



Vol.8 No.1 (2025)

Journal of Applied Learning & Teaching

ISSN : 2591-801X

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

The impact of AI- assisted language learning tools on augmenting university EFL students' speaking skills in Jordan

Naima Ahmad Al-husban^A A

Education and general studies faculty–Arab Open University-Jordan

Keywords

Artificial intelligence;
artificial intelligence-assisted language learning;
ELSA speech analyzer application;
English as a foreign language;
higher education;
rubrics;
speaking fluency;
speaking standards.

Abstract

This study aims to investigate the effect of the ELSA speech analyzer application, an artificial intelligence platform, on developing the speaking skills of English as a Foreign Language (EFL) students. The participants encompassed 34 university EFL students who were randomly assigned to either an experimental group or a control group. The students in the experimental group received instruction and assessment utilizing the ELSA speech analyzer application, while those in the control group were taught using traditional methods, specifically rubrics. Data were collected through the application of both the AI tool and rubrics. The findings indicated that the experimental group outperformed the control group in overall speaking skills and in all standards of speaking, including pronunciation, intonation, fluency, vocabulary, and grammar. Based on these results, it is essential to incorporate AI tools into university course descriptions and the teaching and assessment processes. Moreover, tutors should receive professional development training on efficiently employing AI applications like ELSA to maximize their potential in engaging students through interactive tasks, personalized real-life activities, and adapting course content to align with AI tools. Additionally, it is crucial to enhance students' use of AI applications in learning English, particularly in speaking, and to evaluate students' speaking skills through AI platforms.

Correspondence

n_husban@aou.edu.jo^A

Article Info

Received 18 September 2024
Received in revised form 11 December 2024
Accepted 30 December 2024
Available online 24 January 2025

DOI: <https://doi.org/10.37074/jalt.2025.8.1.13>

Introduction

Speaking English is a fundamental skill that EFL teachers strive to cultivate in students from early education through their university years. Teachers acknowledge the importance of this skill, as students will be required to employ it in several aspects of their lives, particularly in the context of globalization. English serves as the primary language for communication, science, and media. Consequently, it is essential for students to engage and interact effectively with others in English. However, achieving spoken proficiency in English can be challenging, especially in classroom settings where many students are primarily involved in listening activities as asserted by Tagg (2020).

Fear and anxiety are common experiences for students when participating in speaking tasks in front of their peers (Fulcher, 2015). At the same time, EFL tutors face substantial challenges in providing feedback on individual students' oral performance, particularly in large classes with a high student-to-teacher ratio. This issue is compounded at the university level, where speaking skills are sometimes overlooked, and evaluations are conducted without offering constructive feedback. Fan and Yan (2020) highlighted that teaching and assessing speaking skills often require students to make multiple attempts to improve, as psychological factors frequently hinder progress. Furthermore, the complexity of speaking skills, as elaborated by Huettig et al. (2022), can lead to their deprioritization in language instruction.

The psycholinguistic theory posits that second or foreign language speakers must possess cognitive skills that enable effective language production, anticipate conversational messages, and navigate social conversational dynamics. Assessing these speaking skills poses a considerable challenge for educators due to the multifaceted nature of the evaluation, which encompasses pronunciation, intonation, vocabulary, grammar, and fluency. This complex process demands significant time and expertise to objectively evaluate EFL (English as a foreign) students' proficiency in each criterion. Moreover, the interaction between speaker and listener, involving both verbal and nonverbal behaviors, further complicates the assessment process (Huettig et al., 2022).

Winaryo and Gusdian (2021) investigated speaking fluency levels using a rubric with three specific criteria: speech rate, pause rate, and the occurrence of disfluent syllables. Their findings revealed that participants generally exhibited an intermediate level of speaking proficiency. However, prolonged silent pauses and notable pronunciation challenges were prevalent among the participants. Additionally, students often reported feeling self-conscious and inadequately prepared, expressing a need for more opportunities to practice. These results underscore the importance of providing further practice opportunities to minimize silent pauses and foster self-correction during speaking tasks.

Traditionally, proficiency assessments of EFL learners' speaking skills have relied on human raters (De Jong, 2023). While effective, this method introduces challenges related to reliability, as scores must remain consistent across

different raters. Hana et al. (2022) identified potential issues such as leniency or severity, inconsistent scale usage, central tendency, and bias among raters. Furthermore, the time-intensive nature of human-led assessments often hampers their efficiency. To mitigate these limitations, Isaacs (2018) proposed a hybrid scoring method that integrates human ratings with automated measures. Supporting this approach, Gu et al. (2021) demonstrated the utility of automated analytic feedback, particularly using the TOEFL iBT (Internet-based Test of English as a Foreign Language) 'speech-rater' engine, to enhance the reliability and efficiency of speaking assessments.

The above discussion highlights a significant challenge in the research and development of speaking assessments: the need to explore practical, readily available technological tools that both teachers and students can use to practice speaking skills anytime and anywhere, while enabling teachers to objectively assess students' performance. Mills et al. (2023) emphasize that this trend has been accelerated by the COVID-19 pandemic, which prompted the closure of educational institutions and the transition to online instruction and assessment. Higher education institutions are increasingly recognizing the potential of generative artificial intelligence to perform assessment tasks traditionally carried out by human raters.

This emerging issue underscores the potential of AI-assisted language learning to enhance speaking instruction for both students and tutors. Artificial intelligence is rapidly transforming various aspects of life, including education, and presents innovative solutions as well as challenges for addressing teaching, learning, and assessment challenges (Tan, 2020; Rudolph et al., 2023a; Rudolph et al., 2023b). Zou et al. (2020) emphasize the need for software that develops speaking proficiency, provides personal feedback, and analyzes speakers' voices and speech according to the same standards used by teachers. Such tools can create a less intimidating environment, allowing students to engage with AI-assisted language learning in a supportive setting. Gamage et al. (2023) highlight how using AI applications supports students by analyzing their performance data, offering recommendations based on strengths and weaknesses, and tailoring instruction to meet individual needs. According to Popenici (2023), AI exhibits significant potential for tutors, including automating assessment, personalizing teaching, and providing individualized assistance. Similarly, Rudolph et al. (2023a) highlighted that intelligent tutoring systems are one of the most outstanding benefits of AI in transforming education. Consequently, numerous researchers have explored this area, examining both the advantages and drawbacks of AI applications depending on the approach taken.

Emerging technologies play a pivotal role in enhancing learners' pronunciation skills by delivering personalized feedback, real-time assessments, and adaptive learning opportunities (Mohammadkarimi, 2024). These tools also expand accessibility and flexibility within language learning environments, providing learners with dynamic opportunities to improve their skills. Similarly, language learning applications such as ELSA have proven highly effective in engaging students with adaptable content

aligned to English language curricula and suitable for various proficiency levels. By fostering a supportive and risk-free environment, these tools empower students to practice without fear of mistakes, encouraging more autonomous and confident learning experiences (Hana et al., 2022).

Artificial intelligence (AI) is broadly defined as a computerized system or virtual assistant capable of mimicking human-like activities, including learning, adapting, self-correcting, and performing complex tasks (Popenici, 2023). Recent advances in AI technology, such as speech evaluation programs utilizing automatic speech recognition, have demonstrated significant potential in supporting EFL learners to enhance their speaking skills. These tools are lauded for their user-friendly nature, engaging interfaces, and ability to improve pronunciation, vocabulary, and fluency. Furthermore, they alleviate the anxiety often associated with speaking in front of peers by providing a low-pressure practice environment (Zou et al., 2023; Dai & Wu, 2023; Tai & Chen, 2024). As a result, AI-driven speech evaluation tools have gained widespread popularity among educators and learners alike.

AI speech evaluation tools serve as interactive conversation partners for EFL learners, enabling them to practice a range of speaking topics in a stress-free setting. These tools also provide detailed feedback on speaking proficiency, covering pronunciation, fluency, vocabulary, and grammar (Jeon et al., 2023). Examples include Duolingo, an AI language-learning platform known for its engaging and interactive approach, and ELSA, a specialized AI-powered tool for improving speaking skills. ELSA offers features such as pronunciation correction and fluency enhancement, accessible through platforms like [https://speechanalyzer.elsaspeak.com/] (Kholis, 2021). Together, these innovations have made significant contributions to language education by enhancing learning outcomes and addressing learners' individual needs.

Numerous studies have explored the effectiveness of AI-based speech evaluation programs in developing EFL learners' speaking skills. Zou et al. (2023) investigated how feedback provided by such tools impacts skill development among 40 EFL learners in China. The results indicated that most participants believed their speaking abilities improved due to the tool's feedback. These findings were based on mean differences between experimental and control groups. However, Zou et al. (2023) also highlighted the importance of ensuring that AI feedback systems accurately reflect users' progress to maintain credibility and utility.

In a related study, Zainuddin and Mohamad (2024) examined the influence of the SpeechAce tool on ESL students' speaking skills. Their findings revealed that students found the tool both beneficial and user-friendly. Participants demonstrated a positive attitude toward using SpeechAce, and teachers observed noticeable improvements in students' pronunciation. The tool's ability to deliver detailed, efficient feedback saved time and fostered learner autonomy by enabling its use outside the classroom. However, the study did not evaluate longitudinal changes in speaking skills or directly compare feedback from the tool with traditional assessments.

Hanna et al. (2022) compared the impact of ELSA and Duolingo on students' speaking abilities. Involving 30 participants divided into two groups, the study found that students using ELSA demonstrated slightly better speaking skills than those in the control group. Similarly, Mohammadkarimi (2024) examined AI's role in improving English pronunciation among 78 elementary and pre-intermediate learners, alongside 19 experienced teachers. Pre- and post-test results showed significant pronunciation accuracy improvements in the experimental group. Participants expressed positive attitudes toward AI-driven tools, citing enhanced pronunciation, increased confidence, and heightened engagement. Nonetheless, concerns emerged about feedback interpretation and the tools' sensitivity to subtle pronunciation differences.

Tai and Chen (2024) and Jayasinghe (2024) evaluated the effectiveness of generative AI (GAI) chatbots in improving EFL speaking skills. Their study involved 85 students who were divided into three groups: individual interaction with the coolE bot, paired interaction with the bot, and traditional teacher-led classroom interaction. Results indicated positive outcomes for all groups using the coolE bot, whether individually or in pairs. Students appreciated the tool's relevance to their interests, its motivating nature, and its ability to create a supportive, anxiety-free environment. The tool fostered rapport, boosted confidence, and reduced speaking-related anxiety. Similarly, Kim et al. (2021) reported that AI chatbots improved speaking proficiency compared to face-to-face instruction. Despite challenges such as misunderstandings caused by unfamiliar vocabulary, students valued the opportunity for low-pressure practice with AI, likening it to conversing with a native speaker.

The potential of AI-assisted language learning applications, such as Duolingo, was also explored by Qiao and Zhao (2023). Their study on Chinese learners revealed that the experimental group using AI tools showed significant improvements in speaking skills compared to the control group. The authors called for further research to investigate the long-term effects and underlying mechanisms contributing to these improvements.

The impact of ELSA on pronunciation skills was independently studied by Alwin et al. (2024) and Al-Shallakh (2024). Both studies concluded that ELSA effectively enhanced pronunciation by delivering detailed feedback for accurate sound production, reinforcing its value as a practical tool for language learning.

It can be concluded that AI speech evaluation tools significantly improve students' speaking fluency, pronunciation, and intonation. Participants expressed positive views, noting that these tools offer comprehensive feedback based on international standards and detailed reports on various speaking proficiency aspects. Furthermore, most practice sessions occurred outside the classroom, creating a stress-free environment.

This study utilized the ELSA-speech analyzer as a mobile app and desktop tool for assessing spontaneous English speech. It allows students to engage in activities like interviews and presentations while preparing for tests such as IELTS

(International English Language Testing System), TOEIC (Test of English for International Communication), or Test of English as a Foreign Language (TOEFL). Students can record their speeches for analysis, receiving personalized AI-driven feedback (Anguera et al., 2023).

This tool has been selected because of its support of non-native speakers, especially learners of English as a foreign language. The ELSA speech analyzer's automatic speech recognition engine achieves low word error rates for non-native and low-proficiency speakers. It also provides a speaking score aligned with International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL) exam scales. The overall score combines various aspects of communication, such as pronunciation, intonation, fluency, grammar, and vocabulary. Users receive feedback to improve their speaking skills, and a customized learning program addresses specific weaknesses. The tool's engagement is enhanced by human-like voices that ask questions. Likewise, Rudolph et al. (2023a) suggested [rephrase] that teachers employ AI applications to reduce their workloads, facilitate innovation in classroom, and the tools are designed to assist teachers by automating assessment and feedback mechanisms to provide students with further insights about the opportunities of developing their skills.

Despite the promising potential of artificial intelligence (AI) speech evaluation tools for enhancing students' speaking proficiency, there is a pressing need for large-scale empirical studies to comprehensively assess their effectiveness in the context of English as a Foreign Language (EFL) learning. These tools offer students the advantage of receiving immediate feedback and targeted pronunciation practice. However, EFL learners may perceive the evaluation process as inequitable if they receive low scores, which can be attributed to the tools' heightened sensitivity to the speech patterns of native speakers (Belda-Medina & Calvo-Ferrer, 2022). This situation underscores the need for a more in-depth examination of the evaluation process and the acceptance of its stringent nature when assessing learners. Wu (2024) has underscored this issue, highlighting the significance of students' educational levels in moderating the impact of AI tools on learning outcomes. Furthermore, Wu (2024) found that students who utilized AI tools did not demonstrate superior learning outcomes compared to their peers who did not employ such tools. Consequently, the uncertainty surrounding the influence of AI tools on students' speaking proficiency represents a critical area that warrants further investigation.

This study seeks to address the identified gaps in the existing literature by comparing the speaking proficiency outcomes of two distinct groups of students. One group received instruction and assessment through traditional evaluation methods, while the other group utilized AI speech evaluation tools. The analysis was conducted using the reports generated by the AI speech evaluation. The motivation for this research arises from the limitations observed in prior studies, which failed to incorporate feedback reports from these tools and did not compare them with traditional assessments conducted by educators. To ensure consistency, both groups were evaluated according to the same speaking

proficiency standards. The primary objective of this study is to assess the impact of the ELSA speech analyzer application on students' speaking skills. Accordingly, this research aims to address the following research questions:

- How is the ELSA speech analyzer application different from rubrics in terms of assessment approach?
- Is there a statistically significant difference between the speaking skill means of students who were instructed using ELSA speech analyzer application and those instructed using traditional approaches to teaching speaking skills?

Method

The design of the study

A quasi-experimental design was employed in this study. It is based on an independent variable, which employs an ELSA speech analyzer as an AI-assisted language learning tool during teaching and assessing the participants in the study while teaching speaking, and the dependent variable, which is EFL students' speaking skills.

The participants of the study

The participants in the current study were 34 female students in the first year of their enrollment at universities in Jordan; they were purposefully selected from some private universities. Their age ranged from 18 to 19 years old. 17 of them were randomly assigned to the intervention group, and the other 17 students were designated to the control group. Their language proficiency placement test scores were taken into account by random distribution into the two groups. The same EFL tutor, with five years of experience, taught them.

The data collection procedures

To answer the research questions, the researcher designated some speaking topics embedded in the communication skills course, such as emotional intelligence, future career, and protecting data. A quasi-experiment was conducted, and two groups were randomly assigned. The experiment took place from April to August 2024. The intervention followed the following procedures:

1. The general speaking proficiency level for both groups was determined based on the results placement exam conducted by the university, which showed that all students' speaking proficiency ranged from middle to low level. Speaking topics were selected from the assigned learning outcomes of the English communication course, there was no need to establish its validity, as there are juries of specialized experts who developed it.

2. The experimental group and its tutor have been trained on how to employ the ELSA speech analyzer application. This application uses AI and ELSA stands for English Language Speech Assistant. It listens to the learners' spontaneous speech and provides immediate feedback on pronunciation, fluency, grammar, and vocabulary. It also indicates the possible score they would obtain in major English-speaking exams and is available on mobile devices and via web browsers.
3. The sampled students were asked to speak about the different topics of speaking to ensure the reliability of the AI tool and the rubric. Thus, Cronbach alpha coefficient was 0.89, and it was appropriate for the purpose of this study.
4. The EFL tutor received training on utilizing the ELSA speech analyzer application. This AI-powered application offers multiple functions to improve English speaking skills by providing a real-time transcription of a speech, detailed analysis of criteria like pronunciation, intonation, pace, and grammar, opportunities to practice speaking for presentations, meetings, interviews, and through self-assessments, support for group conversations and online meetings, and user-friendly interface by simply turning on microphone and start recording speech.
5. The tutor created an account in the application and trained students to create their own accounts. He then explained to them how to use the application to practice speaking for presentations and meetings. Then, he started recording and checking the detailed feedback against each speaking standard, and proposals to improve students' performance.
6. The tutor created a WhatsApp app group to listen to students' comments about their use of the ELSA Speak application.
7. When students engaged with speaking topics on their individual accounts and reviewed the detailed feedback provided, the (EFL) tutor recorded the speaking performance of the experimental group for the assigned tasks. In contrast, the control group was instructed to discuss the same speaking topics and was evaluated using a speaking rubric. Rubrics serve as one of the assessment tools employed to evaluate students' performance through structured observation. The rubric encompasses criteria that tutors utilize to assess students' progress in achieving their respective proficiency levels (Dawson, 2017). The application of the rubric in this study was aligned with the ELSA Speak standards, which include pronunciation, intonation, fluency, grammar, and vocabulary.

8. Reports were extracted from the tool, and the differences in speaking skills between the experimental and control groups were analyzed.

Data collection instruments

ELSA Speak application

This AI-powered application is utilized to evaluate the speaking skills of students in the experimental group during both the pre-test and post-test phases. Students were encouraged to use the application while practicing their speaking skills prior to submitting their recordings for final assessment. This approach allows them to benefit from the comprehensive feedback provided by the application, which addresses aspects such as pronunciation, intonation, fluency, vocabulary, and grammar. The application offers tutorials tailored to students' proficiency levels for each standard; it not only assesses students but also aids in their development and practice of speaking skills according to their current capabilities. Furthermore, it is user-friendly and provides an overall assessment in various international grading systems, including the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL). Consequently, this application enables tutors to allocate more time to delivering innovative instruction rather than merely evaluating students' speaking abilities. Figure 1 illustrates the general template of results generated by the ELSA Speak application (Anguera et al., 2023; Alwin et al., 2024).

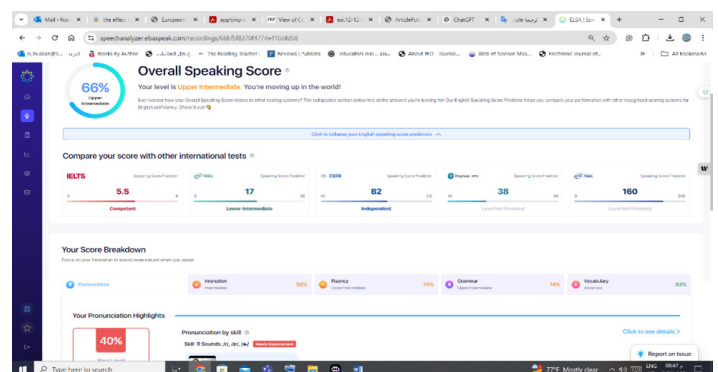


Figure 1. The standard template of the ELSA speak–speech analyzer.

To validate the tools, a rater reviewed the reports generated by the application. Initially, the rater manually assessed the students' speaking skills, which had been recorded using the ELSA speech analyzer application. Subsequently, the results of the human assessment were compared with the detailed feedback generated by the application. The agreement coefficient between the rater's two assessments was calculated to be 0.836, which is considered appropriate for this study and ensures reliable results, according to Wang et al. (2020).

The speaking skill assessment rubric was designed by reviewing widely recognized standards in language proficiency and assessment, such as Teaching English to Speakers of Other Languages (TESOL). It was specifically aligned with the standards that ELSA Speak focuses on,

which include pronunciation, intonation, fluency, grammar, and vocabulary. A group of experts in TEFL validated the rubric and provided inputs that were incorporated into it. The human assessment rubric for speaking skills was formulated to determine the effectiveness of AI assessment compared to human assessment on students' speaking performance.

Data analysis

A descriptive and inferential analysis was conducted to investigate the effect of delivering speaking instruction through the ELSA speech analyzer application on the speaking skills of EFL students. The Statistical Package for the Social Sciences (SPSS) was used, along with detailed feedback from the ELSA speech analyzer application used in this study. Descriptive statistics were presented, including percentages, frequencies, means, and standard deviations. As this study utilized a quasi-experimental design, an Analysis of Covariance (ANCOVA) was employed as an inferential statistic to determine if there were significant differences in the mean scores between the experimental and control groups.

Results

Question one

To answer research question number one, how is the ELSA speech analyzer application different from the rubric in terms of assessment approach?

In this section, two tools were employed to evaluate the proficiency of students in the primary speaking skills standards. For the control group, a speaking skills rubric was adopted, which aligned with the established benchmarks in language proficiency and assessment, including Teaching English to Speakers of Other Languages (TESOL). As part of the manipulation, the students were furnished with the rubric to ensure their comprehension of the criteria used for assessing their speaking abilities. Table 1 provides a comprehensive breakdown of the anticipated proficiency levels for each standard, ranging from one to four.

The tutor instructed the students to discuss the previously assigned tasks. Following this, students were given feedback on their current performance levels for each standard, as outlined in the rubric. Due to time constraints, there was no opportunity for additional practice or detailed feedback on each standard.

In contrast, the ELSA speech analyzer application provided students in the experimental group with various engaging speaking activities. For instance, the speech analyzer feature allowed students to record their speech about the task, which the tutors explained to them at any time. In addition to detailed assessments, grades, and feedback on each standard, the application also provided performance highlights for each sound pronounced. Furthermore, students were given access to transcripts and extra practice materials to help improve their pronunciation. For example, Figure 2 illustrates the overall grade of students in fluency

Table 1. Speaking assessment rubric based on reviewing widely recognized standards in language proficiency and assessment like Teaching English to Speakers of Other Languages (TESOL).

Standard	Exceed expectations 4	Meet expectations 3	Need improvement 2	Dissatisfied 1
Pronunciation	Pronunciation is clear and accurate, with native-like accuracy.	Pronunciation is generally clear, with minor errors that do not impede understanding.	Pronunciation is sometimes unclear, with noticeable errors that occasionally impede understanding.	Pronunciation is unclear, with frequent errors that significantly impede understanding.
Intonation	Uses varied intonation effectively to convey meaning and emotion.	Uses appropriate intonation with minor lapses that do not affect overall communication.	Uses limited intonation, affecting the expressiveness and engagement of the speech.	Uses flat or inappropriate intonation, affecting clarity and engagement.
Grammar	Demonstrates strong command of grammar with few or no errors.	Demonstrates good command of grammar with occasional errors.	Demonstrates some command of grammar but with frequent errors.	Demonstrates limited command of grammar with pervasive errors.
Vocabulary	Uses a wide range of vocabulary accurately and effectively.	Uses a good range of vocabulary, with minor inaccuracies.	Uses limited vocabulary, with frequent inaccuracies.	Uses very limited vocabulary, with frequent inaccuracies and misusage.
Fluency	Speaks fluently with natural pacing and minimal hesitation.	Speaks fluently with occasional pauses or hesitations.	Speaks with noticeable pauses and hesitations that affect the flow.	Speaks with frequent pauses and hesitations, significantly affecting the flow.
20				

assessment, along with detailed feedback. Additionally, the application provided students with information on their pace, pausing, and hesitations and the accompanying recording script revealed any unacceptable pauses made.



Figure 2. The feedback elements of fluency standards as displayed in the ELSA speech application.

In intonation, the application analyzed students' recordings and showed them their pitch variation, volume variation, and word prominence with a detailed explanation of the level the students are placed, using each criterion, as shown in Figure 3.

In grammar, the application determines all the grammatical errors made by the students in the transcript. Mistakes are highlighted in red, corrections are provided in green, and detailed feedback is given explaining why the error occurred and how to avoid making it in the future, as shown in Figure 4.

Whereas, in terms of vocabulary standards, the application provided AI recommendations that described the students' strengths and weaknesses. It identified areas for



Figure 3. The feedback elements of intonation standards as displayed in the ELSA speech application.

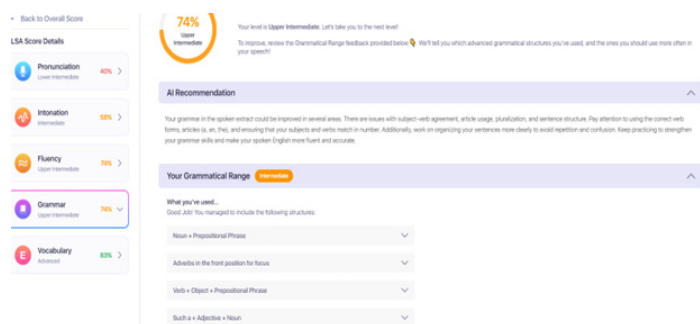


Figure 4. The feedback elements of grammar standards as displayed in the ELSA speech application.

improvement and expanded their vocabulary by offering more advanced alternatives to their current choices. Additionally, the application offered tutorials to enhance the students' vocabulary and other areas, as illustrated in Figure 5.

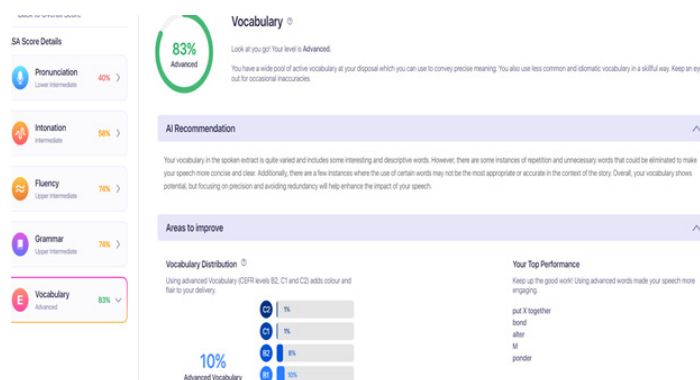


Figure 5. The feedback elements of vocabulary standards as displayed in the ELSA Speech application.

The tutor assisted students in creating their own accounts using their Gmail, either on their laptops or mobiles, and showed them the tools and activities that the applications provided students. The teacher also modelled how they could use it to practice IELTS speaking skills and self-assess their speaking before sending their final attempt to the tutor. In addition to that, the application provides challenging exercises in all the speaking standards to leverage students' speaking levels, all exercises and activities are engaging. As a person asks questions, the students record their answers, and at the end, the application provides each

student with a thorough report with feedback and detailed recommendations on how to correct their mistakes or develop their pronunciation and intonation. All of these points are not available while assessing students using rubrics. Students in the control group had one chance to speak; then the tutor showed him his performance level in the rubric from 4 (Exceed expectations) to 1 (Dissatisfied), as Table 1 shows.

Based on the general description of the nature of applying each assessment tool, it is important to investigate the effect of these tools on the performance of students in each standard. Table 2 shows the descriptive statistics displaying the differences between the two groups in their performance in each standard in the post-assessment.

Table 2. The descriptive statistics of EFL students' speaking skills performance in experimental and control groups in all standards.

Speaking skills standards	Group	Samples	mean	Standard deviation
Pronunciation	Experimental	17	3.84	0.24
	Control	17	2.88	0.33
Intonation	Experimental	17	3.61	0.47
	Control	17	2.65	0.49
Fluency	Experimental	17	3.51	0.51
	Control	17	2.58	0.51
Vocabulary	Experimental	17	3.57	0.51
	Control	17	2.60	0.51
Grammar	Experimental	17	3.64	0.24
	Control	17	3.10	0.43
Total	Experimental	17	18.17	0.44
	Control	17	13.81	0.56

As shown in Table 2, the overall mean of the experimental group outperformed the mean of the control group by 18.17 to 13.81, respectively. Likewise, the experimental group has higher mean scores in all the speaking skills standards, namely, pronunciation, intonation, fluency, vocabulary, and grammar. This consistency in improving all speaking skill standards implies the effectiveness of the intervention of using ELSA speech analyzer application on improving EFL student. Furthermore, it is apparent that students developed their speaking skills in pronunciation, intonation, and vocabulary more than grammar because traditional instruction focused mainly on structural delivery, and the main approach implemented is a grammar-translation approach while teaching, whereas pronunciation and intonation instruction received less attention by tutors because they need much time and experience.

Question two

A statistical test was implemented using statistical Package for Social Sciences (SPSS 28). Thus, we conducted a descriptive and inferential analysis to investigate the impact of delivering speaking skills instruction through ELSA speech analyzer application on EFL students' speaking skills. The author used Statistical Package for the Social Sciences (SPSS) and data employing the speaking rubric and the ELSA speech analyzer application. Descriptive statistics, such as percentages, frequencies, means, and standard deviations, were presented. Since this study employed a quasi-experimental design, we employed Analysis of Covariance (ANCOVA) as an inferential statistic to determine if there were significant differences in the mean scores between the

experimental and control groups.

Table 3. Means and standard deviation for the experimental and control groups on speaking skills.

Group	Pre-test			Post-assessment		
	N	Mean	Std. Deviation	N	Mean	Std. Deviation
Experimental	17	12.6	0.94	17	18.17	0.44
Control	17	12.	1.12	17	13.81	0.56
Total	34			34		

As shown in Table 3, the means of the experimental and control groups in the previous assessment of speaking skills were not significantly different from each other. On the other hand, after the manipulation, the mean of the experimental group was 18.17, with a standard deviation of 0.44. The mean of the control group, on the other hand, was 13.81, with a standard deviation of 0.56. This indicates that there was a significant difference in the mean scores between the experimental group and the control group on the post-assessment. To examine if the means differences between the experimental and control groups were statistically significant, analysis of covariance (ANCOVA) was computed. The results are delivered in Table 4.

Table 4. Analysis of covariance (ANCOVA) for the differences in speaking skills performance between the experimental and control groups.

		Levene's Test for Equality of Variances		T-test for Equality of Means						
		F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
test	Equal variances assumed	838.858	.000	-6.452	33	.000	0.256	.65644	-6.47213	-5.50374
	Equal variances not assumed			-6.452	25.746	.000	212.500	.77644	-7.28427	-4.09073

* Significant level at $\alpha = 0.05$

Table 4 unveils that there were statistically significant differences between the experimental and the control groups in the EFL students' speaking skills performance at ($\alpha = 0.05$), since (F) value was (838.858) with statistically significant level (0.000) which is significant at ($\alpha = 0.05$).

In order to determine the favor of differences between the experimental and control group, adjusted means for both the experimental and control groups in the post-test were calculated as shown in Table 5.

Table 5. The adjusted means and standard deviations for both the experimental and control groups in the post-test.

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Experimental	18.17	0.31	13.4	13.5
Control	13.81	0.28	9.6	9.8

As Table 5 shows, the adjusted mean of the experimental group in the post-test was (18.17) out of (20), and it is higher than the adjusted mean of the control group in the post-

test which was (13.81) out of (20). Thus, the null hypothesis "there are no statistically significant differences at ($\alpha = 0.05$) in the mean scores of EFL students' speaking skills between the experimental group and the control group attributed to the use of ELSA speech analyzer application and the conventional ones" was rejected. Instead, the alternative hypothesis was accepted, which refers to existing statistically significant differences at ($\alpha = 0.05$) to the effectiveness of the use of ELSA speech analyzer application on female EFL students' speaking skills in favor of the experimental group. These findings indicate insights for educators and policymakers in higher education who are seeking further context for implementing the integration of AI-assisted language learning with language instruction. Specifically, all universities have strategic outcomes aimed at producing fluent graduates who are ready to engage in the labor market.

Furthermore, the application in this study did not enforce specific topics, but rather the tutor had the flexibility to adapt the speaking topics in their course while utilizing the target tool. As a result, individual differences could be guaranteed because the tool suggests a remedial program for each student based on their individual feedback report.

Discussion and implications

This study investigates the impact of the ELSA speech analyzer application, an AI-assisted tool, on improving speaking skills among language learners. Drawing on the psycholinguistic theory proposed by Huettig et al. (2022), which suggests that speakers of second or foreign languages must develop cognitive abilities to generate languages, predict and anticipate the flow of conversation, and employ social conversational skills, this research explores the potential of AI tools to enhance these competencies. The theory supports the use of AI-assisted language tools in teaching and assessing speaking skills, given that traditional methods for speaking improvement often require extensive time and effort.

A quasi-experimental design was employed to collect and analyze data. The findings revealed a significant effect of the ELSA speech analyzer application on students' speaking skill development. The experimental group showed substantial improvement, with mean scores rising from 12.6 out of 20 in the pre-assessment to 18.17 out of 20 in the post-assessment. Additionally, students exhibited marked progress in pronunciation, intonation, and vocabulary, with mean improvements of 3.84, 3.61, and 3.557, respectively, whereas grammar showed comparatively less improvement. These findings suggest that the manipulation played a significant role in enhancing students' speaking skills. The manipulation involved providing students with detailed information about their current level of proficiency, as determined by internationally recognized English language assessment tests such as the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL). This information motivated students to regularly utilize the application, as it offered a clear indication of their proficiency level based on a highly regarded assessment test.

Furthermore, the application provided students with tailored activities related to the topics they were discussing. It also offered a series of materials for English exam preparation, aimed at improving students' vocabulary, fluency, and accuracy. As a result, students' confidence in dealing with speaking exams grew. The application also facilitated engagement in interactive activities with an artificial intelligence system in a stress-free environment. This allowed students to feel at ease while speaking, recording their responses, and receiving feedback on their pronunciation. The feedback included specific information on mispronounced letters and targeted videos to address specific areas of concern, such as intonation, fluency, and word choice. The application also provided extra practice exercises to help students improve in these areas.

In contrast, students in the control group were instructed to speak about the same topics as the experimental group students but were only provided with a rubric to assess their performance level. The rubric did not provide specific guidance on how to improve pronunciation, intonation, fluency, grammar, or vocabulary. Additionally, students in the control group felt a sense of apprehension due to the fear of making mistakes. Consequently, they did not engage in sufficient practice, despite being provided with the rubric in advance to familiarize themselves with the assessment standards.

These findings are consistent with previous studies conducted by Isaacs (2018) and Gu et al. (2021), which demonstrated the effectiveness of automatic speech analysis in enhancing speaking skills. Additionally, this study aligns with the research of Zou et al. (2024), which highlighted the benefits of AI-assisted language learning tools, such as the ELSA speech analyzer application, in providing personalized feedback, adaptive learning experiences, and interactive assessments to improve speaking skills. Hana et al. (2022) compared ELSA to Duolingo and found ELSA to be superior due to its ability to make learning enjoyable, adaptability to curriculum, and suitability for students of all speaking levels. Mohammadkarimi (2024) found that ELSA was effective in improving pronunciation and saving time by providing detailed feedback. Furthermore, the studies conducted by Chen (2024), Kim et al. (2021), Qiao and Zhao (2023), Alwin et al. (2024), and Al-shallakh (2024) demonstrated the significant impact of AI tools on the development of students' speaking skills. What sets this study apart from previous ones is its focus on comparing the effect of ELSA with traditional rubric-based methods for teaching and assessing students' speaking skills.

Unlike previous studies, which only focused on one or two standards, such as Al-shallakh (2024), and Wu (2024), this study examines all speaking skills standards. The selection of university students as participants addresses challenges identified in previous research, such as the failure to recognize the sounds of EFL speakers by the application (Zou et al., 2024; Kim et al., 2021), which leads to inauthentic evaluation. By exposing this sample of students to the nature of the program and providing them with opportunities for practice, the study allows for constructive feedback that accurately reflects their performance. Additionally, the students are not discouraged when they receive low grades, as they have

the ability to repeat the assessments multiple times. The focus of this study is on learning and improvement, rather than grades alone. Using an AI-assisted language learning program enables students to access all of their recordings, allowing them to track their progress in each standard and practice speaking continuously.

As a result, this study provides evidence that students in foreign or second language contexts can benefit from AI-assisted language learning in the same way native English speakers do. Given the high sensitivity of English language accents, EFL students have the potential to speak like native speakers if they practice and address their weaknesses.

These significant results should serve as motivation for EFL tutors to familiarize themselves with various AI-assisted language tools. Such tools can support tutors in efficiently and effortlessly conducting speaking skills assessments on a regular basis. Furthermore, tutors can encourage and guide EFL students to engage in independent practice using AI tools during their free time. This approach reduces the time-consuming of conducting speaking skill assessments in the classroom, while also minimizing subjectivity in assessment. These platforms offer EFL teachers valuable constructive feedback based on standardized assessments. With the assistance of AI platforms, teachers can implement remedial programs to address students' weaknesses and focus on improving specific speaking problems. In summary, the findings of this study have important implications for both EFL teachers and students, highlighting the potential impact of AI platforms on language teaching and learning.

The study recommends that EFL students embrace AI platforms as a valuable resource to motivate their development of speaking skills. It is essential to strategically incorporate these platforms within the course description, rather than treating them as optional. These engaging and motivating platforms empower students to learn, providing continuous feedback without causing any anxiety or concern about low performance. This allows students to feel in control of their learning and fosters a sense of autonomy. AI has proved to be a valuable tool for enhancing speaking skills, especially in the areas of speech recognition and pronunciation feedback.

AI-powered speech recognition technology accurately analyzes learners' spoken English, offering instant feedback on pronunciation accuracy, intonation, fluency, vocabulary, and grammar. Through identifying specific errors, learners receive individualized feedback and insights on how to improve. This real-time feedback increases learners' awareness of their speaking performance and helps them make the necessary adjustments to sound more natural and fluent. Similarly, EFL tutors can incorporate AI platforms into their instructional process. By using the data provided by these tools, tutors can effectively tailor their instruction to address specific language learning challenges, differentiating their teaching strategies to meet the diverse needs of their students providing that students are aware of how these tools could be used appropriately and specifically outside classrooms (Broadhead, 2024) and at the same time, tutors and teachers employ it as an assistant and try to preserve the students' critical thinking skills while developing any skill

It can be deduced that integrating AI into language teaching and learning processes opens up opportunities for innovation in teaching and assessment. This integration has the potential to help students achieve desired learning outcomes. Further research could explore the effects of AI tools in different contexts, with varied samples and skills. Additionally, studies could focus on the long-term impact of AI on students' learning and the effects of professional development programs that incorporate AI on tutors' proficiency in using these tools.

Conclusion

The present study examined the impact of the ELSA speech analyzer application on the speaking abilities of English as a Foreign Language (EFL) students. The results indicated that the use of the ELSA speech analyzer application significantly improved the speaking skills of the EFL students. Specifically, the experimental group demonstrated superior performance compared to the control group in terms of pronunciation, intonation, vocabulary, and grammar.

The students found the AI application to be engaging and motivating, as it allowed them to conveniently practice speaking topics and integrate it with their course materials. Consequently, teachers were able to allocate more time to teach essential aspects of the language due to the accurate feedback and analysis provided by the AI platform. This facilitated both students and teachers in identifying areas for improvement and addressing them. The study offers valuable insights into the integration of AI-assisted language learning tools to enhance language programs with the guidance of tutors.

Despite the study's significant contribution to the field of technology-supported English language instruction, particularly regarding the development of speaking skills, as well as its potential novelty within the region, it is important to recognize the limitations associated with the interpretation of the results. Notably, the study utilized a quasi-experimental design, which may constrain the generalizability of the findings to broader contexts.

To enhance external validity, future research should include a more diverse sample of EFL students, encompassing individuals from various disciplines and educators. By doing so, researchers can gain a comprehensive understanding of the issues addressed in the study and facilitate a more thorough analysis. Furthermore, a recent study conducted by Ogunleye et al. (2024) cautioned against the inappropriate use of artificial intelligence (AI) bots in higher education, as their misuse may impede the learning process. Specifically, AI applications should be employed to enhance speaking skills rather than for assessment purposes. When evaluating non-native speakers with accents, AI systems may render biased evaluations, potentially demotivating students from pursuing further language acquisition. Additionally, Al-Awawdeh et al. (2023) noted the limited integration of AI in higher education, particularly in the context of teaching English as a foreign language in course descriptions as well

as in the learning management system (LMS). This situation places an additional burden on universities in the region to provide professional development for academic staff on the effective integration of AI-assisted language learning applications in both teaching and assessment processes, as these tools have become integral to the educational landscape in the current era.

It can be highlighted in this context that the findings of this study hold considerable significance for the field of language acquisition. That is, the results reinforce prominent theories like the psycholinguistic theory that emphasizes generating languages, predicts and anticipates the flow of conversation, and employs social conversations. The effect of AI-powered tools, such as ELSA, in developing speaking skills calls for the development of new theoretical frameworks that account for the role of AI in language learning.

References

- Al-Awawdeh, N., Al-shaboul, I. A., & Khasawneh, M. A. S. (2023). Advancing foreign language teaching with AI-assisted models; insights from lecturers and university administrators. *Journal of Namibian Studies: History Politics Culture*, 33, 1491-1506. <https://doi.org/10.59670/jns.v33i.798>
- Al-Shallakh, M. A. I. (2024). Embedding artificial intelligent applications in higher educational institutions to improve students' pronunciation performance. *Theory and Practice in Language Studies*, 14(6), 1897-1906. <https://doi.org/10.17507/tpls.1406.31>
- Alwin, W. R., Manurung, K., Lebagi, D., & Mukrim, M. (2024). The use of English Language Speech Assistant (ELSA) speak application to improve pronunciation. *Jurnal Onoma: Pendidikan, Bahasa, dan Sastra*, 10(3), 2880-2887. <https://doi.org/10.30605/onoma.v10i3.3926>
- Anguera, X., Proença, J., Gulordava, K., Tarján, B., Parslow, N., Dobrovolskii, V., & Girard, R. (2023). ELSA speech analyzer: English communication assessment of spontaneous speech. *Proceedings of 9th workshop on Speech and Language Technology in Education (SLaTE)* (pp. 95-96)
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, 2(17). <https://doi.org/10.3390/app12178427>
- Broadhead, L. A. (2024). Insidious chatter versus critical thinking: Resisting the Eurocentric siren song of AI in the classroom. *Journal of Applied Learning and Teaching*, 7(2), 28-37. <https://doi.org/10.37074/jalt.2024.7.2.9>
- Dai, Y., & Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, 36(5-6), 861-884. <https://doi.org/10.1080/09588221.2021.1952272>
- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment &*

Evaluation in Higher Education, 42(3), 347-360. <https://doi.org/10.1080/02602938.2015.1111294>

De Jong, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, 9(1), 541-560. <https://doi.org/10.1146/annurev-linguistics-030521-052114>

Fan, J., & Yan, X. (2020). Assessing speaking proficiency: a narrative review of speaking assessment research within the argument-based validation framework. *Frontiers in Psychology*, 11, 330. <https://doi.org/10.3389/fpsyg.2020.00330>

Fulcher, G. (2015). Assessing second language speaking. *Language Teaching*, 48(2), 198-216. <https://doi.org/10.1017/S0261444814000391>

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). <https://doi.org/10.37074/jalt.2023.6.2.32>

Gu, L., Davis, L., Tao, J., & Zechner, K. (2021). Using spoken language technology for generating feedback to prepare for the TOEFL iBT® test: A user perception study. *Assessment in Education: Principles, Policy & Practice*, 28(1), 58-76. <https://doi.org/10.1080/0969594X.2020.1735995>

Hanna, A. N., Harmayanthi, V. Y., & Astuti, S. (2022). The effect of ELSA speak app towards students' speaking skill. *Proceeding of international conference on education-01, 01* (pp. 16-20). <https://doi.org/10.37640/ice.01.240>

Huetig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, 224, 105050. <https://doi.org/10.1016/j.cognition.2022.105050>

Isaacs, T. (2016). Assessing speaking. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 131-46). Berlin: DeGruyter Mouton. <https://doi.org/10.1515/9781614513827-011>

Jayasinghe, S. (2024). Promoting active learning with ChatGPT: A constructivist approach in Sri Lankan higher education. *Journal of Applied Learning and Teaching*, 7(2), 141-154. <https://doi.org/10.37074/jalt.2024.7.2.26>

Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, 206, 104898. <https://doi.org/10.1016/j.compedu.2023.104898>

Kim, H. S., N, Kim & Y, Cha (2021). Is it beneficial to use AI chatbots to improve learners' speaking performance?. *Journal of Asia TEFL*, 18(1), 161-178. <http://dx.doi.org/10.18823/asiatefl.2021.18.1.10.161>

Kholis, A. (2021). Elsa speak app: Automatic speech recognition (ASR) for supplementing English pronunciation skills. *Pedagogy: Journal of English Language Teaching*, 9(1),

01-14. <https://doi.org/10.31849/elsya.v6i1.18727>

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. (2024). Exploring the use of artificial intelligence in promoting English language pronunciation skills. *LLT Journal: A Journal on Language and Language Teaching*, 27(1), 98-115. <https://doi.org/10.24071/llt.v27i1.8151>

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). Higher education assessment practice in the era of generative AI tools. *Journal of Applied Learning and Teaching*, 7(1), 46-56. <https://doi.org/10.37074/jalt.2024.7.1.28>

Popenici, S. (2023). 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>

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL context. *Frontiers in Psychology*, 14, 1255594. <https://doi.org/10.3389/fpsyg.2023.1255594>

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>

Tagg, C. (2020). English language and social media. In S. Adolphs & D. Knight (Eds.), *The Routledge handbook of English language and digital humanities* (pp. 568-586). Routledge.

Tai, T. Y., & Chen, H. H. J. (2024). Navigating elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, 105112. <https://doi.org/10.1016/j.compedu.2024.105112>

Tan, S. (2020). Artificial Intelligence in education: Rise of the Machines. *Journal of Applied Learning and Teaching*, 3(1), 129-133. <https://doi.org/10.37074/jalt.2020.3.1.17>

Wang, W., Lin, N., Oberhaus, J. D., & Avidan, M. S. (2020). Assessing method agreement for paired repeated binary measurements administered by multiple raters. *Statistics in Medicine*, 39(3), 279-293. <https://doi.org/10.1002/sim.8398>

Winaryo, S., & Gusdian, R. I. (2021). Measuring English language education department students' speaking fluency

level. *Wanastra: Jurnal Bahasa dan Sastra*, 13(1), 13-18. <https://doi.org/10.31294/w.v13i1.9812>

Wu, X. Y. (2024). Exploring the effects of digital technology on deep learning: A meta-analysis. *Education and Information Technologies*, 29(1), 425-458. <https://doi.org/10.1007/s10639-023-12307-1>

Zainuddin, N. M., & Mohamad, M. (2024). Utilising Speechace to enhance speaking skills among English as a second language pre-university students. *International Journal of Academic Research in Progressive Education and Development*, 13(2), 1206 – 1219. <http://dx.doi.org/10.6007/IJARPED/v13-i2/21458>

Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *Sage Open*, 13(3), 1-9. <https://doi.org/10.1177/21582440231193818>

Zou, B., Liviero, S., Hao, M., & Wei, C. (2020). Artificial intelligence technology for EAP speaking skills: Student perceptions of opportunities and challenges. In M. R. Freiermuth & N. Zarrinabadi (Eds.), *Technology and the psychology of second language learners and users. new language learning and teaching environments*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-34212-8_17