaluation, 32*(4), 381-400.

Herrington, J., & Herrington, A. (1998). Authentic assessment and multimedia: how university students respond to a model of authentic assessment. *Higher Educational Research & Development, 17*, 305-322.

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2), 1-8. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.2

Kohnen, A. M. (2013). The authenticity spectrum: The case of a science journalism writing project. *English Journal, 102*(5), 28–34.

McAvinia, C. (2016). *Online learning and its users: Lessons for higher education*. Chandos Publishing.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*, 13-23.

Paul, R., & Elder, L. (2019). *The miniature guide to critical thinking concepts and tools.* Rowman & Littlefield.

Preston, R., Gratani, M., Owens, K., Roche, P., Zimanyi, M., & Malau-Aduli, B. (2020). Exploring the impact of assessment on medical students' learning. *Assessment & Evaluation in Higher Education, 45*(1), 109-124.

Rudolph, J., Tan, S., & Tan, S. (2023a). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Rudolph, J., Tan, S., & Tan, S. (2023b). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), 342-362. https://doi.org/10.37074/jalt.2023.6.1.9

Schuwirth, L. W., & Van Der Vleuten, C. P. (2004). Different written assessment methods: What can be said about their strengths and weaknesses? *Medical Education, 38*(9), 974-979. https://doi.org/10.1111/j.1365-2929.2004.01916.x

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Van Merrienboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Educational Technology Publications.

Wargo, K. (2020). A conceptual framework for authentic writing assignments: Academic and everyday meet. *Journal of Adolescent & Adult Literacy, 63*(5), 539-547.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46*(7), 41-47.

Appendix

Critical thinking and communicating assignment: Writing technical proposals

Overview

Type: Team-based Project (Groups of 4-5 Students)
Weighting: 30%
Due: Week 13

Learning outcomes

This assignment provides you with an opportunity to showcase your knowledge, skills, and attitudes in technical proposal writing. You will be writing a proposal addressed to your lecturer, who will play the role of a corporate audience. You should write in the academic tone and style and cite sources following the APA 7th edition guidelines.

Task

For this assignment, you are tasked to work on a team project to identify and elaborate on a specific system/design/process/procedure/protocol that needs improvements. Then, perform secondary research and write a literature review about the specific system/design/process/procedure/protocol and solutions that are currently available to improve it. Next, select one or a combination of existing solutions or develop a solution. In the Body of your technical proposal, explain your solution and the rationales for your decision and the specific objectives that you would like to achieve by the improved system/design/process/ procedure/protocol. You should also discuss the contributions of your solution to sustainability as compared to those of the current system/design/process/ procedure/protocol in the body of your proposal. In the Methods section, elaborate on and justify the method(s) that you propose to evaluate the efficiency and sustainability of the proposed improved system/design/process/ procedure/protocol.

Based on the selected topic by the team, write a proposal of not more than 2000 words (from the Introduction through the Conclusion) with the following sections:

1. Introduction
2. Literature Review
3. Body
4. Methods
5. Conclusion

Submission

Please ensure that your assignment is typed in either Times New Roman, Arial, or Calibri font, size 12, with double spacing between lines. Please submit an electronic copy of your assignment to the designated assignment Dropbox folder before the deadline mentioned earlier. It is important to note that your assignment will be subject to an automated check using Turnitin software, and any instances of plagiarism or collusion will result in severe penalties. The softcopy of your assignment should include your proposal, and the file should be named as follows: (assignment name) _ (group name) _ (names of group members).

For late submissions of any assignment without incurring penalties, a written application must be submitted prior to the deadline. If the right for a late submission is not granted, the assignment may still be submitted up to 4 days after the original deadline, but with a penalty of 15% per day. Please note that submissions made 4 days after the original deadline will not be awarded any marks.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Book review. Popenici, Stefan (2023). Artificial intelligence and learning futures. Critical narratives of technology and imagination in higher education. Routledge.

Jürgen Rudolph[A]

A Director of Research, Kaplan Singapore

DOI: https://doi.org/10.37074/jalt.2023.6.2.27

Introduction

The poignant premise of Stefan Popenici's excellent book *Artificial Intelligence and learning futures* is that the English-speaking world faces multiple crises in higher education: ideological, intellectual, managerial, and ethical. Popenici argues against an all-too-common 'solutionism' or techno-chauvinism – by those terms, he and others refer to the prevailing belief that technology, particularly AI, is the ultimate solution to all these crises. His book delves into such an uncritical and naïve mindset within higher education and explores its underlying reasons.

AI deeply influences our lives, shaping our knowledge, perceptions, and worldviews. To reshape the narrative of higher education, Popenici argues that we must discern our aspirations and what influences them. By doing so, we can determine how to steer universities and their stakeholders away from a surveillance-driven, authoritarian dystopia. Such a 're-storying' (see Chapter 9) could pave the way for an educational vision that fosters a sustainable future and for teaching and learning in higher education institutions.

Artificial intelligence and learning futures consists of three sections subdivided into three chapters each. Although the book has only over 200 pages (including extensive references and a helpful index), it feels much more voluminous. Such a sentiment is not without reason: The publisher (Routledge) has put many words on the book's pages (a rough calculation of the words on a randomly chosen single page resulted in 550 words, which, assuming the generalisability of that random sample, would mean that the book has way in excess of 100,000 words), and the content is challenging – in the best sense of the word. Due to my being extremely busy, it took me a few weeks to read the book (and quite a few months to write this long-overdue book review). However, this bite- (or byte-)sized approach gave me a lasting impression of the book's major ideas and significantly changed my thinking on AI and higher education. In any event, the book does not render itself particularly well to speed-reading. Its three main sections are titled "Education, Artificial Intelligence, and ideology", "Higher learning", and "The future of higher education". Before critically evaluating it, I am offering my biased summary of its nine chapters.

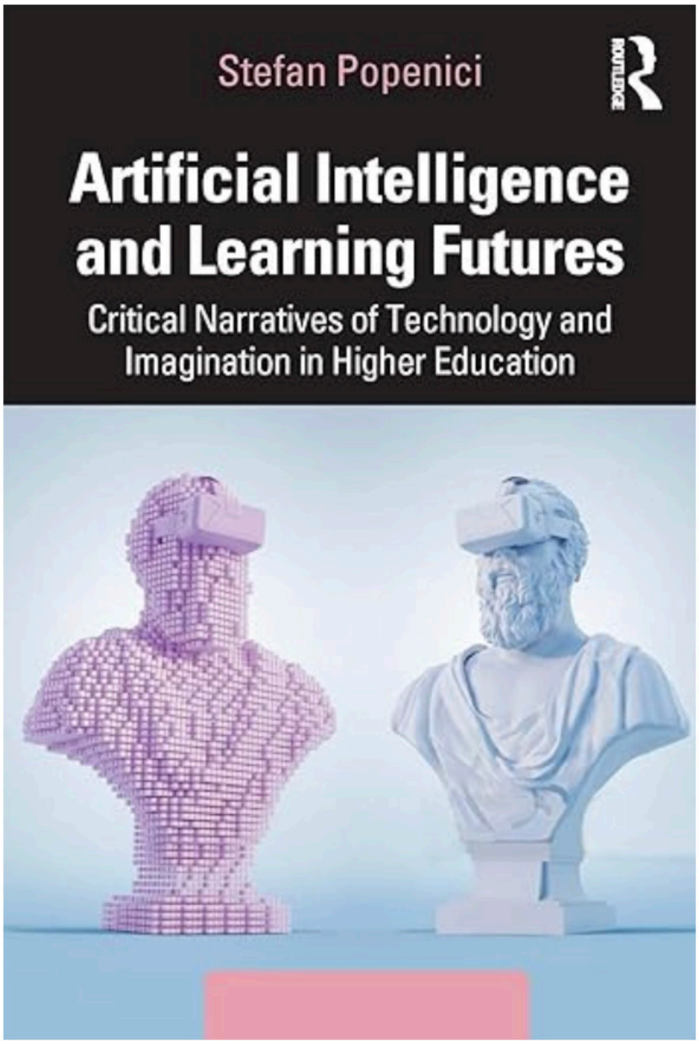


Figure 1: Book cover.

First, however, an introduction of the book's author, Stefan Popenici, is in order. He has a 25-year tenure in higher education encompassing teaching, research, and leadership across universities in Europe, North America, Southeast Asia, New Zealand, and Australia. Popenici is a respected scholar and sought-after public speaker whose significant contributions to education earned him the 'Merit of Education' Order from the President of Romania. His research centres on the implications of artificial intelligence in higher education teaching and learning, as well as on quality assurance and student involvement. I should disclose that, together with Shannon Tan and Samson Tan, I interviewed the author (Popenici et al., 2023), and he has also written an excellent opinion piece for JALT (Popenici, 2023) – so this review is not without bias, and it could even be regarded as a companion piece to the aforementioned articles.

Education, Artificial Intelligence, and ideology

The first section of *Artificial intelligence and learning futures* delves into the concept of 'intelligence', tracing its ideological roots and its connection to the development of AI, particularly its application in education. It emphasises the significance of understanding the ideology behind intelligence for a comprehensive grasp of the swiftly evolving AI field, noting AI's ties to the 'Californian ideology' and its blend of vast opportunities and risks. Succinctly put, Popenici characterises the 'Californian Dream' as a tech-utopianism rooted in Ayn Rand's extreme individualism, advocating for unchecked resource exploitation and offering technological solutions as an escape route to other planets amid escalating climate crises. As higher education globally adopts AI for various purposes, it is crucial to examine the relationship between technology's colonising force and the inherent colonisation in the American model. The first section consists of three chapters: "The ideological roots of intelligence", "Imagination, education, and the American dream", and "The narrative construction of AI".

The first Chapter, "The ideological roots of intelligence", shows that the discourse on using intelligence – and later, AI – in modern society and higher education is deeply rooted in the eugenics theories of the late 19th and early 20th centuries. Galton propagated appalling racist ideologies, advocating for the control and elimination of what he deemed 'inferior groups', which included non-whites, the poor, and those he classified as 'criminals and semi-criminals'. His works laid a pseudo-scientific but highly influential foundation for the investigation of intelligence. Following in Galton's footsteps was Pearson, who justified widespread genocide against the First Nations in America using a dehumanised rationale. In another instance, he argued against integrating Jewish populations, labelling them as mentally and physically 'inferior' to the native populace.

This eugenic perspective gained traction in the United States in the 1920s, spurred by the intellectual frenzy surrounding the pseudo-scientific laws of heredity, which aimed to 'perfect' humanity. This period saw the emergence of theories that advocated controlling the reproduction of the 'feebleminded' and 'unintelligent', a stance widely adopted by American politicians, industrialists, and academics during the first half of the 20th century.

There is a little-known and deeply disturbing connection between American eugenics theories and the nefarious ideologies of Nazi Germany. Shockingly, even the Nazis found some US race laws too extreme to adopt, as the book documents. This era firmly established intelligence as a potentially dangerous concept utilised to justify controlling or eradicating groups seen as undesirable or threatening to the political elite. Since its inception by Galton, the concept of intelligence has been weaponised to serve the monopolies of social, economic, and political power. Popenici argues that even in the 21st century, the concept remains closely tied to eugenics. These extremely problematic ideological roots have opened up opportunities for misuse, for instance, by promoting class and ethnic discrimination.

The ideas formed in the eugenics era have shaped the development and implementation of AI, machine learning, and data analytics. Coined by John McCarthy in 1956, the term 'artificial intelligence' represents not just a set of algorithms but an ideological and political project inherently linked to its eugenic and elitist roots. The current trajectory of AI development, largely fuelled by American military investments and the interests of a techno-elite, risks perpetuating class and ethnic discrimination, posing significant challenges to fostering equity, transparency, and democracy. This direction is especially concerning in the field of education.

The second Chapter, "Imagination, education, and the American dream", links the challenges in defining intelligence to AI's susceptibility to exaggerations and misuse. It is refreshing that Popenici chooses a decidedly sobering and unenthusiastic definition, citing Bartoletti (2020, p. 23):

> To put it simply, AI is (so far at least) about machines performing a task that humans perform and which is possible only because we, humans, have taught them to do so. The thing we program them to do is to recognize and act upon the correlation between things (*intelligere*); things that for us, humans, make up some part of what constitutes life and experience.

The Internet, facilitated by companies such as Meta (formerly known as Facebook) and YouTube (owned by Alphabet, the parent company of Google), has become the most potent propaganda machine in history. These platforms' algorithms purposefully accentuate content that sparks outrage and fear, keeping users engaged. Major tech corporations like Alphabet and Meta refrain from revealing the inner workings of their algorithms. While we can ascertain the input and output of these systems, the intermediate process, termed the *black box* of technocracy, remains inaccessible to all but a handful of engineers (see Pasquale, 2015).

Popenici provides evidence for the existence of a Silicon Valley (or machine) religion that worships digitalism. Digitalism embodies the belief in surpassing human limitations, including death, through technological advancements. It envisions a redemption from the constraints of human brains and ageing bodies, akin to religious salvation. The Silicon Valley religion harbours an anti-human sentiment deeply embedded in major tech companies' technologies and business strategies that serve only a few 'chosen' ones, the technological and economic elites.

This machine religion positions AI as the beacon of hope via a narrative entrenched in 'solutionism'. AI has long been marketed as a catalyst for devising solutions to create a better world. This 'solutionist' approach simplifies complex social issues into computable problems solvable by implementing the right algorithms. However, AI algorithms have sometimes produced discriminatory and erroneous outcomes with severe real-world repercussions, as evidenced by Amazon's Rekognition software misidentifying black members of the US Congress as criminals. It is also apt that Popenici reminds us of an unsung Soviet hero, Stanislav Petrov, who, by disregarding a false alarm from a satellite warning system, the most advanced technology of its time, prevented a nuclear Holocaust by not pressing the famous red button, risking his life in a totalitarian state and showcasing that technology can be fallible and that human judgment remains necessary.

The third Chapter, "The narrative construction of AI", reminds us, amongst other things, that the current AI solutionism is nothing new. It is, in fact, reminiscent of the incredible hype around MOOCs a decade ago (see Rudolph, 2014). MOOCs quite spectacularly failed in their promise of free and inclusive higher education. They were hyped as the remedy for perceived issues in the educational system, spurred by Silicon Valley narratives that technology could fix a 'broken' education system and provide free higher education for all. This led to a frenzied competition among universities to allocate substantial resources to this trend. However, a reality check shows MOOCs' limited impact. Despite the initial optimism, it became evident that MOOCs failed to revolutionise higher education. Contrary to the aspirations of democratising education, the courses mostly attracted individuals who were already graduates, thereby not addressing the educational needs of the underprivileged in dire need of educational opportunities, or the most significant problems confronting higher education.

Higher learning

The second section of the book is titled "Higher learning". It explores the ongoing identity crisis in higher education by examining the consequences of incorporating neoliberal ideologies in academia, including the rise of anti-democratic tendencies that prioritise profits over educational objectives. This trend, amplified by technology advancements, fosters increased surveillance of students and faculty alongside the commodification of education. The sector is also grappling with the influences of Americanisation and the inconsistent integration of market strategies, exacerbating the stress of audit cultures and the metrification of academic pursuits. Finally, the book's middle section strives to contextualise the role of AI within educational goals, aligning it with core human values such as an appreciation for learning, beauty, and passion.

In Chapter 4, "Automation of teaching and learning", Popenici persuasively posits that in the early 21st century, universities globally face a paradox where they are simultaneously more numerous and important than ever before yet grappling with an unprecedented crisis of confidence and identity. This crisis is compounded by the neoliberal model of metrification, which, despite failing notably in sectors like finance, for instance, during the 2007-2008 Great Financial Crisis, continues to permeate the education sector. "Bean counters and accountants decide what is real, but behind them is a grab of power with an unprecedented capacity to colonise and incorporate all aspects of our lives" (p. 94). The approach of treating education as a marketable commodity emerges as a destructive force for the future of society.

The metrification of higher education poses a stark shift from genuine learning to a focus on test results and quantifiable outcomes, leaving education bereft of depth and relevance. Current educational environments pressure even the most eager students to learn merely for test performances, side-lining a comprehensive education's intrinsic joy and vigour. This pervasive metrification emphasises quantifiable judgments and fosters a culture of surveillance and metrics at the expense of substantial academic growth, instilling a culture of 'efficiency' and suspicion that breeds increasing levels of stress and anxiety in the academic community. It is a fallacy that only what can be counted counts. I was reminded of a quote frequently attributed to Albert Einstein: "Not everything that can be counted counts, and not everything that counts can be counted".

Popenici's critique of metrification certainly struck a chord with me. It is a symptom of the many things that have gone wrong in higher education and a theme that has been explored in many publications (see, for instance, Parker, 2018; Tourish, 2019; Fleming, 2021; Fleming et al., 2021; Parker et al., 2021; Brookfield et al., 2024). I would have been inclined to see the beginnings of the neoliberal era and its by-product of metrification in higher education with the Reagan and Thatcher governments (see Fleming, 2021). Popenici's additional bringing of the Democratic presidency of Bill Clinton into the metrification picture was nothing short of revelatory.

Since the Clinton era, the US education model has become infatuated with performance-based accountability, emphasising universities' return on investments (ROI). Institutions like the OECD and the World Bank have been fervent advocates of this model, fostering a broader Americanisation of the global education landscape. However, this model seems to cultivate an elite corporate class while relegating learning to a secondary role and leaving the majority in a struggle for survival amidst a backdrop of falling education standards and an ongoing identity crisis in universities.

In this chapter, Popenici returns to the book's pervasive theme of AI's false promise as a panacea. AI, exemplified by IBM's Watson, has been touted as a universal solution for numerous societal issues ranging from healthcare to education. Despite the relentless promotion and high expectations, the advancements have been considerably less impressive.

In Chapter 5, "Surveillance, control and power – the AI challenge", Popenici refers to OECD's important and critical study on EdTech (2015). The report reveals that while moderate computer usage in schools can be beneficial, "resources invested in ICT for education are not linked to improved student achievement in reading, mathematics or science" (p. 104). Importantly, in countries where Internet use for school assignments is less prevalent, improvements in reading were observed to be faster.

Popenici underscores that the liberalisation and privatisation of education have restricted intellectual freedom, encouraging rigid managerial control structures and predisposing universities to mediocrity and groupthink tendencies. In the contemporary university setting, EdTech is being employed as "technologies of domination", as AI vigorously encourages students and teachers to acquiesce to manipulative oversight and surveillance, fostering a culture of apathy and resignation (p. 115). The sixth Chapter picks up from the third, where Popenici already touched on the perils of surveillance in higher education. Incorporating pervasive surveillance as a fundamental component is "poisonous for any educational project" (p. 60). Rather than relying on software solutions to combat plagiarism, it is more prudent to address the issue's root causes. Viewing students as potential "thieves" and "criminals" requiring constant monitoring contradicts the essential ethos of higher education (p. 60).

Chapter 6 is titled "Beauty and the love for learning". It is a critically important argument of Popenici's book that the current predicament in the higher education sector goes way beyond its technological issues; it is deeply embedded in a series of interlinked crises that span the intellectual, moral, and ontological spheres. These crises manifest themselves as a stark escalation in climate change, driven by unrestrained exploitation of the Earth's resources and the exacerbation of political and cultural instability, fostering wars and escalating violence globally. Moreover, the COVID-19 pandemic has unveiled deep-rooted issues in even the most developed nations, showing a startling lack of compassion, civility, and education, as market profits have overtaken the importance of life itself. Despite the availability of scientific solutions, significant segments of the world's population seem disinterested or unable to grasp the gravity of the situation, displaying an inability to comprehend how science could construct a collective response to our looming challenges. The 'polycrisis' has become a 'perma-crisis'.

At this critical juncture, the focus of higher education should be realigned. Instead of solely advancing technology, it should contemplate the true objectives of higher learning: fostering a civil society, nurturing a wise citizenry, and working towards creating sustainable futures for everyone. Universities should emphasise genuine education over mere credentialism and engage actively in shaping a better collective future. However, there is a worrying move in a direction that unduly focuses on technological solutions that miss the essence of education. In Popenici's view, an example of this is the booming industry centred around plagiarism detection, which operates on a fundamental distrust of students and represents a glaring failure in the educational process. These measures, often lacklustre in their effectiveness, signify a pivot towards commercialisation and a detachment from the true objectives of education.

Popenici argues that the increasing reliance on EdTech, particularly AI, seems to be driven by a baseless optimism without considering its limitations and the responsibilities it entails for educators. This blind faith threatens to overlook the integral aspects of humanity – love, beauty, passion, and inspiration – turning the educational process into a commercial, hollow, and artificial endeavour that side-lines the innate human elements that should be at its core.

Therefore, it is imperative to approach the integration of AI in education with cautious scepticism, scrutinising the claims of financial consultants and corporate behemoths who stand to profit from the EdTech market. Furthermore, an in-depth analysis of the 'EdTech imaginary' is necessary to critically evaluate how new technologies influence teaching and learning processes, especially considering the opaque nature of AI and its implications in data collection and utilisation.

To circumvent our ontological decline, it is vital to resist being overwhelmed by the dazzling technological advances to the point where calculative thinking becomes the sole accepted method of reasoning. With repeated references to the German philosopher Martin Heidegger, Popenici posits that preserving diverse forms of critical thinking is essential to prevent succumbing to the bewitching allure of technological advancements, which threaten to eclipse other vital forms of intellectual engagement.

The future of higher education

The book's final section is titled "The future of higher education". It delves into the potential role of imagination in education and examines the intersection of intelligence, imagination, and AI. While considering the prospective directions of education, it underscores that the principal hurdles at the outset of the 21st century for universities and open societies are political, educational, and cultural rather than technological. Our current period of technological acceleration is marked by a concerning global surge in authoritarian ideologies and a growing socioeconomic and cultural divide. It foresees AI becoming a central element in the future trajectory of education, posing opportunities and challenges for educational institutions. The final section proposes vital guidelines for responsibly incorporating AI in higher education, aiming to guide educators and students towards an ethical and productive integration of AI systems in the quest for purposeful education.

In Chapter 7, "Imagination and education", Popenici dismisses the belief that the eugenic foundations of AI are relegated to history as naïve and unrealistic, as they can be observed in current political strategies and technological projects. Currently, the discussion surrounding AI in education is largely governed by commercial narratives, neglecting essential education objectives such as empathy and compassion and thus risking the fostering of a society prone to cruelty and greed, devoid of human values. The major challenges facing humanity are cultural and moral rather than technological. This context reflects a society engulfed in transient and superficial information, promoting fleeting interests, which can be seen as a colonisation of the educational sphere by corporate interests, pushing tech solutions and neoliberal policies. Although Popenici does not appear to refer to Habermas (1985) in this discussion explicitly, it has the latter's discourse on the overall colonisation of the lifeworld (*Lebenswelt*) written all over it.

The modern higher education narrative is dominated by a business-oriented approach, side-lining the importance of imagination, inspiration, and empathy. The corporate sector's influence has led to an education system deeply entrenched in technocracy and commercial interests, prioritising profitability and efficiency over fostering imaginative and empathetic minds. This shift signals a dangerous colonisation of the educational imagination, where the field is now overwhelmed with cynical and profit-driven approaches, likening educational institutions to profit-centric businesses, thereby undermining the rich history and contributions of academic institutions to society.

In Chapter 8, "Scenarios for higher education", Popenici reiterates his point that in light of escalating crises, it is crucial to scrutinise the actual effectiveness of EdTech and AI in enhancing education, particularly when the successful outcomes have been claimed ad nauseam, but the actual quality of higher education seems to be deteriorating. Since the 1990s, there has been a repetitive narrative surrounding AI as a revolutionary tool for educational institutions. However, in Popenici's argument, this alleged 'innovation' appears to be a tactic to dominate and monetise educational spaces, with little real progress observed over the decades.

It is telling that Popenici finds a speech by the comedian Sacha Baron Cohen to be of particular value in this context. Cohen highlighted the pernicious effects of social media platforms, which use algorithms to amplify content that engenders fear and outrage, often disseminating falsehoods and promoting hate at a pace faster than truth can spread, thus serving as formidable propaganda machines. Popenici also observes a worrying trend of diminishing human intelligence as AI technologies advance. The essence and purpose of universities are under threat, with a shift towards commercialisation and the diminishing of critical thinking and democratic values amongst graduates. This is exacerbated by university leaders adopting a corporate approach, prioritising business interests over intellectual growth and enriching learning environments.

The looming threat of the commodification of education is evident in the adoption of business models from exploitative corporations and the employment of AI primarily for profit maximisation and cost reduction rather than fostering higher cognitive skills and meaningful learning. Lastly, AI's pervasive influence in daily life, as demonstrated by Facebook's (now Meta) developments, underscores the potential for manipulation and the propagation of harmful stereotypes and discriminatory content, posing a considerable risk to societal values. The current educational landscape is marred by an uncritical embrace of technology, thus contributing to the concurrent decline in intellectual engagement, civic relevance, and true educational value.

In the book's final chapter, "Re-storying higher learning", Popenici cleverly summarises that the marketing plot of AI propels the idea of an unavoidable tech-driven evolution in education, promising a personalised approach akin to a Netflix (or Amazon) model of higher education. In contrast, he offers five guiding principles for the responsible and constructive use of AI.

The first guiding principle urges higher education institutions to meticulously evaluate the budgetary and ethical aspects of data collection and aggregation, considering potential costs, privacy concerns, and legal implications while actively involving students in the process before implementing AI systems. The second principle emphasises that universities must constantly scrutinise AI's inherent biases and corporate influences, urging them to prioritise intellectual curiosity and critical thinking over adopting exploitative EdTech, thereby preventing a potential drift towards authoritarianism and narrowed educational experiences. The third principle states that while AI can serve as a valuable tool for facilitating access to information and administrative tasks in education, it should not replace human educators, as it lacks the ability to offer a nuanced, meaningful, and creatively stimulating education that fosters independent thinking and responsible citizenship.

The fourth principle emphasises that the effectiveness and fairness of AI in education are deeply influenced by the social, economic, cultural, and political contexts, and without considering these variables, the use of AI can potentially be more detrimental than beneficial in achieving educational and organisational objectives. The final principle highlights the necessity to carefully select the appropriate EdTech/AI solution that aligns with the institution's goals and fosters a meaningful education that cultivates responsible, engaged, and mentally agile members of society without diverting focus from the core mission of nurturing well-rounded individuals.

Towards the end of his book, Popenici proposes a pledge for academics to adopt AI in education. This pledge emphasises committing to an equitable, compassionate, and reflective pedagogy prioritising students' learning needs and rights. It encourages fostering an environment of intellectual curiosity and moral development, thereby nurturing mutual trust between students and educators. Moreover, the pledge advocates the protection of students from intrusive data practices, coupled with a continuous effort to enhance the quality and inclusivity of education, fulfilling the profound responsibilities inherent in the role of academics.

A critical evaluation

My review is extensive by design, as outstanding books like this one are few and far between, necessitating a widespread exploration of its significant themes. I am unable to pinpoint any shortcomings within this masterful work. It is a treasure trove of well-curated references, showcasing Popenici's generous and due acknowledgement of his sources. With its firmly grounded humanistic approach, the book vehemently opposes authoritarian inclinations found in both the political right and left. Bridging critical analysis with humour and a palpable sense of hope, *Artificial intelligence and learning futures* provides a rich, multifaceted reading experience.

Despite the advent of the ChatGPT 'revolution' after the book's completion, its relevance remains unscathed. Popenici's insightful critiques resonate profoundly with the ongoing developments in chatbot technologies and related phenomena, a trajectory he has pursued further in subsequent publications (Popenici, 2023; Popenici et al., 2023). I wholeheartedly endorse this pivotal book, urging a broad readership encompassing governmental and educational leaders, academics, postgraduate students, and anybody interested in the domains of AI and EdTech to immerse themselves in its contents.

In an era where the prevailing crises are increasingly conspicuous, the necessity to redefine the essence of a well-rounded higher education has never been as important as it is now. Surpassing the confines of mere job readiness or intellectual prowess, we must cultivate an educational ethos centred on holistic development. It is imperative to move beyond the mere quantitative assessment propagated by current neoliberal and technocratic ideologies to cultivate a generation of individuals who are not just well-informed but truly well-educated.

Additional references

Bartoletti, I. (2020). *An artificial revolution: On power, politics and AI*. Black Spot Books.

Brookfield, S. D., Rudolph, J., & Tan, S. (2024). *Teaching well: Understanding key dynamics of learning-centered classrooms*. Routledge. Ahead of print.

Fleming, P. (2021). *How universities die*. Pluto Press.

Fleming, P., Rudolph, J., & Tan, S. (2021). 'Never let a good crisis go to waste'. An interview with Professor Peter Fleming on dark academia, the pandemic and neoliberalism. *Journal of Applied Learning and Teaching, 4*(2), 110-120. https://doi.org/10.37074/jalt.2021.4.2.14

Habermas, J. (1985). *The theory of communicative action*. Beacon Press.

OECD. (2015). *Students, computers and learning. Making the connection*. OECD Publishing. https://www.oecd.org/publications/students-computers-and-learning-9789264239555-en.htm

Parker, M. (2018). *Shut down the business school*. Pluto Press.

Parker, M., Rudolph, J., Shukaitis, S., & Tan, S. (2021). Strategic utopianism and the avoidance of dualisms: An interview with Martin Parker. *Journal of Applied Learning and Teaching, 4*(Sp. Iss. 1), 60-73. https://doi.org/10.37074/jalt.2021.4.s1.3

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Popenici, S. (2023). The critique of AI as a foundation for judicious use in higher education. *Journal of Applied Learning and Teaching, 6*(2), 1-7. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.4

Popenici, S., Rudolph, J., Tan, S., & Tan, S. (2023). A critical perspective on generative AI and learning futures. An interview with Stefan Popenici. *Journal of Applied Learning and Teaching, 6*(2), 1-21. Advanced online publication. https://doi.org/10.37074/jalt.2023.6.2.5

Rudolph, J. (2014). *Massive Open Online Courses (MOOCs) as a disruptive innovation in higher education?* Master of Education dissertation, University of Adelaide. https://www.academia.edu/15137335/Massive_Open_Online_Courses_MOOCs_as_a_Disruptive_Innovation_in_Higher_Education

Tourish, D. (2019). *Management studies in crisis: Fraud, deception and meaningless research*. Cambridge University Press.

Vol.6 No.2 (2023)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

Content Available at : http://journals.sfu.ca/jalt/index.php/jalt/index

Book review. Rajaram, Kumaran (Ed., 2023). *Learning intelligence: Innovative and digital transformative learning strategies. Cultural and social engineering perspectives*. Springer.

Jürgen Rudolph[A]

A Director of Research, Kaplan Singapore

DOI: https://doi.org/10.37074/jalt.2023.6.2.3

Introduction: the future of learning

Learning intelligence: Innovative and digital transformative learning strategies aims to provide guidance and frameworks for navigating complex learning environments in the context of digital transformation and innovation. The term "learning intelligence" pertains to an institution's ability to innovate, transform, and enhance its learning capabilities. This includes its ability to establish and articulate learning objectives, understand the rapidly changing learning demands and practices, and measure learning outcomes effectively.

Unusually, the book has one editor and only two authors: Kumaran Rajaram (the editor) and Samson Tan. The duo addresses the challenges and opportunities of nurturing future knowledge workers in higher education. They advocate a shift in the learning culture across different contexts and disciplines. Their book proposes that we ask ourselves: "Why are we teaching people, what we are teaching, and why do we value our current system of educating human beings as the best, and as the most wholesome, accurate way of assessing the intelligence of a human being?" (p. 14).

Learning intelligence aims to address four key trends. (1) There is a significant rise in cooperative learning approaches, where students actively participate in shaping their own educational journey. This is crucial for preparing students for future challenges, enabling them to take control of their learning process throughout their lives. (2) Technology is becoming a crucial tool that allows learning to take place without the constraints of time or location. (3) Teachers are beginning to adopt more tailored and personalised learning strategies for each student. (4) The objective of evaluations is shifting from merely achieving high grades to a more comprehensive understanding of learning.

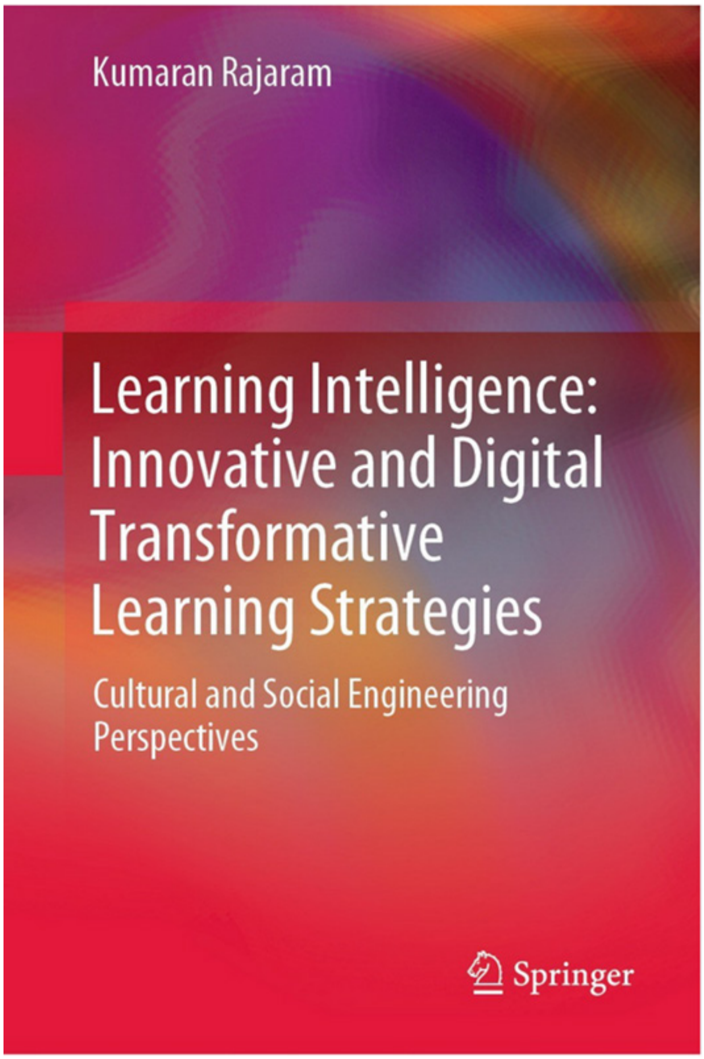


Figure 1: Book cover.

The book consists of ten chapters, organised into four main sections: Future of learning, cultural and social engineering of learning, innovation and transformation in learning, and digital transformation and data analytics in learning. Each chapter covers a specific topic related to learning intelligence, such as teaching and learning strategies, cultural intelligence, social-psychological intervention, disruptive innovation, blended learning, authentic learning, immersive technologies, artificial intelligence, assessment and feedback, and concluding thoughts.

Rajaram and Tan's book draws on a wide variety of theories, models, concepts, and examples from different fields and domains to illustrate the key ideas and principles of learning intelligence. It also provides practical suggestions and recommendations for implementing effective learning interventions and designing engaging learning experiences. The book is intended for a wide range of readers, such as academics, senior management of higher education institutions, corporate leaders, policy makers, researchers, students, and lifelong learners.

The beautiful hard-cover book has a whopping 481 pages. The detailed contents page alone spreads over nine pages and the book contains numerous figures and tables. The book comes at a hefty price of €109.99 (the e-book is available at €93.08). This is a rich tome and it constitutes a rather heroic effort by Kumaran Rajaram and Samson Tan. There are impressive reference lists at the end of each chapter. The breadth and ambition of this book are certainly admirable, and a brief review cannot do justice to the book's many fascinating aspects, hence forcing me to be selective, especially in my discussion.

A brief introduction of the authors is in order. Dr Kumaran Rajaram is a Senior Lecturer with the Leadership, Management and Organization Division at the Nanyang Business School, Nanyang Technological University, in Singapore. Dr Samson Tan is Director of Regional Strategy & Operations (Singapore) of Civica Asia Pacific; he was previously Head, Centre for Innovation in Learning, of the renowned National Institute of Education, Singapore. I should disclose that Samson and I have co-authored two articles on generative AI and higher education (Rudolph et al., 2023a, b).

The book comprises ten comprehensive chapters that are organised in six parts, laden with concrete evidence and presenting pragmatic approaches for individuals engaged in promoting transformative and forward-thinking educational strategies within socio-cultural environments. I recommend this book to anyone curious about the possible future of higher education trajectories.

Overview

The first chapter, "Future of learning: Teaching and learning strategies", introduces the book's aim – to offer guidance for navigating the complexities of learning environments amid digital innovation. This introductory chapter addresses the current obstacles and opportunities in cultivating future global leaders in higher education, advocating a learning culture shift across various contexts. It outlines the need for change, driven by innovative strategies relating to digital transformation and cultural and social engineering. It identifies research gaps to be covered in subsequent chapters and concludes by underscoring the book's value to various stakeholders while outlining its structure.

The book's second part is about "Cultural and social engineering of learning". It comprises a chapter on cultural intelligence in teaching and learning, while the third chapter focuses on the development of cognitive empathy. The second chapter, "Cultural intelligence in teaching and learning", investigates the role and significance of cultural intelligence (CQ) within higher education. The chapter asserts that CQ is a pivotal skill for effective cross-cultural interaction and collaboration among educators and learners, discussing both challenges and advantages of nurturing CQ. It describes CQ as the adaptability to varied cultural scenarios, introducing a four-dimensional model for its development. The importance of CQ in creating inclusive, engaging learning experiences for diverse groups is emphasised, providing suggestions for its integration into curriculum and pedagogy. It concludes by summarising CQ's implications and identifying future research directions. I found the discussion of Western pedagogical techniques in Asian contexts and the learning styles of Confucian Heritage Cultures (CHC) in this chapter particularly useful.

Chapter three, "Social-psychological intervention: Development of cognitive empathy", delves into the role of cognitive empathy and its significance in higher education. The chapter asserts that cognitive empathy, an ability to comprehend others' thoughts and feelings without necessarily agreeing with them, is integral for effective interaction among diverse learners and educators. Rajaram highlights the role of cognitive empathy in establishing positive educational relationships and fostering active participation in multicultural settings. He provides suggestions for embedding cognitive empathy into educational frameworks.

This chapter reminds us why Goleman, the father of emotional intelligence, considered empathy particularly important. In addition, it highlights the importance of design thinking: "Design thinking is an iterative process that seeks to comprehend the user, challenge assumptions and redefine problems in an attempt to identify alternative strategies and solutions that might not be instantly apparent with our initial level of understanding" (p. 128). The authors of *Learning intelligence* are very good at design thinking themselves, demonstrated by the many helpful figures and tables in their text.

The book's third part focuses on the innovation and transformation of learning and has three chapters devoted to it. The fourth chapter is Samson Tan's first chapter contribution (out of three) to the book. Titled "Exploiting disruptive innovation in learning and teaching", Tan (2023a) discusses the role of disruptive innovation in the realm of higher education. He asserts that such innovation is fundamental for maintaining value and impact in the education sector, discussing the related challenges and opportunities. Disruptive innovation is defined as a process leading to new products or services that outperform and eventually supplant existing ones. The chapter presents a four-stage model of disruptive innovation, explaining each stage's distinct activities and challenges. Tan underscores the necessity for educators and learners to embody disruptive innovation, offering practical recommendations for its implementation. Chapter four brims with intriguing concepts that Tan effortlessly synergises. For instance, the Gartner Hype Cycle is discussed in conjunction with the diffusion of innovation model (p. 170).

The fifth chapter, "Blended learning", examines the significance of blended learning in higher education. It proposes that blended learning enhances educational quality and outcomes when conducted as an effective approach that synergises face-to-face and online learning. The chapter underscores the necessity for educators to design diverse, creative, blended learning experiences and for learners to be self-regulated and collaborative. I found Table 5.1 particularly instructive. The table describes and exemplifies various types of blended learning (such as rotation, station rotation, lab-rotation, individual rotation, flipped classroom, flex, a la carte a.k.a. self-blend, enriched virtual, and low-, medium-, and high-impact blends) and discusses their impact on students' learning.

Chapter six, "Authentic learning digital transformation and innovations", investigates the significance of authentic learning in higher education. It posits that authentic learning, a learner-centred approach facilitating real-world knowledge application, can be effectively implemented through digital transformation and innovations. Rajaram underscores the need for educators to devise diverse, creative, authentic learning experiences and for learners to actively participate. Chapter six features another great table (6.1) that discusses the primary functions of educational technology tools such as K[hat sign]m Alive, Kahoot, Quizizz, Slack, Trello, Google Classroom, Tiki-Tokio, Hypothes.is, and Socrative.

Part four concentrates on digital transformation and data analytics in learning. It contains two excellent chapters by Samson Tan, one on immersive technologies and another on the currently super-hot topic of AI in education. Chapter seven, "Harnessing immersive technologies for innovation in teaching and learning", delves into the role of immersive technologies in higher education. Tan (2023b) posits that these technologies, which create or augment a virtual or augmented reality environment, are potent tools for generating immersive, interactive learning experiences that enhance educational outcomes. Tan highlights the need for educators to design engaging immersive learning experiences and for learners to actively participate in them. Chapter eight, "Harnessing Artificial Intelligence for innovation in education", examines the relevance of AI in higher education. Tan (2023c) asserts that AI offers intelligent, personalised learning experiences, thereby improving educational outcomes. He emphasises the necessity for educators to design engaging AI-based learning experiences and for learners to actively participate.

Part five on assessment and feedback for learning contains a single chapter. The ninth chapter, "Assessment, assessment rubrics and feedback", addresses the critical roles these elements play in enhancing education quality within higher education. The chapter defines assessment as a process for evaluating student progress, introducing a model based on purpose, method, criteria, and quality. It highlights the importance of proficiently designed and facilitated assessment rubrics and feedback, stressing the necessity for learner participation and self-regulation.

The concluding chapter forms the sixth and final part of *Learning intelligence*. Rajaram revisits the book's central themes and their implications for stakeholders like academics, institutional leaders, and students. He re-emphasises the need for learner-centric, real-world approaches and a shift in learning culture. The chapter summarises key points from each chapter, spotlighting concepts, strategies, and models of learning intelligence. Rajaram also provides suggestions to incorporate learning intelligence into curricula, pedagogy, and assessment.

Miscellanea

Tan begins chapter 4 with an important insight:

> With the dawn of the twenty-first century, the world has been in chaos, turmoil and a changing environment that is chaotic and difficult to predict. In the midst of rapid technological advancements, geopolitical shifts, dramatic demographic changes, ecological disasters and immigration, lives are being disrupted at a level of severity and frequency that seems to only increase (Tan, 2023a, p. 149).

Thomas Friedman's (2005) optimistic popularization of the 'flat world' metaphor may have been influenced by him being a billionaire (Sirota, 2006). Piketty's *Capital in the twenty-first century* (2015) and Susskind's *A world without work* (2021) show that an unsustainable global inequality is in the ascendancy. With the recent pandemic, there was also at least a temporary trend toward de-globalisation, which, if were to continue, would lead to decreased diversity.

One of Rajaram's examples for the globalisation of higher education is the cooperation between Yale and the National University of Singapore (NUS) (p. 69). This is a great example of the writer's curse: you give an example and shortly after that, it becomes obsolete. As widely reported, the Yale-NUS partnership has ended (Bloom, 2021).

Rajaram describes social engineering, a term oftentimes used pejoratively, in surprisingly positive terms as the "science of masterfully directing human beings to take action in some aspects of their lives" (p. 226). He relates social engineering to the 'nudge theory' of Nobel-prize-winning authors Thaler and Sunstein. The theory posits that positive reinforcement and indirect suggestions can influence the behaviour and decision-making of groups or individuals more effectively than direct instruction, enforcement, or prohibition. Thaler and Sunstein's (2008) theory's central tenet is that by understanding how people think, we can design choices that help them make better decisions. "Social engineering" is a term often used in the context of manipulating people into performing actions or divulging confidential information, for instance, in the context of malicious activities like scams or cyberattacks. Critics of social engineering argue that it can be seen as manipulative and infringing on individual autonomy, particularly when used without transparency or consent (Mitnick & Simon, 2002). Furthermore, it raises ethical questions about who gets to decide what behaviours should be promoted (Sunstein, 2014).

The authors do an admirable job by ploughing through a plethora of journal articles and other academic literature. Every reader has their own biases and preferences. In my case, I kept thinking of Stephen Brookfield's and Stephen Preskill's work when reading about critical thinking, discussion, and leadership (Brookfield, 1987, 2012; Brookfield & Preskill, 2005, 2016; Preskill & Brookfield, 2009; Brookfield et al., 2019, 2022; Preskill et al., 2023).

We could discuss many small things in the spirit of 'agreeable disagreement'. For instance, was Covid-19 a black swan or a grey rhino event (Rudolph et al., 2021)? Or, to use another example, is the late Clayton Christensen's concept of disruptive innovation so useful for higher education (Rudolph, 2014)? That the book raises many questions plainly shows how rich and admirable Rajaram's and Tan's tour de force is.

Style matters

I could not help but notice the differences in style between the two authors, Kumaran Rajaram and Samson Tan. The Singaporean communication style has occasionally been described as succinct. In my view, however, such a statement is always an over-generalisation, as it does not consider Singapore's main ethnic groups – Chinese, Malays, and Indians – and individual differences. In my 30-year experience in the island nation, there is a statistical tendency for the Chinese to be more succinct and for Indians to be more elaborate. A succinct style is characterised by brevity and conciseness. Individuals who use this style tend to get straight to the point and use fewer words to express their thoughts or ideas.

On the other hand, an elaborate style is characterised by detailed and complex expressions. Individuals who use this style tend to provide more context, use more words, and include more details in their communication. These communication styles are associated with direct and informal versus indirect and formal cultural norms (Gudykunst & Ting-Toomey, 1988). Neither style is inherently better or worse than the other. Style is, by and large, a matter of taste. It was amusing to me that these styles also seem to apply to Rajaram and Tan. Upon reflection, my preference tends toward succinctness, which may be one reason I enjoyed Tan's writings more than Rajaram's.

When observing the style of the book, there are very minor things that many readers may not even notice: for instance, the issue of personal pronouns. Rajaram uses "I", and "we" and occasionally also talks about himself in the third person (e.g. "research scholar Rajaram"). I found Sword's (2012, p. 18) research on academic journal styles in different disciplines fascinating in this context:

> "The high percentages in medicine, evolutionary biology and computer science... confound the commonly held assumption that scientists shun the pronouns I and we in their research writing. By contrast, only 54 percent of the higher education researchers in my data sample and only 40 percent of the historians use first-person pronouns" (Sword, 2012, p. 18).

In academic writing, a single author using 'we' can imply a pluralis majestatis (majestic plural) or a pluralis modestatis (modest plural). If in daily life, somebody spoke about themselves in the third person (he/she/they), we could easily suspect some psychological issues. Perhaps it is time to put such academic conventions to bed and unabashedly use 'I'?

Conclusion

I do not want to appear overly critical, as I like the book, and to state it clearly, I am happy to recommend it. It is the very fact that the book stimulates so many of these questions that its authors deserve much credit for. Rajaram makes this excellent point in the book's preface: "The one true goal of education is to leave a person asking questions". I highly commend the authors of this educative book, as it covers a large ground and helps us rethink learning, teaching and assessment in the light of rapidly evolving technologies. Students with teachers like Kumaran Rajaram and Samson Tan can count themselves lucky, as they will help them learn through their domain expertise, teaching, and technological mastery.

Further references

Bloom, D. (2021, December 14). The Yale-NUS closures: Unanswered questions. *The Yale Globalist*, https://globalist.yale.edu/2021-2022-issues/the-yale-nus-closures-unanswered-questions

Brookfield, S. D. (1987). *Developing critical thinkers*. Jossey-Bass.

Brookfield, S. D. (2012). *Teaching for critical thinking: Tools and techniques to help students question their assumptions*. John Wiley & Sons.

Brookfield, S. D., & Preskill, S. (2012). *Discussion as a way of teaching: Tools and techniques for democratic classrooms*. John Wiley & Sons.

Brookfield, S. D., & Preskill, S. (2016). *The discussion book: 50 great ways to get people talking*. John Wiley & Sons.

Brookfield, S. D., Rudolph, J., & Yeo, E. (2019). The power of critical thinking in learning and teaching. An interview with Professor Stephen D. Brookfield. *Journal of Applied Learning and Teaching, 2*(2), 76-90, https://doi.org/10.37074/jalt.2019.2.2.11

Brookfield, S. D., Rudolph, J., & Tan, S. (2022). Powerful teaching, the paradox of empowerment and the powers of Foucault. An interview with Professor Stephen Brookfield. *Journal of Applied Learning and Teaching, 5*(1), 131-145, https://doi.org/10.37074/jalt.2022.5.12

Friedman, T. L. (2005). *The world is flat: A brief history of the twenty-first century*. Farrar, Straus and Giroux.

Gudykunst, W. B., & Ting-Toomey, S. (1988). *Culture and interpersonal communication*. Sage Publications.

Mitnick, K. D., & Simon, W. L. (2002). *The art of deception: Controlling the human element of security*. Wiley.

Piketty, T. (2014). *Capital in the twenty-first century*. Belknap Press.

Preskill, S., & Brookfield, S. D. (2009). *Learning as a way of leading: Lessons from the struggle for social justice*. John Wiley & Sons.

Preskill, S., Rudolph, J., & Tan, S. (2023). Learning leadership personified. An interview with Professor Stephen Preskill. *Journal of Applied Learning and Teaching, 6*(1), 324-336, https://doi.org/10.37074/jalt.2023.6.1.12

Rudolph, J. (2014). *Massive Open Online Courses (MOOCs) as a disruptive innovation in higher education?* Master of Education dissertation, University of Adelaide, https://www.academia.edu/15137335/Massive_Open_Online_Courses_MOOCs_as_a_Disruptive_Innovation_in_Higher_Education

Rudolph, J., Tan, S., & Aspland, T. (2021). Editorial 4(2): Black swan or grey rhino? Reflections on the macro-environment of higher education during the pandemic. *Journal of Applied Learning and Teaching, 4*(2), 6-12, https://doi.org/10.37074/jalt.2021.4.2.1

Rudolph, J., Tan, S., & Tan, S. (2023a). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching, 6*(1), 364-389. https://doi.org/10.37074/jalt.2023.6.1.23

Sirota, D. (2006, July 31). Billionaire scion Tom Friedman. *HuffPost,* https://www.huffpost.com/entry/billionaire-scion-tom-fri_b_26164

Sunstein, C. R. (2014). *Why nudge?: The politics of libertarian paternalism*. Yale University Press.

Susskind, D. (2021). *A world without work*. Penguin.

Sword, H. (2012). *Stylish academic writing*. Harvard University Press.

Tan, S. (2023a). Exploiting disruptive innovation in learning and teaching. In K. Rajaram (Ed.), *Learning intelligence: Innovative and digital transformative learning strategies: Cultural and social engineering perspectives* (pp. 149-176). Springer Nature Singapore.

Tan, S. (2023b). Harnessing immersive technologies for innovation in teaching and learnings. In K. Rajaram (Ed.), *Learning intelligence: Innovative and digital transformative learning strategies. Cultural and social engineering perspectives* (pp. 305-334). Springer Nature Singapore.

Tan, S. (2023c). Harnessing artificial intelligence for innovation in education. In *Learning intelligence: Innovative and digital transformative learning strategies. Cultural and social engineering perspectives* (pp. 335-363). Springer Nature Singapore.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.