Vol.7 No.2 (2024)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

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

Validation and cut score's estimation for the online learning scale using Item Response Theory

Nuttaporn Lawthong[A]

A Associate Professor, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

Purin Thepsathit[B]

B Research Assistant, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

Keywords

Cut score;
higher education;
online learning scale;
psychometric properties;
student-centered learning.

Correspondence

nuttaporn.l@chula.ac.th [A]

Article Info

Received 20 February 2024
Received in revised form 15 August 2024
Accepted 28 August 2024
Available online 2 September 2024

DOI: https://doi.org/10.37074/jalt.2024.7.2.28

Abstract

This study aims to validate the Online Learning Scale (OLS) and estimate the OLS's cut score. The sample was 1,459 undergraduate students across different universities and field studies in Thailand. The OLS was newly developed by synthesizing related literature, focus groups, and observations and was designed to measure online learning quality from student-centered perspectives. The 24 OLS items were validated using Classical Test Theory (CTT) and Item Response Theory (IRT). The OLS' cut score was estimated by analyzing the scale's lowest points of the test information function. The effectiveness of OLS' cut score was investigated by dividing groups of samples by cut score and analyzing their difference using Chi-square test and ANOVA with other variables. The results indicated that the OLS showed appropriate psychometric properties: 1) The Cronbach's alpha reliability was .94 and discrimination indices were .62 to .79, 2) CFA fit indices indicated good fit (CFI = .953, RMSEA = .053), and 3) The IRT difficulty parameters were arranged in order and in the range of -3.0 to 3.0, while the discrimination parameters exceeded 0.65 in all items. The estimated two cut scores (29 and 48 points) effectively distinguished the online learning quality delivered to the students.

Introduction

The spread of the COVID-19 pandemic has brought about sudden and unprecedented changes to all levels of education. One of the most prominent changes is the adoption of full-scale online or remote learning, replacing face-to-face learning. Even though online learning is not entirely a novel concept, it was only employed as additional learning support outside the classroom context or a tool to develop additional knowledge and skills before the arrival of COVID-19 (Firmansyah et al., 2021; Pal & Vanijja, 2020; Zemsky, 2014). Therefore, the sudden and full-flung changes to online learning affected teachers regarding appropriate learning management, various learning strategies, learning resources, and personal well-being (Mulrooney & Kelly, 2020). For this reason, learners faced unreadiness in many areas, including learning strategies that were different from face-to-face learning in which different teachers had different ways of maneuvering learning methods and adapting themselves to the nature of online learning (Doucette et al., 2021; Pal & Vanijja, 2020). This idea is consistent with the authors' observation of our own classes in the major at the university, which demonstrated that learners mostly learned from lectures since teachers were not well prepared to teach online and lacked the technical skills necessary for learning. Teachers could not interact as they did in a regular classroom due to the lack of body language communication and immediate interactions, which caused difficulties. Also, students sometimes accidentally interrupted teachers' teaching, and teachers sometimes needed to impose rules for students to keep their cameras and microphones on during class time. Students' names were called to enhance class engagement.

Even though instruction was quite problematic, Thai universities must still conform to the educational standards and correspond to the teaching and learning imposed by the ASEAN University Network Quality Assurance when conducting classes. The mentioned educational quality control emphasized learners as the center in which learning activities should provide opportunities for learners to engage in the learning process, reflecting the impacts of learning engagement and student participation on the ability to seek knowledge and ways of self-learning independently (ASEAN University Network, 2020). From the global point of view, the study conducted by Miranda et al. (2021) revealed that tertiary education is currently in Education 4.0, which supports heutagogy, peeragogy, and cybergogy. Students are active learners and can design their learning styles and self-learning, whereas teachers perform the roles of mentors or coaches, collaborators, and points of reference. Therefore, the learning design in higher education should be based on the student-centered approach. The intended outcomes are both soft and hard skills via the assistance of technological tools, learning platforms, and online learning sources.

Although the COVID-19 situation has been increasingly improved and normalized, many universities around the world still adopt online learning, which has been proven to be an effective major learning platform, and there is a tendency to expand in the future. From learners' perspectives, some learners learned effectively online, thanks to the considerable convenience, and it was especially beneficial to students from deprived families. In addition, it increased diversity in class (Dos Santos, 2022; Doucette et al., 2021; Jones & Sharma, 2020). Online learning is advantageous for self-directed learners because it contains a massive pool of available online resources for seeking additional knowledge and working with others on online platforms (Geng et al., 2019). Besides, students can revise lessons via recorded videos of online meetings. This also facilitates self-directed and student-centered learning, leading to life-long learning (Karatas et al., 2021; Chuwuedo et al., 2021; Sallah et al., 2019). Therefore, to summarize, there is a possibility that online learning will persist in higher education because the outcomes of all research bodies seem to go in the same direction, which indicates that online learning could be a major learning platform with abundant benefits to students.

Table 1. Existing scales related to online learning.

Author	Variable	Instrument	Sample	Validation
Artino and McCoach (2008)	perceived task value and self-efficacy for online learning	**Online learning value and self-efficacy scale** - 11 items, 7-point Likert-type scale *("completely disagree"* to *"completely agree")* - 2 subscales: 1) task value, and 2) self-efficacy for learning with self-paced, online training	204 personnel from the U.S. Navy (in study 1) and 1,127 undergraduate U.S. Naval Academy students	- validity based on test content (experts) - validity based on internal structure (EFA, CFA) - validity based on relations with other variables (correlations, regression) - Cronbach's alpha reliability
Barnard et al. (2009)	self-regulation in online learning	**Online self-regulated learning questionnaire (OSLQ)** - 24 items, 5-point Likert-type scale *("strongly disagree"* to *"strongly agree")* - 6 subscales: 1) environment structuring, 2) goal setting, 3) time management, 4) help seeking, 5) task strategies, and 6) self-evaluation	434 students enrolled in blended or hybrid class, and 204 students enrolled in online courses	- validity based on internal structure (EFA, CFA) - validity based on relations with other variables (CFA across two type of sample) - Cronbach's alpha reliability
Barrot et al. (2021)	challenges in online learning	**Online learning challenges rating scale** - 37 items, 6-point Likert-type scale ("*not at all/negligible"* to "*to a very great extent"*) - 7 subscales: 1) self-regulation challenges, 2) technological literacy and competency challenges, 3) student isolation challenges, 4) technological sufficiency challenges, 5) technological complexity challenges, 6) learning resource challenges, and 7) learning environment challenges	200 university students	- Cronbach's alpha reliability
- Bolliger and Martindale (2004) - Bolliger and Halupa (2012)	online learning satisfaction	**Student satisfaction questionnaire in online environments** - 24 items, 5-point Likert-type scale *("strongly disagree"* to *"strongly agree")* - 5 subscales: 1) instructor, 2) technology, 3) course setup, 4) interaction, 5) outcomes, and 6) overall satisfaction.	- 303 graduate students - 42 doctoral students	- validity based on internal structure (CFA) - Cronbach's alpha reliability
Cheon et al. (2021)	readiness for online learning	**Online Learning Readiness Self-Check (OLRSC)** - 23 items, 7-point Likert-type scale *("strongly disagree"* to *"strongly agree")* - 6 subscales: 1) learning management, 2) interaction with peers, 3) technology management, 4) space management, 5) interaction with instructors, and 6) motivation management	505 online learners with diverse background	- validity based on internal structure (EFA, CFA)
Deshwal et al. (2017)	online learner experience	**Online learning experience scale** - 20 items - 4 subscales: 1) pragmatic-pleasurable experience, 2) use-social experience, 3) hedonistic-exhaustive experience, and 4) sociability experience	150 students who enrolled in online learning platforms and online academic content from universities	- validity based on internal structure (EFA) - reliability - validity based on relations with other variables (regression)
Hung et al. (2010)	readiness for online learning	**Online Learning Readiness Scale (OLRS)** - 18 items, 5-point Likert-type scale *("strongly disagree"* to *"strongly agree")* - 5 subscales: 1) computer/internet self-efficacy, 2) self-directed learning, 3) learner control, 4) motivation for learning, and 5) online communication self-efficacy	1051 undergraduate students with different majors	- validity based on internal structure (CFA) - composite reliability - validity based on relations with other variables (convergent and discriminant validity)
Kaufmann et al. (2016)	perceptions about classroom climate	**Online learning climate scale (OLCS)** - 15 items, 7-point Likert-type scale *("strongly disagree"* to *"strongly agree")* - 4 subscales: 1) instructor behaviors, 2) course structure, 3) student connectedness, and 4) course clarity	263 undergraduate students	- validity based on test content (experts) - validity based on internal structure (EFA) - Cronbach's alpha reliability - validity based on relations with other variables (convergent validity)
Sun and Rogers (2021)	self-efficacy in online learning	**Online learning self-efficacy scale (OLSS)** - 31 items, 6-point Likert-type scale *("strongly disagree"* to *"strongly agree")* - 4 subscales: 1) technology use self-efficacy, 2) online learning task self- efficacy, 3) instructor and peer interaction and communication self- efficacy, and 4) self-regulation and motivation efficacy	366 undergraduate and graduate students	- validity based on test content (experts) - validity based on internal structure (EFA, CFA) - Cronbach's alpha reliability
Tsai et al. (2020)	self-efficacy in online learning	**Self-efficacy questionnaire for online learning (SeQoL)** - 25 items, 11-point Likert-type scale *("cannot do at all"* to *"highly confident can do")* - 5 subscales: 1) self-efficacy to complete an online course, 2) self-efficacy to interact socially with classmates, 3) self-efficacy to handle tools in a CMS, 4) self-efficacy to interact with instructors in an online course, and 5) self-efficacy to interact with classmates for academic purposes	406 undergraduate and graduate students	- validity based on internal structure (CFA, Latent profile analysis) - validity based on relations with other variables (concurrent and convergent validity) - reliability
Tzafilkou et al. (2022)	students' digital competence	**Students' digital competence scale (SDiCoS)** - 28 items, 5-point Likert-type scale ("*strongly disagree"* to *"strongly agree")* - 6 subscales: 1) search, find, access, 2) develop, apply, modify, 3) communicate, collaborate, share, 4) store, manage, delete, 5) evaluate, and (6) protect	156 undergraduate and graduate students	- validity based on internal structure (CFA) - validity based on relation with other variables (discriminant validity)
Wei and Chou (2020)	online learning perception	**Online learning perception scale (OLPS)** - 23 items, 5-point Likert-type scale ("*strongly disagree"* *to "strongly agree")* - 5 subscales: 1) accessibility, 2) interactivity, 3) adapt-ability, 4) knowledge acquisition, and 5) ease of loading.	356 undergraduate students	- validity based on internal structure (EFA) - Cronbach's alpha reliability
Yeşilyurt (2021)	readiness for online learning	**Readiness for online learning scale (ROLS)** - 9 items, 5-point scale - 2 subscales: 1) computer literacy and 2) computer-based self-confidence	306 undergraduate students	- validity based on internal structure (EFA, CFA) - validity based on relation with other variables (t-test, ANOVA)
Yilmaz and Karataş (2018)	online learning interaction	**Perceptions of online interaction scale** - 30 items, 5-point Likert-type scale - 3 subscales: 1) learner-content interaction, 2) learner-instructor interaction and 3) learner-learner interaction	177 undergraduate students	- validity based on test content (experts) - validity based on internal structure (EFA, CFA) - Cronbach's alpha reliability

Even though scales related to online learning have been developed, there is still a scarcity of measurement scales for online learning quality. The scales related to online learning that the researchers found are provided in Table 1. Table 1 revealed that there is no specific study aimed at developing a scale intended to measure the quality of online learning that aligns with student-centered learning. The existing scales focused on readiness, self-efficacy, self-regulation, challenges, satisfaction, classroom climate, experience, digital competence, perception, and interaction. Moreover, the validation of the psychometric properties of the existing scales still relied on the Classical Test Theory (CTT) by analyzing the internal consistency reliability, the validity based on internal structure using exploratory and confirmatory factor analysis, and validity based on other variables using the correlation coefficient. The limitation of CTT is that it is a sample-dependent measurement model that lacks invariance regarding psychometric properties when applied to other research samples (Hambleton & Jones, 1993).

Moreover, there is a research gap in estimating a cut score for interpreting outcomes of such online learning scales. The online learning scale should establish two or more cut scores, dividing the score range to partition the distribution of scale scores into several categories. These categories are used for descriptive purposes or to distinguish among students who received different quality online learning programs (American Educational Research Association [AERA] et al., 2014). Information acquired from the developing scale in this study will help universities design higher levels of quality in online courses, enhance the standard of online learning to respond to students' needs, improve students' learning outcomes, and provide validity for interpreting the measurement results.

Literature review

Student-centered online learning

During the initial period of effective online learning, academics focused on strategies that effectively deliver learning material; later, they realized there should be a specific online learning instruction, bringing about changes that turned the attention from delivering learning material to studying online learning instructional methods, as well as from teacher-centered to student-centered ways of learning (Ehlers, 2009; Junus et al., 2015). Student-centered learning refers to the type of learning where students are active learners who play a part in the learning design and initiators of knowledge via activities and group work. In contrast, teachers are facilitators who create a comfortable learning atmosphere, so students feel free to provide comments. By so doing, teachers provide the necessary guidance, feedback, motivation, and formative assessment to give rise to a meaningful learning process, which is in contrast to teacher-centered learning in which teachers are key people responsible for students' learning process by designing learning and communicating knowledge, while learners are passive learners (Froyd & Simpson, 2008; Keiler, 2018; Mascolo; 2009; Overby, 2011; Wright, 2011). To accomplish student-centered learning in higher education (ASEAN University Network, 2020; Miranda et al., 2021), learners must seek knowledge, design learning strategies, and learn independently. They also need to participate in various activities cooperatively and collaboratively. There are online sources to assist their learning, and teachers are supporters who can relate existing knowledge to new knowledge to make lessons more intriguing for students to maintain their attention (Pornkul, 2009). Online learning encourages the development of student-centered learning, allowing students to set a goal and work on learning to achieve it as self-directed learners (Teo & Divakar, 2021).

Even though online learning can take the form of synchronous and asynchronous learning, it is observable that during online learning in Thai universities during and after the COVID-19 pandemic, teachers have employed teacher-centered learning via lectures on online programs and relied on attendance checking and tests as an assessment of learning outcomes. However, after teachers became more accustomed to online learning platforms, they adopted assistive features such as breakout rooms, polls, chats, whiteboards, and screen sharing in their teaching. Besides, there have been learning materials like video clips or learning slides to allow students to preview and review after classes. There have also been adjustments in learning materials such as books that could be scanned to access the file, which is convenient for students to study using different gadgets. Teaching videos are sometimes uploaded to cloud storage for the sake of revision. Assessments have been remodified through essay writing, open-book exams, or performance assessments via recorded online presentations. However, for specific fields that depend on practice and performance, like medicine, engineering, or fine arts, students were required to come to the university as exceptions during the lockdown period, under strict preventive measures such as wearing masks or attending classes in small groups. This goes in the same direction as a study by Marinoni et al. (2020) that asserted that online learning cannot completely substitute in-class learning for skills requiring practice and performance, such as science or arts. Moreover, some teachers invited guest speakers to give specific talks to compensate for field trips during lockdown.

It is noticeable that blended learning in colleges after the COVID-19 pandemic still plays a significant role even though on-site learning has come back, particularly in subjects that rely on lectures and special occasions such as a student's absence. Teachers would be asked to open the online meeting program and record the meeting so that students can review the lessons and catch up with their friends. At the graduate level, online education is the primary learning channel, and they only come to the university to take exams with other students. It is observable by teachers that online learning opens opportunities for students who reside or work in distant locations to obtain graduate education, which corresponds to the study conducted by Pal and Vanijja (2020) that points out the benefits of online meeting applications in both synchronous learning via the live feature, and asynchronous learning via recorded videos and shared files.

The relevant literature and observations suggest that after teachers and students have become familiar with online learning, there has been a shift from teacher-centered to student-centered, consistent with a study by Doucette et al. (2021). They found that the advantage of online learning is that it allows students to review lessons via video clips. The breakout rooms can be used to stimulate discussions and collaborative learning. Also, online learning has back channels such as chat or private chat that facilitate communication with teachers and among peers, which is even better than in-class learning.

Test theory

The model used to measure the scale's psychometric properties was based on Classical Test Theory (CTT) and Item Response Theory (IRT). CTT was analyzed using the sum score under the concept that actual abilities could be observed from the sum score and errors. IRT was analyzed using latent variables under the concept that latent variables of test takers' abilities could be measured from questions that were made on the same scale (Edelen & Reeve, 2007). Additionally, the model based on CTT is a sample-dependent model in which psychometric properties of the measurement vary according to the research sample used to analyze. This contrasts with the concept of IRT, which is a sample-independent model. If the empirical data fit with the model, the psychometric properties remain invariant across all samples (Hambleton & Jones, 1993). That is not all; when considering the measurement precision, it is discovered that the IRT model estimates the measurement precision using the item information function that estimates information based on different ability levels. In contrast, the CTT model adopts the same standard error of measurement across all ability levels, possibly generating bias in extremely high, extremely low, and moderate samples' abilities (Jabrayilov et al., 2016). Thus, the current study employed the IRT model to analyze the psychometric properties.

The estimation of the cut score using IRT

A cut or cutoff score refers to the score point used as a criterion to separate students into different groups. There are two concepts of setting a cut score, which are the test-centered and the student-centered. Aside from dividing students, a cut score is used in the outcome reports to obtain feedback necessary for improving students at diverse levels, affecting stakeholders' decisions (Tiyawongsuwan, 2011; Ziesky et al., 2008). There are various ways to perform both concepts of a cut score. This current OLS development studied a cut score of the developed scale to allow teachers or individuals who deal with online learning to use the feedback acquired from the scale in the adjustment or quality enhancement of online learning in tertiary education. Because OLS is a newly developed measurement, the test-centered cut score in which experts decide and determine the minimum passing criteria may not be an appropriate method. Therefore, the student-centered cut score was adopted using the contrasting-group method (Lin, 1989; Ziesky et al., 2008), which determines the cut score from the curve of the distribution of students' scores using the test informative function obtained from the analysis based on the IRT. This also helps eliminate the limitations in the contrasting-group method, in which the score distribution possibly showed more than one cut score followed by more than one criterion. The test information function converts students' scores into a standardized value, making it feasible to determine the cut score.

However, according to Zumbo (2016), the difference between cut-score setting and standard-setting definitions is discussed. His study reveals that the definition of the cut score may refer to an arbitrariness of setting the cut score to determine whether to pass or fail. In contrast, the definition of standard-setting refers to the process of finding the cut-score point of a tool that meaningfully categorizes respondents into two or more groups. The cut score should indicate the minimum level of performance to pass each group. Therefore, the OLS study on estimating cut score was focused on determining the standard-setting of OLS.

There has been a controversy among experts regarding the use of test information in determining the cut score. Some claim that cut scores should be at a high information value to determine the cut score accurately and reliably. However, a study by Wyse and Babcock (2016) suggested that the high information value that could divide students with accuracy and reliability should be far from the cut scores. The cut score should be estimated at the point of low information value (Wyse, 2017). This notion aligns with the interpretation of the test information graph, which is a high information value that refers to the fact that the measurement can divide abilities with a high level of precision (Edelen & Reeve, 2007). In other words, the measurement can determine the point of the ability at the high information score with precision. Hence, the current research adopted the lowest point guideline of the test information graph to determine the cut scores. If the graph rose and plummeted before peaking again, the lowest point was determined as the cut score. However, when adopting the student-centered cut score method, we should ensure that the samples used show enough differences in the quality of online learning to ensure the precise estimation of the cut score. Therefore, the CTT was employed using the discrimination index to investigate differences in the samples' responses. If the value was high, it could be interpreted that there were heterogeneous online learning qualities (Hambleton & Jones, 1993).

From the study of relevant literature, it is crucial to develop the online learning scale (OLS) that measures the quality of online learning, which necessarily corresponds to a learner-centered approach and current online learning contexts, to obtain information that can be utilized to improve learning instruction. The OLS is a newly developed scale and should provide evidence of psychometric properties in reliability and validity analyzed by CTT and IRT to ensure its quality. The cut score estimation will also be conducted to interpret the measurement results and use them to improve online learning classrooms. The scale will be up-to-date and facilitative to the current online learning at the tertiary level, which has the potential to expand continuously. Therefore, the research questions were: 1) what are the psychometric properties of the OLS measurement? 2) how many groups should there be in the quality estimation of online learning?

3) what should be the value of cut scores for the OLS?

Research objectives

1) To validate the Online Learning Scale (OLS) for undergraduate students.

2) To estimate the OLS's cut score to distinguish online learning quality for blended and online learning in higher education.

Method

This research adopted a research and development (R&D) design (Gall et al., 2007; Gustiani, 2019), which is widely used to develop an educational assessment tool or product. The process of OLS development started with studying related literature, focus groups, and observation to collect information for designing OLS, then developing a preliminary OLS. The OLS underwent a pilot study and revision before being tested with the research sample to reveal its psychometric properties and validate its finalized version. Lastly, OLS's cut score was estimated to interpret OLS implementation effectively.

Sample

1,459 undergraduate students in Thailand were research samples. This study used stratified random sampling to cover samples' distribution in five types of universities, including public universities, Rajamangala University of Technology, Rajabhat universities, private universities, and community colleges, and four types of study fields, including health sciences, sciences and technology, humanities, and social sciences. Most students came from public universities, accounting for 70.53%, and 49.55% came from social sciences majors (the summary table of university types and study fields is provided in the appendix). The majority gender of samples was female (72.72%), and the average age of the samples was 20.84 (*SD* = 2.33, *Min* = 17, *Max* = 43).

Instrument

The instrument adopted in the current research is the Online Learning Scale (OLS). OLS was developed and synthesized from a literature review, a focus group of experts, and online classroom observation in a Thai university. This process aimed to develop an operational definition of online learner-centered learning that aligns with the current learning instruction. The operational definition of the current study is the following:

> "Online Student-Centered Learning means students are mainly responsible for their learning, in which learners participate in the decision-making process to seek the most optimal way to enhance their learning and skills via analyzing problems and contributing factors of a specified incident. They are required to seek knowledge, participate in both online and offline activities, work in a team, perform their duties, exchange opinions with peers and teachers, plan, execute the plan, and summarize their learning outcomes. Ultimately, they will have to present their learning outcomes to their teachers, peers, and the public via different media by producing a multimedia product suitable for online presentations."

To implement online student-centered learning, teachers' roles should include incorporating the existing knowledge into new knowledge and turning it into a lesson using discussion questions or applications to stimulate learners. Teachers support and facilitate group work, projects, and research studies that students are intrigued about. They also need to relocate time to provide advice and feedback and answer learners' questions outside the class time. They can invite guest speakers to compensate for some fieldwork. It is crucial to ensure that learners are clearly informed about the submission platforms and deadlines.

Online student-centered learning should involve both synchronous learning in a live environment and asynchronous learning, which relies on various applications and video recordings of the in-class teaching so that students can review lessons. Content files, activities, or video clips should be uploaded if they are interested in additional information. Online learning that adopts such meeting applications as Zoom, Microsoft Team, and Google Meet allows users to use breakout rooms to enable small discussions to exchange opinions or do other activities. Screen sharing allows users to give presentations, and students can learn from their peers. Also, students can answer questions by unmuting their microphones or sending messages on the chat box. Some learning content that heavily depends on practices should be taught on-site to enable actual practices, but they can integrate online learning into it.

After the operational definition was thoroughly coined, an item pool of 25 questions was generated from the relevant literature aligning with the operational definition. The OLS was developed using the 4-point Likert-type scale (never, rarely, sometimes, and often). This was adapted from *Frequency – five points* (never, rarely, sometimes, often, and always) in Likert-type scale response anchors of Vagias (2006), but the highest response (always) was eliminated because each content in OLS cannot always be performed in the classroom, it must be combinedly used according to learning activity, subject, or topic. The OLS items were analyzed for content validity by experts, and a pilot study with a sample consisting of 61 students was conducted to test its psychometric properties regarding reliability and the discrimination index before putting it into actual samples.

The analysis results of the psychometric properties in the pilot study revealed that Item 8 (*learners spend most of their time watching lectures in online learning*), a reverse-scale item, had a lower discrimination index of less than .30. Therefore, it had to be omitted before the scale was implemented with actual samples. The omission would not affect the measurement content because the content representation of this item was displayed in other items remaining in the OLS. In summary, there were 24 items of OLS in total, as shown in the appendices.

Data collection and data analysis

The OLS data was gathered using Google Forms. The link of OLS was sent electronically to multiple universities, determined by the types of universities and study fields (majors) mentioned in the sample section. The Institutional Review Board (IRB) approval of the ethical code of the research instrument was SWUEC 279/64s based on the Declaration of Helsinki, the Belmont Report and CIOMS guidelines. To analyze the OLS's psychometric properties, the items were analyzed in terms of the internal consistency reliability using Cronbach's alpha coefficient based on the CTT, test information, the expected a posteriori (EAP) reliability, the discrimination index, and the discrimination parameter (***a***) and difficulty parameter (***b***) based on the IRT.

The validity of the instrument was presented in three types as follows. The first type was content validity, in which relevance and representation between the items and objective to be measured, which were core components of content validity of the measurement, currently recognized as evidence based on test content (AERA et al., 2014), evaluated by eight experts. The second type was construct validity, currently known as evidence based on internal structure. Exploratory factor analysis and confirmatory factor analysis were employed. Lastly, criterion-related validity, currently known as evidence based on relations to other variables (AERA et al., 2014), was analyzed by correlation coefficient with other variables. The four variables were adapted from existing scales, including student-teacher engagement scale of Klincumhom (2013) (11 items, r = .80 - .90), online learning characteristics scale of North Carolina Community College System (n.d.) and Penn State University (n.d.) (12 items, r = .95) (*reliability was analyzed in this study because the original value was not provided*), self-directed learning scale of Yang et al. (2020) (5 items, r = .96), and self-assessment scale of Yan (2018) (10 items, r = .79 - .90). The four scales were sent to the sample together with the OLS. These four scales are provided in the appendix.

To estimate the effectiveness of the cut score, the Chi-square test was employed to test differences between students according to the cut scores, university types and majors, and online learning outcomes, using ANOVA to test differences between students and the psychological variables of online learning. The software for data analysis was R with multiple packages. First, the *psych* package (Revelle, 2024) was used to analyze Cronbach's alpha coefficient and the discrimination index in the CTT. Second, the mirt package (Chalmers, 2023) was used to analyze the psychometric properties based on the IRT. Third, the *lavaan* package (Rosseel et al., 2023) was used to analyze factor analysis.

Result

Validation of OLS psychometric properties

Evidence based on test content

To investigate the content validity of the OLS, eight experts were consulted to assess the relevance and representation between OLS items and the operational definition. These experts were selected from various academic fields, including health sciences, sciences and technology, humanities, social sciences, and instructional media and technology specialists. Additionally, all of them had prior experience in online teaching. To evaluate the items, the eight experts assigned three points to items they found to be relevant and well-represented, two points to items where they were uncertain about their relevance and representation, and one point to items lacking relevance and representation.

Subsequently, the researchers collected the experts' ratings and calculated the average score by dividing the total points earned by the number of experts. Items with scores exceeding 2.5 were considered to exhibit relevance and representation. Following this assessment, the researcher made necessary improvements and adjustments to the items based on the feedback provided by the experts before implementing the OLS. The scores from eight experts were between 2.50 – 3.00, which indicated that the items were relevant and represented the quality of online learning.
Evidence based on internal structure

Evidence based on internal structure was analyzed using exploratory factor analysis (EFA) to investigate the number of factors of the OLS and confirmatory factor analysis (CFA) to confirm the EFA result. To ensure the measurement quality, apart from designing random sampling to cover the student population, the researchers inspected the discrimination index of the OLS using corrected item-total correlation to ensure diversity of the samples' responses under the criterion of at least .30 (Ferketich, 1991). The analysis revealed that all 24 items in OLS obtained discrimination index values from .62 to .79, as illustrated in Table 2. Additionally, the internal consistency reliability of the OLS was analyzed using Cronbach's alpha coefficient to consider eliminating some items in the complete scale under the criterion of at least .70 (George & Mallery, 2003, as cited in Gliem & Gliem, 2003). It was discovered that the OLS obtained a reliability value of .96.

The results of EFA showed that the OLS possessed one factor, determined by the high factor loading value loaded in factor 1, as shown in Table 2. Later, to confirm one trait of OLS, the results of CFA revealed that the Chi-Square value was 1274.71 (df = 247), and the p-value was .000, reflecting the misfit between the model and the empirical data. Nevertheless, because of the sensitivity to a large sample size of the Chi-square test, other fit indicators were taken into consideration (Weston & Gore, 2006). It was discovered that the CFI value was .953, which exceeded the basic criterion of .90. The GFI was .922, which exceeded the criterion of .90. The AGFI value was .905, which was again over .90. The RMSEA value was .053, which was lower than .06. The RMR value was .02 which was lower than .05. This supported the conclusion that the model fit with the empirical data. The CFA showed that the factor loading values of all 24 items were positive, ranging from .614 to .788, with a statistical significance of .05 in all items, as shown in Table 2 and Figure 1.

Validating OLS's psychometric properties

The analysis of the psychometric properties of the measurement based on the IRT consists of the analysis of the threshold (difficulty) parameter, the discrimination parameter, test information, item information, and the Expected a Posteriori (EAP) reliability. To elaborate, the threshold parameter refers to the difficulty of the cut score in each score range of the items. There were four score ranges in the OLS; therefore, there were three thresholds (b_1,b_2,b_3). The threshold parameter values should be in order (monotonicity) and should be in the range of -3.0 to 3.0, whereas the discrimination parameter value (*a*) should be at least 0.65 (Baker, 2001; Bichi & Talib, 2018). These were the criteria used to qualify the items.

To analyze model fit, the Partial Credit Model (PCM) and the Graded Response Model (GRM) were analyzed based on the AIC, BIC, log2 likelihood, whose value was supposed to be lower, and the Chi-Square test. If the p-value had a statistical significance, it meant GRM had a better fit than PCM. The analysis results revealed that AIC, BIC, and log2 likelihood values of GRM (AIC = 60222.76, BIC = 60730.17, and log2 likelihood = -30015.38) were lower than those in PCM (AIC = 61766.34, BIC = 62152.18, and log2 likelihood = -30810.17), while the p-value had a statistical significance (p-value < .000), meaning the GRM showed a higher level of fit than that in PCM. For this reason, GRM was adopted to analyze the psychometric properties of the OLS, as shown in Table 2.

The analysis results of the threshold parameter and the discrimination parameter of the OLS revealed that the threshold parameter values ranged from -2.595 to 1.533. It was discovered that all items exceeded the basic criterion scores of -3.0 to 3.0 and followed the threshold arrangement (as shown in the item characteristic curve in Figure 2). The discrimination parameter values ranged from 1.752 to 3.107, which exceeded the criteria. The analysis of the item information revealed that all items were distributed according to different abilities (θ). Some items indicated a high level of item information, such as Items 1, 6, 8, 20, and 22, as shown in the item information function in Figure 2.

The analysis of the test information function of the OLS demonstrated that the highest information values in the ability (θ) range of -3.0 to 1.5, meaning the measurement was suitable for estimating all ranges from low to high ranges of online learning quality (θ) from -4 to 3 and the standard errors were in the same ranges. After considering the lower standard errors (SE) (red dashes) and the peaks, we could divide students into three groups. The analysis of the EAP reliability showed that the value was .95, which was extremely high. EAP is comparable to reliability in the CTT, which uses the same interpretation methods (Adams, 2005). The test information function and the reliability graph based on the IRT of the OLS are shown in Figure 3.

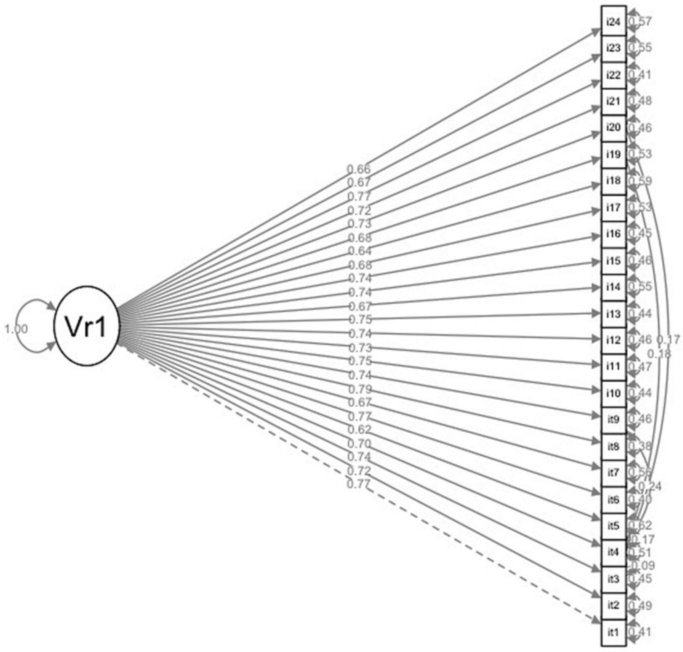

Figure 1. The result of the confirmatory factor analysis of the OLS measurement.

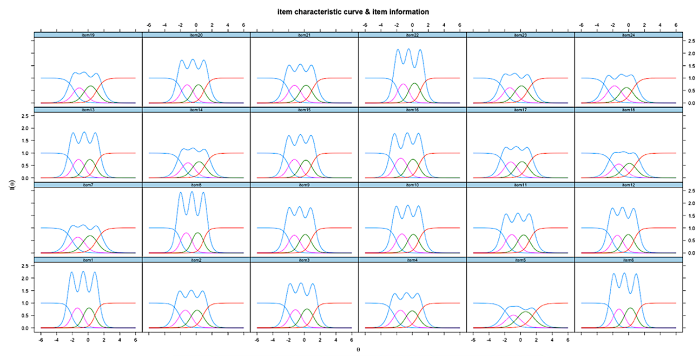

Figure 2. The item characteristic curve and the item information function of the OLS.

Table 2. The summary of psychometric properties' validation of the Online Learning Scale (OLS).

Item	Discrimination Index	EFA			CFA					IRT			
		Factor		Uniqueness	Factor Loading								
		1	2		b	SE	β	Wald	R^2	*a*	b_1	b_2	b_3
1	.77	.771		.379	1.000		.770		.594	2.979	-2.098	-0.627	0.852
2	.72	.717		.486	0.975	.033	.718	29.321*	.515	2.403	-2.121	-0.637	0.847
3	.75	.750		.343	1.006	.033	.744	30.558*	.553	2.640	-1.838	-0.422	1.089
4	.70	.703	.357	.378	0.949	.033	.700	28.438*	.490	2.315	-2.291	-0.759	0.696
5	.62	.621	-.328	.442	0.930	.038	.617	24.617*	.381	1.752	-1.546	-0.256	1.533
6	.77	.770		.405	1.024	.032	.773	32.077*	.598	2.929	-1.887	-0.536	0.973
7	.67	.665		.556	0.925	.034	.666	26.845*	.443	2.050	-2.039	-0.585	1.069
8	.79	.785		.383	1.029	.031	.787	32.802*	.620	3.107	-1.987	-0.549	0.909
9	.75	.751		.340	1.019	.034	.738	30.317*	.545	2.663	-1.934	-0.599	0.822
10	.75	.747		.432	1.011	.033	.747	30.751*	.557	2.719	-2.058	-0.597	0.804
11	.73	.732		.411	1.004	.034	.731	29.980*	.534	2.474	-1.862	-0.344	1.125
12	.73	.736		.412	0.998	.033	.735	30.182*	.540	2.648	-2.093	-0.766	0.663
13	.74	.744		.445	1.017	.033	.746	30.732*	.557	2.665	-1.940	-0.501	0.934
14	.67	.674		.519	1.001	.037	.674	27.233*	.454	2.098	-1.722	-0.382	1.093
15	.74	.735		.446	1.016	.034	.738	30.328*	.545	2.582	-1.994	-0.512	0.888
16	.74	.739		.424	0.968	.032	.739	30.356*	.546	2.651	-2.289	-0.669	0.796
17	.69	.688		.477	0.993	.036	.685	27.756*	.469	2.164	-1.938	-0.540	0.906
18	.65	.647		.512	0.987	.038	.644	25.850*	.415	1.940	-1.892	-0.606	0.778
19	.69	.689		.467	0.993	.036	.684	27.714*	.468	2.142	-1.780	-0.466	1.089
20	.74	.740		.416	1.014	.034	.734	30.117*	.539	2.594	-1.864	-0.440	0.994
21	.72	.724		.427	0.996	.034	.719	29.360*	.516	2.433	-1.970	-0.517	0.941
22	.77	.768		.394	1.029	.032	.769	31.855*	.591	2.908	-1.857	-0.478	1.002
23	.67	.667		.536	0.942	.035	.670	27.056*	.449	2.100	-2.047	-0.683	0.915
24	.65	.658		.475	0.893	.034	.655	26.367*	.429	2.068	-2.595	-0.972	0.435

Note: * means p-value < .05, ** means p-value < .01, *** means p-value < .001.

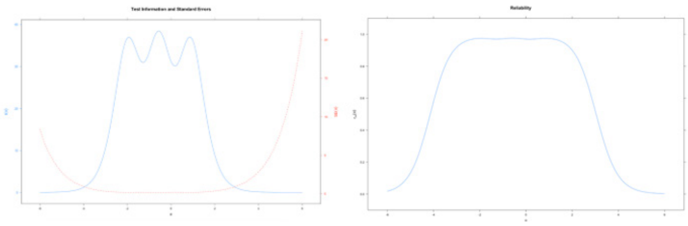

Figure 3. The test information function (left) and the EAP reliability graph (right) of the OLS.

Evidence based on relations to other variables

Evidence based on relations to other variables was established by calculating correlation coefficients between OLS scores and four distinct variables scores: student-teacher engagement (STE), online learning characteristics (OLC), self-directed learning (SDL), and self-assessment (SA). This analysis aimed to demonstrate convergent validity, also known as validity based on relations to other variables. Convergent validity suggests that the developed scale should exhibit a moderate correlation coefficient with these other relevant variables (AERA et al., 2014).

The analysis utilized the Pearson correlation coefficient to examine the relationships between OLS and the abovementioned five variables. The results indicated a statistically significant moderate positive correlation at a significance level of .01. Specifically, OLS demonstrated correlation coefficients of .656 with online learning assessment, .621 with student-teacher engagement, .650 with online learning characteristics, .625 with self-directed learning, and .603 with self-assessment. These findings are summarized in Table 3.

Table 3. The correlation coefficients between OLS scores and five variables.

Variables	OLS	OLA	STE	OLC	SDL	SA
OLS	1.000					
OLA	0.656**	1.000				
STE	0.621**	0.606**	1.000			
OLC	0.581**	0.650**	0.543**	1.000		
SDL	0.544**	0.625**	0.519**	0.918**	1.000	
SA	0.585**	0.672**	0.560**	0.920**	0.884**	1.000

Note: * means p-value < .05, ** means p-value < .01, *** means p-value < .001.

Cut score's estimation

A sum score of the students' responses was collected to estimate the information distribution and the result showed that the sum score of the OLS ranged between 0 and 72, with a mean score of 44.019 (SD = 14.019) and a median of 46 points. The graph showing the distribution of the sum score of the OLS measurement is shown in Figure 4.

The estimation of a cut score relied on the analysis of the OLS information by looking for the lowest point of information after peaking in the information curve that is also the highest in the ability (θ) from -4 to 4. According to Figure 3, the measurement precision was high during the ability (θ) range of -4 to 3, and there were three peaks in the test information function. The two lowest points in between the three peaks were analyzed. They showed that the information graph hit the lowest at the ability (θ) level of -1.3 and 0.2, with the information values of 30.96 and 30.09, respectively. The sum score of students' responses in the OLS reflecting the online learning quality at the locations 1.3 and 0.2, not exceeding ± 0.03, showed that the mean scores of the responses were 22.92 and 48.24, respectively. Thus, the numbers were rounded to determine the cut scores of 23 and 48, respectively. Therefore, there were three levels of interpretation, including learners who needed assistance while learning online (0-22 points), learners receiving standard online learning quality (23-47 points), and learners receiving effective online learning quality (48 points or more). The test information function that estimates the cut scores is shown in Figure 5.

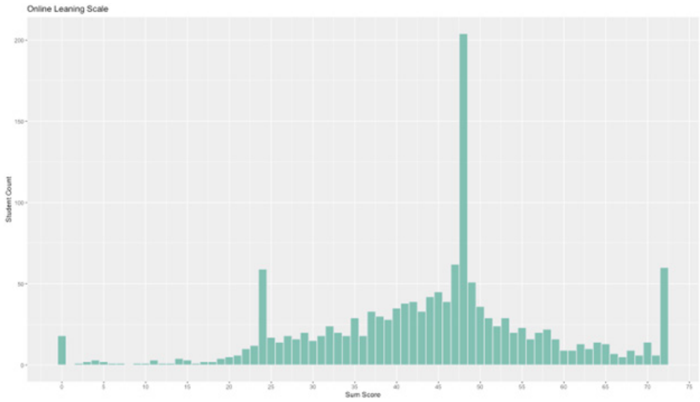

Figure 4. Sum score of the OLS measurement.

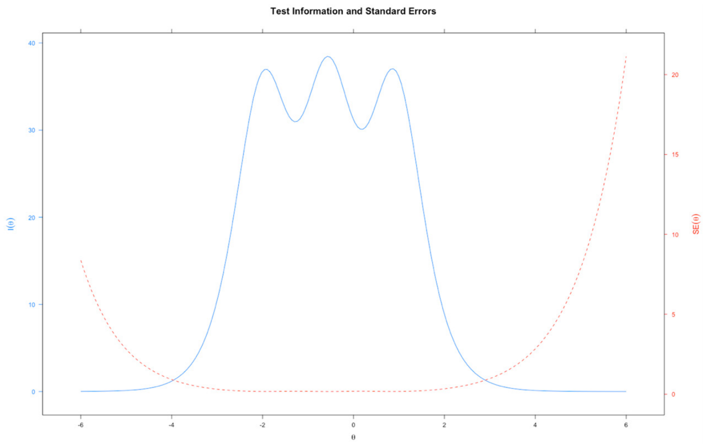

Figure 5. The estimation of the cut scores in the information graph.

Investigation of cut score's effectiveness

After the cut scores were determined, the effectiveness of the cut scores was investigated by dividing students' responses into three groups according to the estimated cut scores and analyzing their basic statistics. It was discovered that Group 1 consisted of 72 people (*M* = 11.514, *SD* = 8.711), while Group 2 was made up of 722 people (*M* = 37.057, *SD* = 7.537), and Group 3 had 665 people (*M* = 55.295, *SD* = 8.195), as shown in Figure 6.

After dividing the responses into three groups, differences in the dependent variables between the samples were based on university types, majors, and the outcomes of online learning responses in each learning outcome. These included the perception of online learning outcomes compared to in-class learning, cognitive, performance, and affective aspects, using the Chi-square test to analyze this portion. The sums of students varied because the answer “Other” from some respondents was omitted from each learning outcome. The Chi-square test results of the three groups of students divided according to universities and majors revealed no difference in the university types. However, there was a difference in majors, as shown in Table 4 and Table 5.

Table 4. The analysis of the Chi-square test classified by universities.

Group	Universities: Public university	Rajamangala University of Technology	Rajabhat universities	private universities	community colleges	**total**
1	51 (5.0%)	10 (4.8%)	5 (3.1%)	4 (11.4%)	1 (5.9%)	71 (4.9%)
2	519 (50.4%)	103 (49.0%)	77 (47.5%)	13 (37.1%)	7 (41.2%)	719 (49.5%)
3	459 (44.6%)	97 (46.2%)	80 (49.4%)	18 (51.4%)	9 (52.9%)	663 (45.6%)
total	1029 (100%)	210 (100%)	162 (100%)	35 (100%)	17 (100%)	1453 (100%)
Pearson Chi-square = 7.035, Cramer’s V = 0.049, p-value = 0.533						

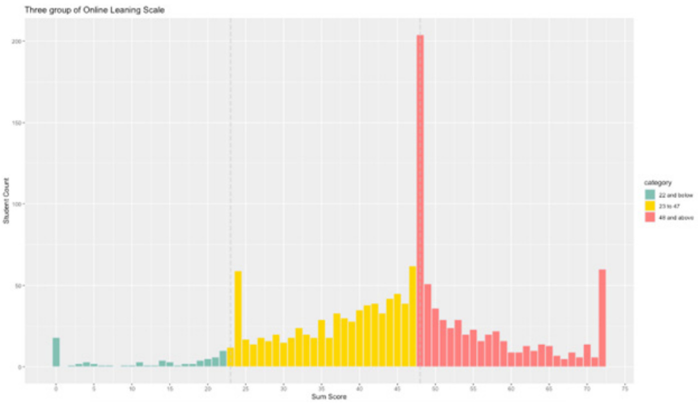

Figure 6. Basic statistics of online learning quality of three groups.

Table 5. The analysis of the Chi-square test classified by majors.

Group	Majors: health sciences	sciences and technology	humanities	social sciences	**total**
1	2 (2.7%)	5 (5.2%)	38 (6.7%)	27 (3.6%)	72 (4.9%)
2	22 (30.1%)	43 (44.3%)	**296 (52.4%)**	**360 (49.8%)**	721 (49.5%)
3	**49 (67.1%)**	**49 (50.5%)**	231 (40.9%)	336 (46.5%)	665 (45.6%)
total	73 (100%)	97 (100%)	565 (100%)	723 (100%)	1458 (100%)
Pearson Chi-square = 24.328, Cramer’s V = 0.091, p-value < .001					

The Chi-square test was used with students’ responses, in which the perception of online learning outcomes in comparison to those of in-class learning was categorized into “better,” “worse,” and “not changed.” The cognitive, performance, and affective aspects could be divided into “more,” “less,” and “not changed.” The result showed that all three groups of students showed differences in all aspects with a statistical significance of .05, as shown in Table 6.

The effectiveness of the cut scores was also analyzed using one-way ANOVA with the four variables in psychometric properties, including student-teacher engagement, online learning characteristics, self-directed learning, and self-assessment. The adopted data was the mean score of all four variables with a score range between 0-3 points. The result revealed that the variance of three groups of students varied in all variables, with a statistical significance of .001, as shown in Table 7.

Table 6. The analysis of the Chi-square test classified by learning outcomes.

Learning Outcomes	Group	better	not changed	worse	Total
Learning outcomes received from online classroom compares to traditional classroom	1	13 (4.0%)	19 (3.5%)	40 (7.1%)	72 (5.0%)
	2	149 (46.0%)	**280 (51.0%)**	**280 (49.6%)**	709 (49.3%)
	3	**162 (50.0%)**	250 (45.5%)	245 (43.4%)	657 (45.7%)
	total	324 (100%)	549 (100%)	565 (100%)	1438 (100%)
	Pearson Chi-Square = 11.150, Cramer’s V = 0.062, p-value = .025				
	Group	more	not changed	less	Total
Cognitive aspect outcomes received from online learning	1	8 (3.9%)	11 (2.3%)	52 (6.9%)	71 (5.0%)
	2	85 (41.1%)	222 (46.9%)	**401 (53.3%)**	708 (49.4%)
	3	**114 (55.1%)**	**240 (50.7%)**	299 (39.8%)	653 (45.6%)
	total	207 (100%)	473 (100%)	752 (100%)	1432 (100%)
	Pearson Chi-Square = 11.150, Cramer’s V = 0.062, p-value < .001				
	Group	more	not changed	less	Total
Performance aspect outcomes received from online learning	1	8 (3.2%)	15 (3.4%)	49 (6.5%)	71 (5.0%)
	2	114 (45.2%)	198 (45.0%)	**402 (53.1%)**	708 (49.3%)
	3	**130 (51.6%)**	**227 (51.6%)**	306 (40.4%)	653 (45.8%)
	total	252 (100%)	440 (100%)	757 (100%)	1432 (100%)
	Pearson Chi-Square = 21.804, Cramer’s V = 0.087, p-value < .001				
	Group	more	not changed	less	Total
Affective aspect outcomes received from online learning	1	7 (2.6%)	32 (4.7%)	33 (6.5%)	72 (4.9%)
	2	108 (40.3%)	**327 (48.3%)**	**284 (55.7%)**	719 (49.4%)
	3	**153 (57.1%)**	318 (47.0%)	193 (37.8%)	664 (45.6%)
	total	268 (100%)	677 (100%)	510 (100%)	1455 (100%)
	Pearson Chi-Square = 28.903, Cramer’s V = 0.100, p-value < .001				

Table 7. The results of ANOVA analysis with four psychological variables.

Variables	ANOVA F	df1	df2	*p*	Levene's test F	df1	df2	*p*
1. Student-teacher engagement	185.902***	2	187.679	<.001	10.900	2	1456	<.001
2. Online learning characteristics	183.981***	2	186.799	<.001	18.169	2	1456	<.001
3. Self-directed learning	151.842***	2	187.091	<.001	20.585	2	1456	<.001
4. Self-assessment	169.273***	2	185.960	<.001	29.616	2	1456	<.001

Note: * means p-value < .05, ** means p-value < .01, *** means p-value < .001.

The analysis of the post hoc test was used to compare each group based on the variables, using the Games-Howell post hoc test because each variable had a diverse level of fluctuation. It was obvious in Levene’s test with the statistical significance. The result demonstrated that Group 1 differed from Groups 2 and 3 with a statistical significance of .001, and Group 2 differed from Group 3 with a statistical significance of .001 in all variables. The analysis result of the post-hoc test is shown in Table 8.

Table 8. The result of post hoc test.

Post hoc test	Group		Group 1	Group 2	Group 3
Student-teacher engagement	1	mean difference	-	0.853 (p < .001)	1.312 (p < .001)
	2	mean difference		-	0.458 (p < .001)
	3	mean difference			-
Online learning characteristics	1	mean difference	-	0.653 (p < .001)	1.137 (p < .001)
	2	mean difference		-	0.484 (p < .001)
	3	mean difference			-
Self-directed learning	1	mean difference	-	0.611 (p < .001)	1.081 (p < .001)
	2	mean difference		-	0.470 (p < .001)
	3	mean difference			-
Self-assessment	1	mean difference	-	0.689 (p < .001)	1.131 (p < .001)
	2	mean difference		-	0.442 (p < .001)
	3	mean difference			-

Discussion

The analysis results of OLS validation revealed agreement between CTT and IRT in the case of the low discrimination index below .3 of the items in the initial scale developed before the pilot study. This reflected that both theories provided an alignment of analysis results. However, in the current study, IRT provided more information about scale development than the use of CTT, meaning that the researchers could analyze the difficulty parameter of each threshold, and the information ensured the monotonicity of the difficulty parameter of each item. This further supports the quality of the OLS. Also, the IRT offered the highest level of information on the students' abilities of each item. The information function could be further utilized to analyze the appropriate cut scores for the respondent-centered method. Therefore, although the results of the psychometric properties analysis based on both theories went in the same direction, the scale developed based on the IRT could offer more beneficial information if the basic requirements of the IRT were met.

The EFA and CFA analysis results were consistent, reflecting internal validity and featuring crucial components of online learning instruction. The researchers also found that the discrimination index and CFA's factor loading were highly related. The five items with the most standard component weights (β) were Item 8 stating that learners participated in thought-provoking activities, giving rise to learning and skill development (β = .787, *SE* = .031), followed by Item 6 stating that learners analyzed problems and contributing factors of a specified incident (β = .773, *SE* = .032), Item 1 stating that teachers integrated existing knowledge into a new lesson (β = .770, *SE* = .032), Item 22 stating that learners reflected important knowledge gained from lessons (β = .769, *SE* = .032), and Item 10 stating that learners were given distinct responsibilities in group works, projects, and research studies in online learning (β = .747, *SE* = .033), respectively. Of all five items, it is apparent that the results corresponded to other studies related to online learner-centered learning because, in four out of the five mentioned items, learners played a major role in self-learning via learning activities, problem-solving, self-assessment, and self-monitoring learning.

The analysis results of the cut scores determined based on the peaks on the test information curve revealed that there were three peaks that could be used to divide the respondents into three groups, including learners who needed assistance while learning online (0-22 points), learners receiving online standard learning (23-47 points), and learners receiving effective online learning (48 points or more). To divide the respondents into the mentioned groups, the effectiveness of the cut scores was analyzed using the Chi-square test and ANOVA with learning-related variables to ensure the effectiveness of the cut scores, discovering that majors were quality variables of online learning instruction that showed differences with the statistical significance of .05. The analysis results demonstrated that students in health sciences and sciences and technology belonged to Group 3, while those in humanities and social sciences could be categorized in Group 2, which corresponded to learning instruction characteristics and availability of technology. In other words, students in sciences and technology were technologically ready due to the Thai university learning context, based on observations and inquiring university officials. It was found that there had been blended learning, which integrated online learning even before the arrival of COVID-19 in sciences as well as sciences and technology. Additionally, most students in both majors are self-directed. Because most students went to tutorial schools before entering universities, they were accustomed to analyzing their strengths and weaknesses, selecting their learning styles, and self-assessment. According to the study by Dispongsant (2022) and Wonglertwisawakorn et al. (2019), most subjects that students studied in tutorial schools for university admission were Physics, Chemistry, Biology, Math, and English, and 84.4% of them studied sciences. On the contrary, humanities and social sciences students might be unfamiliar with the mentioned learning styles. Therefore, they needed time to adapt and monitor their learning online.

The analysis results of the effectiveness of the cut scores using the Chi-square test with variables concerning self-study outcomes in the cognitive, psychomotor, and affective domains. The comparison with the traditional learning revealed that Group 3 students obtained a higher cognitive level along with the other areas with a statistical significance of .05. This reflected that the measurement was effective in setting Group 3 apart from Group 1 and 2, which were in line with the effectiveness of the cut scores using ANOVA with student-teacher engagement, online learning characteristics, self-directed learning, and self-assessment. It was discovered that respondents from each group perceived different outcomes in each variable with a statistical significance of .001, and the outcomes were arranged from low to high according to the nature of the group. Group 1 had lower outcomes than Group 2, and Group 2 had lower outcomes than Group 3, reflecting the effectiveness of the cut scores using the test information function. This is supporting evidence that the cut scores could be used to divide the online learning qualities.

It is suggested that the validation of the scale's quality and the estimation of cut scores reflect that the OLS contains appropriate psychometric properties supported by the two theories. Furthermore, the division of online learning quality into three groups could divide students effectively. University teachers could adopt the OLS with their students to explore areas in online learning that students need assistance with. It was found that 5% of the students, as could be seen from the total of Group 1 students in each variable, as shown in Table 4. Teachers could also assess their own online instruction quality from the items in the scale. Moreover, the researchers found the significance of conducting an in-depth study of Group 1 students, 3-5% of the total, in order to study concerns and problems that happened during online learning. Identifying the Group 1 students could help them learn better in an online environment (Mulrooney & Kelly, 2020). The educational institute should also oversee and assist teachers whose classes contain Group 1 students.

Conclusions and recommendations

In this study, the researchers aimed to validate OLS for undergraduate students and to estimate the OLS's cut score to distinguish online learning quality for blended and online learning in higher education. The data to validate OLS are from the responses of 1,459 undergraduate students. The OLS, a 4-point Likert scale, consists of 24 items and one trait. The OLS is validated by analyzing the psychometric properties based on the IRT approach, as well as evidence based on test content (experts), internal structure (EFA and

CFA), and relations to other variables (correlation). The estimation of cut scores is based on test information, which is divided students into three categories. The estimated cut scores are validated by the Chi-square test of online learning outcomes and variables related to online learning, including student-teacher engagement, online learning characteristics, self-directed learning, and self-assessment.

Regarding policy implementation, the researchers recommend that the results of the Chi-square test that investigates differences among the three groups of students, based on the types of universities and majors, illustrate that more differences were discovered in the type of majors than the type of universities. Hence, assistance should focus on majors that require more assistance, such as humanities and social sciences, rather than health, science, and technology. To promote effective online learning in Thai universities, the effort should not only focus on the teaching approach but also on infrastructures to support online learning environments, such as IT devices, high-speed Internet access, and digital literacy (Mulrooney & Kelly, 2020; Lawthong et al., 2024).

For OLS applications, users should not use this scale to judge online course quality; instead, OLS should be used to acquire information to find areas for improvement and enhance the quality of online learning. In other words, OLS should be used for formative, not summative, purposes. Other evidence should be included to determine the design of effective online learning for higher education instructors. The strength point of OLS is that instructors can receive feedback on their teaching from student perspectives who respond to this scale, not from the instructor's own perceptions.

Future studies should focus on students' actual online learning outcomes because the current study estimates learning outcomes from students' self-perception by perceived scales. Moreover, all 24 items in the OLS have high reliability; therefore, the OLS has the potential to be developed into a short form. Also, accuracy and consistency indices should be investigated (Wyse & Hao, 2012) to study the effectiveness of the OLS's cut scores.

Acknowledgement

The authors would like to acknowledge the National Research Council of Thailand (N`RCT) for the research funding.

References

Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation, 31*, 162-172. https://doi.org/10.1016/j.stueduc.2005.05.008

Artino, A. R., & McCoach, D. B. (2008). Development and initial validation of the online learning value and self-efficacy scale. *Journal of Educational Computing Research, 38*(3), 279–303. https://doi.org/10.2190/EC.38.3.c

ASEAN University Network. (2020). *ASEAN University network quality assurance: Guide to AUN-QA assessment at programme level version 4.0.* ASEAN University Network.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation
.

Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S. L. (2009). Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education, 12*(1), 1-6. https://doi.org/10.1016/j.iheduc.2008.10.005

Barrot, J. S., Llenares, I. I., & del Rosario, L. S. (2021). Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines. *Education and Information Technologies, 26*, 7321-7338. https://doi.org/10.1007/s10639-021-10589-x

Bichi, A. A., & Talib, R. (2018). Item response theory: An introduction to latent trait models to test and item development. *International Journal of Evaluation and Research in Education, 7*(2), 142-151. https://doi.org/10.11591/ijere.v7i2.12900

Bolliger, D. U., & Halupa, C. (2012). Student perceptions of satisfaction and anxiety in an online doctoral program. *Distance Education, 33*(1), 81-98. http://dx.doi.org/10.1080/01587919.2012.667961

Bolliger, D. U., & Martindale, T. (2004). Key factors for determining student satisfaction in online courses. *International Journal on E-Learning, 3*(1), 61-67. https://www.researchgate.net/publication/284041463_Key_factors_for_determining_student_satisfaction_in_online_courses

Chalmers, P. (2023). *MIRT: Multidimensional Item Response Theory.* https://cran.r-project.org/web/packages/mirt/index.html

Cheon, J., Cheng, J., & Cho, M. (2021). Validation of the online learning readiness self-check survey. *Distance Education, 42*(4), 599-619. https://doi.org/10.1080/01587919.2021.1986370

Chuwuedo, S. O., Mbagwu, F. O., & Ogbuanya, T. C. (2021). Motivating academic engagement and lifelong learning among vocational and adult education students via self-direction in learning. *Learning and Motivation, 74*, 1-12. https://doi.org/10.1016/j.lmot.2021.101729

Deshwal, P., Trivedi, A., & Himanshi, H. L. N. (2017). Online learning experience scale validation and its impact on learners' satisfaction. *Procedia Computer Science, 112,* 2455-2462. https://doi.org/10.1016/j.procs.2017.08.178

Dispongsant, P. (2022). Satisfaction expectation and demand in online class of tutoring school during COVID-19 pandemic. *Journal of MCU Nakhondhat, 9*(4), 140-154. https://so03.tci-thaijo.org/index.php/JMND/article/view/260483

Dos Santos, L. M. (2022). Online learning after the COVID-19 pandemic: Learners' motivations. *Frontiers in Education, 7,* 1-13, https://doi.org/10.3389/feduc.2022.879091

Doucette, D., Cwik, S., & Singh, C. (2021). "Everyone is new to this": Student reflections on different aspects of online learning. *American Journal of Physics, 89*(11), 1042-1047. https://doi.org/10.1119/10.0005795

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16,* 5-18. https://doi.org/10.1007/s11136-007-9198-0

Ehlers, U. D. (2009). Web 2.0 - E-Learning 2.0 - quality 2.0? Quality for new learning cultures. *Quality Assurance in Education, 17*(3), 296-314. https://doi.org/10.1108/09684880910970687

Ferketich, S. (1991). Focus on psychometrics aspects of item analysis. *Research in Nursing & Health, 14*(2), 165-168. https://doi.org/10.1002/nur.4770140211

Firmansyah, R., Putri, D. M., Wicaksono, M. G. S., Putri, S. F., Widianto, A. A., & Palil, M. R. (2021). Educational transformation: An evaluation of online learning due to COVID-19. *International Journal of Emerging Technologies in Learning, 16*(7), 61-76. https://doi.org/10.3991/ijet.v16i07.21201

Froyd, J., & Simpson, N. (2008). Student-centered learning addressing faculty questions about student-centered learning. *The Course, Curriculum, Labor, and Improvement Conference, 30*(11), 1–11.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Pearson.

Geng, S., Law, K. M. Y., & Niu, B. (2019). Investigating self-directed learning and technology readiness in blending learning environment. *International Journal of Educational Technology in Higher Education, 16*(17), 1-22. https://doi.org/10.1186/s41239-019-0147-0

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales [Paper presentation]. *2003 Conference, Columbus, OH, United States*. http://hdl.handle.net/1805/344

Gustiani, S. (2019). Research and development (R&D) method as a model design in educational research and its alternatives. *Holistics Journal Hospitality and Linguistics, 11*(2), 12-22.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3). 38-47. https://doi.org/10.1111/j.1745-3992.1993.tb00543.x

Hung, M., Chou, C., Chen, C., & Own, Z. (2010). Learner readiness for online learning: Scale development and student perceptions. *Computer & Education, 55*, 1080-1090. https://doi.org/10.1016/j.compedu.2010.05.004

Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Applied Psychological Measurement, 40*(8), 559-572. https://doi.org/10.1177/0146621616664046

Jones, K., & Sharma, R. (2020). *Reimagining a future for online learning in the post-COVID era*. http://dx.doi.org/10.2139/ssrn.3578310

Junus, I. S., Santoso, H. B., Lsal, R. Y. K., & Utomo, A. Y. (2015). Usability evaluation of the student-centered e-learning environment. *International Review of Research in Open and Distributed Learning, 16*(4), 62-82. http://dx.doi.org/10.19173/irrodl.v16i4.2175

Karatas, K., Sentürk, C., & Teke, A. (2021). The mediating role of self-directed learning readiness in the relationship between teaching-learning conceptions and lifelong learning tendencies. *Australian Journal of Teacher Education, 46*(6), 54-77. https://doi.org/10.14221/ajte.2021v46n6.4

Kaufmann, R., Sellnow, D. D., & Frisby, B. N. (2015). The development and validation of the online learning climate scale (OLCS). *Communication Education, 65*(3), 307-321. https://doi.org/10.1080/03634523.2015.1101778

Keiler, L. S. (2018). Teachers' roles and identities in student-centered classrooms. *International Journal of STEM Education, 5,* 1-20. https://doi.org/10.1186/s40594-018-0131-6

Klincunhom, T. (2013). *A causal model of student engagement and academic achievement by teachers' effects* [Master's thesis]. Chulalongkorn University Intellectual Repository (CUIR). http://cuir.car.chula.ac.th/handle/123456789/43173

Lawthong, N., Thepsathit, P., Poondej, C., Suwathanpornkul, I. & Lerdpornkulrat, T. (2024). Online student-centered learning and assessment during the COVID-19 pandemic through perspectives of higher education teachers and students. *Journal of Education Studies, 52*(1), 1-18. https://so02.tci-thaijo.org/index.php/EDUCU/article/view/265749

Marinoni, G., Van't Land, H., & Jensen, T. (2020). *The impact of COVID-19 on higher education around the world*. IAU Global Survey Report. https://www.uniss.it/sites/default/files/news/iau_covid19_and_he_survey_report_final_may_2020.pdf

Mascolo, M. F. (2009). Beyond student-centered and teacher-centered pedagogy: Teaching and learning as guided participation. *Pedagogy and the Human Sciences, 1*(1), 3-27. https://www.researchgate.net/publication/228694201_Beyond_Student-Centered_and_Teacher-Centered_Pedagogy_Teaching_and_Learning_as_Guided_Participation

Miranda, J., Navarrete, C., Noguez, J., Molina-Espinosa, J., Ramirez-Montoya, M., Navarro-Tuch, S. A., Bustamante-Bello, M., Rosas-Fernandez, J., & Molina, A. (2021). The core components of education 4.0 in higher education: Three case studies in engineering education. *Computers & Electrical Engineering, 93*, 1-13. https://doi.org/10.1016/j.compeleceng.2021.107278

Mulrooney, H. M., & Kelly, A. F. (2020). COVID-19 and the move to online teaching: Impact on perceptions of belonging in staff and students in a UK widening participation university. *Journal of Applied Learning and Teaching, 3*(2), 17-30. https://doi.org/10.37074/jalt.2020.3.2.15

North Carolina Community College System. (n.d.). *Online readiness assessment.* https://vlc.nccommunitycolleges.edu/faculty/online-readiness-checklist/

Overby, K. (2011). Student-centered learning. *ESSAI, 9,* 109-112.

Pal, D., & Vanijja, V. (2020). Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. *Children and Youth Services Review, 119,* 1-12. https://doi.org/10.1016/j.childyouth.2020.105535

Penn State University. (n.d.). *Online learning readiness questionnaire.* http://tutorials.istudy.psu.edu/learningonline/ORQ/ORQ.htm

Pornkul, C. (2009). *Design of learning management integrated reading, analytical thinking and writing* (2nd ed.). CU PRESS.

Revelle, W. (2024). *Psych: Procedures for psychological, psychometric, and personality research.* https://www.researchgate.net/publication/281345624_psych_Procedures_for_Psychological_Psychometric_and_Personality_Research_R_Package_Version_10-95

Rosseel, Y., Jorgensen, T. D., & Wilde, L. D. (2023). *Lavaan: Latent variable analysis.* https://cran.r-project.org/web/packages/lavaan/index.html

Sallah, U. K. M., Zulnaidi, H., Rahim, S. S. A., Zakaria, A. R. B., & Hidayat, R. (2019). Roles of self-directed learning and social networking sites in lifelong learning. *International Journal of Instruction, 12*(4). 167-182. https://doi.org/10.29333/iji.2019.12411a

Sun, Y., & Rogers, R. (2021). Development and validation of the online learning self-efficacy scale (OLSS): A structural equation modeling approach. *American Journal of Distance Education, 35*(3), 184-199. https://doi.org/10.1080/08923647.2020.1831357

Teo, T. C., & Divakar, A. (2021). Understanding the concepts of digital learning approaches: An empirical analysis of schools in developing countries. *Journal of Applied Learning and Teaching, 4*(1), 120-128. https://doi.org/10.37074/jalt.2021.4.1.18

Tiyawongsuwan, S. (2011). *Development of cut score setting procedure on ordinary national educational test in mathematics and science for twelfth students* [Doctoral dissertation]. Chulalongkorn University Intellectual Repository (CUIR). http://doi.org/10.14457/CU.the.2011.577

Tsai, C., Cho, M., Marra, R., & Shen, D. (2020). The self-efficacy questionnaire for online learning (SeQoL). *Distance Education, 41*(4), 472-489. https://doi.org/10.1080/01587919.2020.1821604

Tzafilkou, K., Perifanou, M., & Economides, A. A. (2022). Development and validation of students' digital competence scale (SdiCoS). *International Journal of Educational Technology in Higher Education, 19,* Article 30. https://doi.org/10.1186/s41239-022-00330-0

Vagias, W. M. (2006). *Likert-type scale response anchors.* Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.

Wei, H., & Chou, C. (2020). Online learning performance and satisfaction: Do perceptions and readiness matter? *Distance Education, 41*(1), 48-69. https://doi.org/10.1080/01587919.2020.1724768

Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist, 34*(5), 719-751. https://doi.org/10.1177/0011000006286345

Wonglertwisawakorn, C., Lohtrakul, N., Ippoodom, K., Chensupakarn, C., Manasuth, S. M., Nimitrat, S. N., Silpaiboonpanich, W., Tubtimkeow, P., Banchanurat, S., Piyajitti, I., & Tomanit, G. (2019). Prevalence of stress in high school students data collected at Wannasorn tower. *Vajira Medical Journal: Journal of Urban Medicine, 63*(Supplement), 19-30.

Wright, G. B. (2011). Student-centered learning in higher education. I*nternational Journal of Teaching and Learning in Higher Education, 23*(3), 92-97. https://files.eric.ed.gov/fulltext/EJ938583.pdf

Wyse, A. E. (2017). Five methods for estimating Angoff cut scores with IRT. *Educational Measurement: Issues and Practice, 36*(4), 16-27. https://doi.org/10.1111/emip.12161

Wyse, A. E., & Babcock, B. (2016). Does maximizing information at the cut score always maximize classification accuracy and consistency? *Journal of Educational Measurement, 53*(1), 23-44. https://doi.org/10.1111/jedm.12099

Wyse, A. E., & Hao, S. (2012). An evaluation of item response theory classification accuracy and consistency indices. *Applied Psychological Measurement, 36*(7), 602-624. https://doi.org/10.1177/0146621612451522

Yan, Z. (2018). The self-assessment practice scale (SaPS) for students: Development and psychometric studies. *The Asia-Pacific Education Researcher, 27*(2). 123-135. https://doi.org/10.1007/s40299-018-0371-8

Yang, H., Su, J., & Bradley, K. D. (2020). Applying the Rasch model to evaluate the self-directed online learning scale (SDOLS) for graduate students. *International Review of Research in Open and Distributed Learning, 21*(3), 99-120. https://doi.org/10.19173/irrodl.v21i3.4654

Yeşilyurt, F. (2021). The learner readiness for online learning: Scale development and university students' perceptions. *International Online Journal of Education and Teaching, 8*(1), 29-42. https://doi.org/10.1016/j.compedu.2010.05.004

Yilmaz, A. B., & Karataş, S. (2018). Development and validation of perceptions of online interaction scale. *Interactive Learning Environments, 26*(3), 337-354. https://doi.org/10.1080/10494820.2017.1333009

Zemsky, R. (2014). With a MOOC MOOC here and a MOOC MOOC there, here a MOOC, there a MOOC, everywhere a MOOC MOOC. *The Journal of General Education, 63*(4), 237-243. https://doi.org/10.5325/jgeneeduc.63.4.0237

Ziesky, M. J., Perie, M., & Livingston, S. (2008). *Cutscores: A manual for setting standards of performance on educational and occupational tests*. Educational Testing Service.

Zumbo, B. D. (2016). Standard-setting methodology: Establishing performance standards and setting cut-scores to assist score interpretation. *Applied Physiology, Nutrition, and Metabolism, 41*(6), 74-82. https://doi.org/10.1139/apnm-2015-0522

Appendices

Appendix A. The Online Learning Scale (OLS).

Online Learning Scale's Items	N	R	S	O
1) Teachers integrated existing knowledge into a new lesson.				
2) Learners studied additional video clips, lessons, and exercises.				
3) Teachers made an interesting introduction into the lesson by using applications or discussion questions.				
4) Learners searched for additional knowledge both online and offline to do group works, projects, or research.				
5) Learners studied from guest speakers who were experts in different fields.				
6) Learners analyzed problems and contributing factors of a specified incident.				
7) Learners studied online, on site, and outside the class time.				
8) Learners participated in thought-provoking activities, giving rise to learning and skill development				
9) Learners participated in group works, projects, and research in the stages of planning, implementation, making an outcome summary, and online self-directed study.				
10) Learners were given distinct responsibilities in group works, projects, and research studies in online learning.				
11) Online activities enhanced students' knowledge.				
12) Learners and peers collaboratively studied online.				
13) Learners had a chance to exchange opinions and discuss with friends, both online and offline.				
14) Learners were assigned group works on Breakout Rooms.				
15) Learners talked and exchanged opinions with teachers during online classes.				
16) Learners were allowed to inquire or answer students' questions using several functions such as using the "unmute" feature or the chat box.				
17) Learners produced multimedia to present assigned tasks.				
18) Learners presented their works by screen sharing.				
19) Learners presented group works/ projects/ research to the public via various online channels.				
20) Teachers supported and facilitated group works, projects, and research through online learning.				
21) Teachers allocated time for students' consultations and answering students' questions outside the normal class time.				
22) Learners reflected important knowledge gained from lessons.				
23) Learners studied both online and offline.				
24) Teachers clearly informed students about submission platforms such as e-mail, Line, blogs, and more, and deadlines.				

Note: N = Never, R = Rarely, S = Sometimes, and O = Often

Appendix B. The summary of university types and study fields.

Type of University	Field of study					Total
	health science	science and technology	humanity	social science	other	
public university	63	58	558	350	0	1,029
Rajamangala university of technology	0	17	2	191	0	210
Rajabhat university	0	14	3	145	0	162
private university	0	6	0	29	0	35
community university	8	2	0	7	0	17
others	2	0	2	1	1	6
Total	73	97	565	723	1	1,459

Appendix C. Four scales used in correlation analysis with other variables.

Scale	SD	D	A	SA
Teacher-Student Engagement Scale				
1) Teachers relentlessly attempted to tackle any obstacle students faced in online learning				
2) Teachers were enthusiastic to teach online.				
3) Teachers dedicated themselves to online teaching.				
4) Teachers were always available when students needed advice.				
5) Teachers kept students informed about the objectives of each class.				
6) Teachers made interesting plans for online learning activities.				
7) Teachers designed appropriate online learning media.				
8) Teachers made the class atmosphere facilitative to online learning.				
9) Teachers controlled students' behaviors to suit online learning.				
10) Teachers were able to solve urgent problems during online learning.				
11) Teachers were able to cover all the intended content in an online class in time.				
Online Learning Characteristics Scale				
1) Learners were able to understand new knowledge in a short period of time.				
2) Learners were able to find suitable solutions for unfamiliar problems.				
3) Learners were able to study well when they involved self-study in the process.				
4) Learners preferred both self-study and studying in groups.				
5) Learners were able to discuss with strangers.				
6) Learners required minimum guidance from teachers.				
7) Online learning corresponded to students' ways of life.				
8) Learners were satisfied with different methods of online learning (such as messages, videos, podcasts, online discussions, and video conferences).				
9) Learners tended to establish goals and deadlines for their own works.				
10) Learners refused to give up or quit merely because things seemed difficult.				
11) Learners were confident that they could succeed in online learning.				
12) Learners felt joyful when studying online.				
Self-Directed Learning Scale				
1) Learners could control themselves when studying online.				
2) Learners played a significant role in online learning.				
3) Learners could study or work effectively despite disturbances.				
4) Learners constantly enjoyed studying without teachers' supervision.				
5) Writing and posting messages in online platforms helped students arrange their own thoughts.				
Self-Assessment Scale				
1) Learners enjoyed reviewing lessons and doing extra exercises even though they were proficient in the content.				
2) Learners tended to ask themselves questions to enhance their own understanding in the lessons.				
3) Learners checked the accuracy of each task by comparing their answers with different sources.				
4) Learners asked teachers to provide feedback to develop their works.				
5) Learners asked their peers to provide feedback to develop their works.				
6) Learners could assess their own work qualities.				
7) Learners tended to recheck the accuracy of the works when they felt unsure.				
8) Learners recognized the learning methods that best suited themselves.				
9) Learners observed their own mistakes and weaknesses when doing exercises to fill the loopholes.				
10) Learners paid attention to their own self-assessment to bring about improvements in the future assessments.				

Note: SD = Strongly Disagree, D = Disagree, A = Agree, and SA = Strongly Agree. Four scales are translated from Thai to English.