Harnessing AI for Pronunciation Development: Investigating the Impact of Elsa Speak on EFL Learners' Oral Proficiency

Abbas Hussein Abdelrady[1], Huma Akram[2]*, Abeer Taha Abdelhafeez Mohamed[3], Ghadah Faisal Abaalkhail[4], Ehlam Hussein Elsadig Osman[5], Gadaa Taha Abdal Hafeez Mohamed[6], Wijdan Mohieldeen Mohammed Suliman[1]

[1] Department of English Language & Literature, College of Languages & Humanities, Qassim University, Saudi Arabia. E-mail: ab.altaher@qu.edu.sa, E-mail: w.suliman@qu.edu.sa

[2] School of International Education, North China University of Water Resources and Electric Power, Zhengzhou, Henan, China.

[3] English Department, Applied College, King Khalid University, Saudi Arabia. E-mail: atmohamed@kku.edu.sa

[4] English Department, College of Science and Health Professions, King Saud bin Abdulaziz University for Health Sciences, Saudi Arabia. E-mail: alkhail@ksau-hs.edu.sa

[5] Department of Information Science, College of Arts, Imam Abdulrahman Bin Faisal University, Saudi Arabia.

E-mail: ehosman@iau.edu.sa

[6] Department of Sociology and Social Work, College of Arts, Imam Abdulrahman Bin Faisal University, Saudi Arabia.

E-mail: gtabdulhafiz@iau.edu.sa

Correspondence*: Huma Akram, School of International Education, North China University of Water Resources and Electric Power, Zhengzhou, Henan, China. E-mail: akramhuma@ncwu.edu.cn

Received: October 20, 2025 Accepted: March 4, 2026 Online Published: April 10, 2026

doi:10.5430/wjel.v16n4p311 URL: https://doi.org/10.5430/wjel.v16n4p311

Abstract

Pronunciation plays a pivotal role in oral fluency and overall communicative effectiveness in English as a Foreign Language (EFL) learning. However, pronunciation instruction remains a persistent challenge in higher education contexts, particularly where learners have limited opportunities for individualized feedback and sustained oral practice. Responding to recent advances in artificial intelligence (AI), this study investigated the effectiveness of the AI-powered ELSA Speak application in enhancing EFL learners' pronunciation development. Adopting a longitudinal experimental design, the study involved 60 university-level EFL students, equally divided into experimental and control groups. Learners' pronunciation performance was assessed at three time points using the CAFIS analytic framework. Repeated-measures analyses revealed that students using ELSA Speak demonstrated significantly greater improvement across all CAFIS dimensions compared to the control group. Accuracy showed the most immediate gains, followed by gradual but consistent development in intonation and fluency over time. These findings underscore the value of AI-assisted pronunciation instruction and highlight the importance of adopting multidimensional assessment frameworks to capture nuanced developmental changes. The study contributes empirical evidence from a Saudi higher education context and offers pedagogical insights into the integration of AI-driven tools for pronunciation development in EFL classrooms.

Keywords: Technology integration, artificial intelligence, EFL learners, Elsa Speak, pronunciation development, Language learning

1. Introduction

Pronunciation is a key element in effective oral communication and is widely regarded as a integral part of oral fluency in English (Abdelhalim & Alsehibany, 2025). For EFL learners, it serves as a bridge between linguistic knowledge and effective real-world interaction. Pronouncing well not only enhances listener's comprehension but also foster communicative effectiveness (Al-Shallakh, 2024). Correspondingly, poor pronunciation not only hinders mutual understanding but also erodes confidence in speaking (Aulia & Santosa, 2025), leading to avoidance of oral interaction and stagnation in overall language proficiency (Metruk, 2024). In higher education, where oral communication is integral to classroom discussions or presentations, mastering English pronunciation becomes even more critical (Li & Akram, 2023, 2024). It empowers students to participate actively, articulate their ideas clearly, and compete globally in post-graduation careers (Aryanti & Santosa, 2024). Yet, pronunciation is often overshadowed in many EFL curricula, leaving learners with unaddressed gaps in oral fluency.

Meanwhile, the integration of artificial intelligence (AI) into language settings has grasped substantial consideration (Abdelrady et al., 2025). AI-powered applications have been shown to by facilitating learners' engagement (Akram & Li, 2024), supporting their autonomy (Al-Adwan et al., 2022; Sohail & Akram, 2025), shaping their learning preferences (Ma et al., 2024), bringing creativity (Lin & Chen, 2024), providing immediate and personalized feedback (Akram & Abdelrady, 2023, 2025), and making learning environments dynamic

(Akram et al., 2021). In terms of pronunciation learning, AI technologies enable learners to receive instant evaluations of their speech (Dennis, 2024), compare their production with native-like models (Metruk, 2024), and engage in repeated, self-paced practice (Ou et al., 2024). Such features address many of the limitations of traditional pronunciation instruction and align well with the needs of EFL learners in higher education.

One such AI-powered application gaining popularity in EFL contexts is ELSA Speak. It is prominent for its focus on improving learners' spoken English through automated speech recognition and real-time feedback (Arbain et al., 2023). Utilizing speech recognition technology, it evaluates learners' pronunciation against native speaker models and delivers immediate scores (Pham & Pham, 2025). Its user-friendly interface and gamified elements make it particularly suitable for EFL learners seeking to improve oral fluency (Rusmawaty et al., 2024). Although previous studies report positive effects of AI-assisted tools on EFL learners' pronunciation training, there is a clear gap concerning the use of AI-powered pronunciation applications in the Saudi Arabian higher education context. EFL learners often face pronunciation challenges due to limited English speaking outside the classroom in Saudi context (Al-Bogami & Alahmadi, 2025). It provides them less opportunities for authentic spoken interaction. Despite the growing adoption of educational technologies in higher education, little studies have examined the effectiveness of AI-driven pronunciation tools, specifically, ELSA Speak. In light of this, the present study aims to investigate these research objectives:

1. To examine the role of the ELSA Speak app in enhancing EFL students' pronunciation.
2. To steer a longitudinal assessment of students' pronunciation performance at various intervals of time

2. Literature Review

Several research studies highlight the transformative role of AI in language learning, particularly in the domain of pronunciation development for EFL learners. AI-powered applications such as Elsa Speak have gained attention for their ability to provide real-time, individualized feedback, which traditional classroom instruction often lacks. Aryanti and Santosa (2024) emphasize that AI tools like Elsa Speak create immersive and adaptive learning environments. Their systematic review found that such tools significantly enhance learners' pronunciation skills by offering consistent and targeted feedback. This identification aligns with Godwin-Jones' (2018) opinion, who notes that mobile technologies can extend learning beyond the classroom, supporting autonomous and personalized learning practice. In another study, Sholekhah and Fakhrurriana (2023) highlights the dynamic role of ELSA Speak, revealing the role of gamified elements in securing interactive learning environment by boosting learners' motivation. Aligning this, Kholis's (2021) observation illustrates that this app promotes autonomous learning by allowing students to practice independently, thereby reducing their reliance on teacher feedback and encouraging self-directed learning.

Building on this, Aulia and Santosa (2025) conducted a PRISMA-based review that confirms Elsa Speak's effectiveness in improving pronunciation skills among learners. They attribute these gains to the app's interactive features, including gamification, progress tracking, and speech modeling. Learners also reported increased motivation and engagement, suggesting that AI tools may foster sustained language practice. Comparative studies further validate Elsa Speak's pedagogical value. For instance, Arbain et al. (2023) specified that learners using Elsa Speak outperformed those using non-AI tools in pronunciation assessments. While traditional tools may support vocabulary acquisition, they lack the nuanced phonological feedback necessary for pronunciation development. In a similar vein, a study conducted by Al-Shallakh (2024) at a Jordanian university reported significant enhancement in students' pronunciation after a seven-week intervention with this app. Another study by Karim et al. (2023) found a 17% improvement in speaking ability, underscoring the app's effectiveness in fostering both fluency and accuracy. Moreover, the app's ability to simulate real-life communication scenarios enables learners to practice and enhance their conversational skills, which is particularly beneficial for less proficient students (Rusmawaty et al., 2024).

Despite the growing interest in AI-powered pronunciation tools like Elsa Speak, several critical gaps remain in the existing literature. Most studies emphasize short-term improvements, leaving limited evidence on whether pronunciation gains are retained over time or effectively transferred to spontaneous speech. Furthermore, many studies lack a comprehensive theoretical framework that integrates technological, and behavioral perspectives, which is essential for a holistic understanding of AI's role in language learning. Therefore, a balanced approach that combines AI tools with traditional teaching methods may be most effective in enhancing EFL learners' oral proficiency.

2.1 Theoretical Framework

Pronunciation is a complex, multidimensional construct that underpins second language (L2) oral communicative competence. Unlike other linguistic skills, pronunciation relies on the integration of segmental, suprasegmental, and psycholinguistic processes (Pennington, 2021). To systematically examine how the ELSA Speak app influences pronunciation development, this study adopts the Comprehensibility, Accuracy, Fluency, and Intonation Rating Scale (CAFIS) as its theoretical framework. It is adapted from Derwing and Munro's (2005) holistic pronunciation model, which provides a validated, multidimensional lens to evaluate three core dimensions of pronunciation:

2.1.1 Accuracy

Accuracy refers to learners' control over segmental features, including the production of vowels, consonants, and syllable structures that approximate target-language norms. Its theoretical basis lies in phonetic (Mora-Plaza et al., 2024) and phonological models of second

language acquisition (Abreu & Gathercole, 2012), which emphasize the role of perceptual discrimination and articulatory practice in forming new sound categories. Accurate segmental production is fundamental to intelligibility, as mispronunciations at the segmental level can lead to misunderstanding or communication breakdowns (Derwing & Munro, 2015). Building on this, Gordon and Darcy (2022) suggest that segmental accuracy is often the first pronunciation component to show improvement following focused instruction, particularly when learners receive explicit feedback on sound-level errors. However, accuracy alone does not guarantee natural or fluent speech, underscoring the need to examine additional pronunciation dimensions.

2.1.2 Intonation

Intonation represents the suprasegmental dimension of pronunciation and encompasses pitch movement, stress patterns, and rhythmic organization (Jenkins, 2004). Its theoretical basis lies in prosodic phonology (Hammond, 2020) and communicative intent theory (Fetzer, 2008). These perspectives highlight the role of intonation in signaling meaning, discourse structure, and pragmatic intent (Dahmen et al., 2023). In English, intonation patterns contribute to the distinction between statements and questions, the expression of emphasis, and the management of conversational flow (Liu et al., 2025). For EFL learners, intonation is often challenging due to cross-linguistic influence and limited exposure to authentic spoken input (Adams-Goertel, 2013). Expanding on this, Zuhairya et al. (2024) assert that inappropriate intonation can negatively affect perceived comprehensibility and speaker confidence, even when segmental accuracy is relatively high. As a result, intonation is considered a crucial yet gradually developing component of pronunciation competence.

2.1.3 Fluency

Fluency refers to the temporal dimension of speech, including speech rate, smoothness, and pausing behavior (Derwing & Munro, 2015). Its theoretical basis is rooted in automaticity theory (Suzuki et al., 2025) and cognitive load theory (Bóna & Bakti, 2020). Automaticity theory argues that fluency develops as learners transition from controlled to automatic processing of linguistic forms. The reputation of the practice reduces the cognitive resources needed for production, which leads to smoother speech. While cognitive load theory adds that focusing on accuracy can disrupt fluency, creating a "trade-off" that complicates L2 development. For EFL learners, this trade-off is pronounced in contexts with limited oral practice, as cognitive resources are diverted to monitoring accuracy rather than maintaining flow (Terzioğlu & Kurt, 2022). CAFIS captures this balance by evaluating fluency alongside accuracy and intonation, providing a holistic measure of pronunciation development (Derwing & Munro, 2005).

2.2 ELSA Speak and CAFIS Dimensions

Recent advances in AI technologies have enabled the development of pronunciation tools that align closely with the multidimensional nature of pronunciation competence (Akram et al., 2022; Jalalzai et al., 2025; Sholekhah & Fakhrurriana, 2023). The ELSA Speak application employs automated speech recognition and AI-driven feedback to address each CAFIS dimension systematically. At the segmental level, the app provides immediate feedback on sound accuracy (Liunokas, 2025), allowing learners to identify and correct specific pronunciation errors. This targeted feedback supports the development of segmental accuracy through repeated, individualized practice (Karim et al., 2023). In terms of intonation, ELSA Speak exposes learners to native-like speech models and evaluates pitch variation and stress patterns, encouraging more natural prosodic production (Kholis, 2021). In addition, its visual and auditory feedback features help learners notice deviations from target intonation patterns, thereby facilitating suprasegmental development (Sholekhah & Fakhrurriana, 2023). With respect to fluency, this application not only promotes continuous speech production through structured speaking tasks and repeated practice but also help in reducing hesitation and unnatural pauses (Liunokas, 2025). By integrating these features, ELSA Speak operationalizes pronunciation training in a manner that aligns with the CAFIS framework and proposes the following hypotheses:

1. ELSA Speak application significantly enhances EFL learners' pronunciation accuracy overtime.
2. ELSA Speak application significantly enhances EFL learners' pronunciation intonation overtime.
3. ELSA Speak application significantly enhances EFL learners' pronunciation fluency overtime.
4. ELSA Speak application significantly enhances pronunciation (across CAFIS dimensions) over time for students in the experimental group.

3. Methodology

This study adopted a longitudinal quasi-experimental design to investigate the role of the ELSA Speak application in enhancing EFL students' pronunciation over a 12-week intervention. Following a non-equivalent group design, 60 second-year English program students from a public university of Saudi Arabia were selected through a convenience sampling and assigned to two groups (experimental: n=30; control: n=30) with random distribution prior to the course onset. All participants shared a relatively homogeneous linguistic background and had no prior experience of using the ELSA Speak App for English learning. They were further compared on baseline English proficiency, assessed via pre-test CAFIS scores, revealing no statistically significant differences ($p > 0.05$) between groups, confirming initial equivalence (Twisk & Proper, 2004). Furthermore, the study followed all ethical requirements set out by the World Medical Association's (2013) declaration of Helsinki, which are widely acknowledged standards for research involving human subjects. Research participants were further adequately informed about the research's purpose, methodology, and the academic evaluation of their audio samples. Their participation was voluntary, and informed consent was obtained prior to data collection. They were also advised that they might withdraw from the study at any time without academic or personal consequences. In a similar vein, coded identities were assigned

to each participant to ensure their anonymity, and no personally identifiable information was collected or shared when divulging the results.

3.1 Procedure

Both groups received identical 90-minute weekly in-class pronunciation instruction, focusing English word/sentence stress, rhythmic pacing, and intonation contours, led by the same trained EFL instructor. The only variable difference was the mode of 30-minute daily out-of-class practice, as detailed below:

Experimental Group

At first, the students in the experimental group received pronunciation training through the ELSA Speak app. Subsequently, they were made engaged with ELSA Speak app with 30 minutes per day throughout the entire intervention. The app provided learners with individualized, CAFIS-aligned tasks targeting: (1) **accuracy** (segmental features, consonants, vowel production, and stress placement); (2) **intonation** (suprasegmental aspects, pitch variation, sentence-level prosody, and rhythm); and (3) **fluency** (pacing guidance, pause pattern reduction). Learners received real-time audio-visual AI-driven feedback to identify pronunciation errors and make adjustments in real time and completed tasks sequenced from simple to complex pronunciation activities to progress pronunciation acquisition systematically. Additionally, regular practice schedules and progress indicators within the app encouraged its sustained engagement. In particular, ELSA usage was objectively tracked via the platform's instructor admin dashboard, which captured real-time metrics including daily practice completion, total session minutes, module mastery rates, and weekly engagement frequency. Average compliance across the 12 weeks was 94% (no attrition); participants completed a mean of 30.2 minutes of daily practice (SD = 2.1) and a total of 2174 cumulative practice minutes (SD = 148) over the intervention period. All ELSA tasks were aligned with weekly in-class pronunciation topics to ensure curricular consistency.

Control Group

In contrast, the control group received traditional pronunciation instruction through teacher-led activities without the support of mobile-assisted pronunciation technology. It included repetition drills (for segmental/suprasegmental features) oral reading tasks of EFL textbook passages (Headway Intermediate, 5th Ed.), and written pronunciation worksheets (error correction and pattern recognition). Practice compliance was tracked via weekly worksheet submission and signed practice logs (average compliance: 92%), with the same instructor providing written feedback on completed practice materials (no additional oral feedback beyond in-class sessions).

3.2 Pronunciation Assessment

Students' pronunciation performance of both groups was assessed at the beginning (week 1), mid (week 6), and end (week 12) using the CAFIS' criteria prescribed by Derwing and Munro (2005). From 1 (very poor) to 9 (outstanding or native-like), this scale offers a comprehensive range of rankings, focusing on three core dimensions, accuracy, intonation, and fluency.

3.3 Data Collection Procedure

At each measurement point, participants from both groups undertook a standardized speaking task that included three distinct oral tasks aimed at eliciting comparable oral production for CAFIS rating. Tasks were designed to capture both controlled and spontaneous speech, as follows:

1. **Reading Aloud:** A 100-word passage from a standard EFL textbook (carefully selected for a balanced range of vowels, consonants, and intonation patterns) read at a natural pace.
2. **Picture Description:** A 1–2 minute description of three connected daily-life scenario pictures with no prior preparation.
3. **Free Speaking:** A 1–2 minute spontaneous response to the general topic *"My English Learning Experience"* with no prior preparation.

All speech samples were audio-recorded and anonymized (coded with participant ID only), and stored securely for subsequent CAFIS rating.

3.4 Rating Procedure and Reliability

Three trained EFL instructors (with ≥5 years of pronunciation teaching experience) served as raters for the audio-recorded speech samples. Prior to formal scoring, raters completed **two 90-minute calibration sessions** using a subset of anonymized pilot speech samples (not included in the study dataset) to establish a shared understanding of CAFIS descriptors for each dimension (accuracy, intonation, fluency). Raters scored all samples independently (blind to group assignment and measurement time point) using the CAFIS scoring rubric. **Inter-rater reliability** was calculated for each CAFIS dimension using **two-way mixed intraclass correlation coefficients (ICC)** (Cole, 2024); all ICC values exceeded 0.82 (accuracy: ICC=0.87; intonation: ICC=0.85; fluency: ICC=0.84), indicating excellent inter-rater consistency across raters and measurement time points. Discrepancies in scores (≤1 point on the 9-point scale) were resolved via group consensus among raters.

3.5 Data Analysis

The data were analyzed with SPSS version 26.0. First, descriptive analysis (mean, standard deviation) was run for each CAFIS dimension across each time point for both groups to summarize performance trends. Subsequently, two-way repeated measures ANOVAs were

employed to determine whether observed differences across pre-, mid-, and post-intervention stages were statistically meaningful. This analytic strategy allowed for a clear evaluation of the extent to which sustained engagement with the ELSA Speak app contributed to learners' pronunciation development over time.

4. Results

This section presents the findings related to the impact of the ELSA Speak application on EFL students' pronunciation development across three time points: pre-intervention, mid-intervention, and post-intervention.

4.1 Descriptive Statistics

Descriptive information (see Table 1) displays the mean and standard deviation values for pronunciation accuracy, intonation, and fluency for both groups at each time point. At the pre-test stage, the experimental and control groups demonstrated comparable levels of pronunciation performance across all three dimensions, indicating baseline equivalence. Over time, the experimental group showed a steady increase in mean scores across all pronunciation dimensions,

whereas the control group showed relatively limited progress. Additionally, error bars (see Figure 1) indicate consistent variability in scores, and the gap between groups widened at mid-test and post-test, confirming the intervention's sustained positive effect.

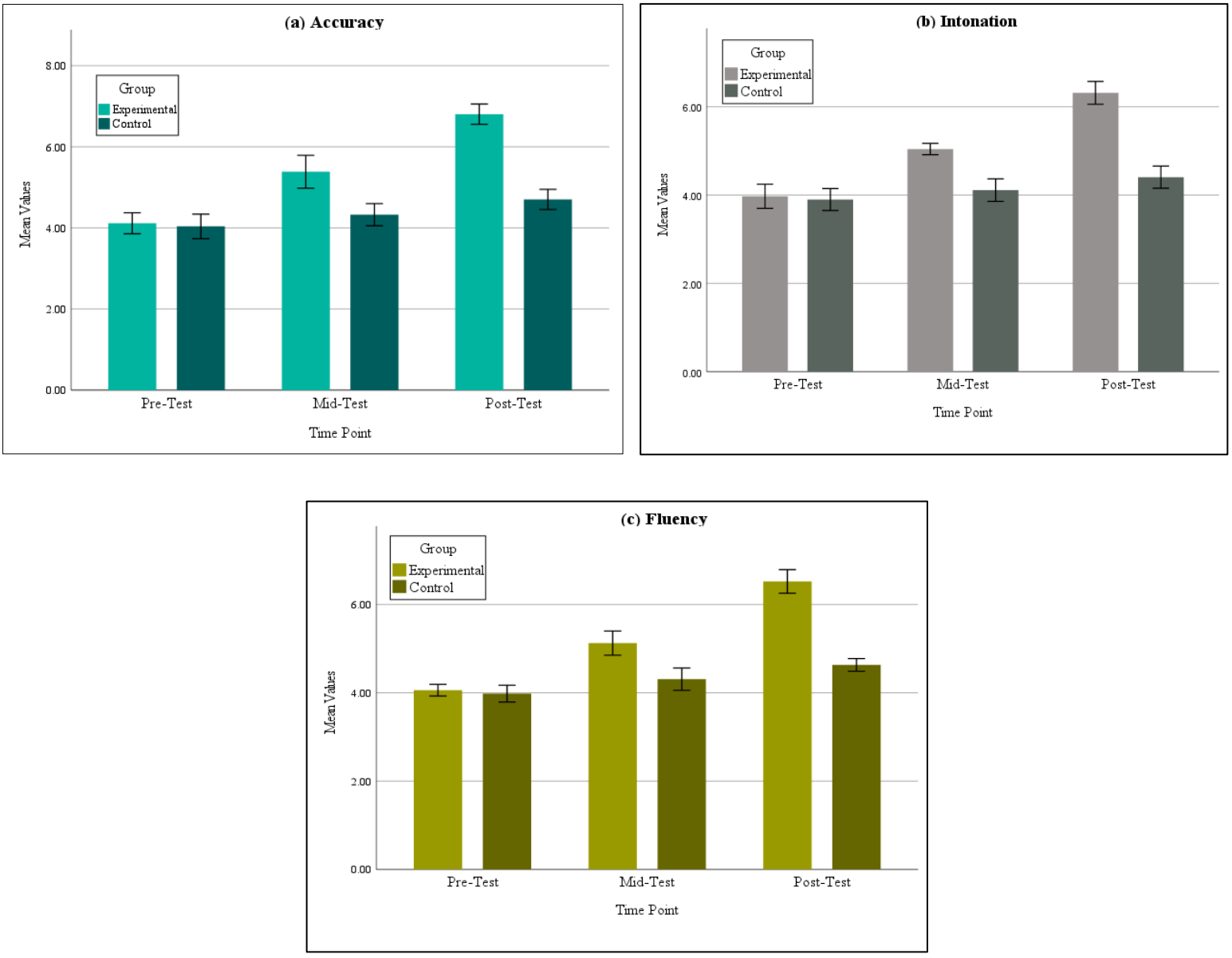

Figure 1. Mean comparison across CAFIS dimensions over time

Table 1. Descriptive results

Dimension	Group	Pre-test	Mid-test	Post-test
		M (SD)	M (SD)	M (SD)
Accuracy	Experimental	4.15 (0.84)	5.46 (0.78)	6.82 (0.70)
	Control	4.12 (0.81)	4.38 (0.79)	4.71 (0.76)
Intonation	Experimental	3.92 (0.89)	5.03 (0.83)	6.35 (0.75)
	Control	3.90 (0.86)	4.14 (0.82)	4.42 (0.80)
Fluency	Experimental	4.08 (0.87)	5.18 (0.80)	6.57 (0.72)
	Control	4.05 (0.85)	4.33 (0.81)	4.60 (0.77)

4.2 Two-Way Repeated Measures ANOVA

To examine the effect of ELSA Speak app on students' pronunciation over time, prior to conducting inferential analyses, assumptions of normality and homogeneity of variance were checked and met for all CAFIS dimensions. Two-way repeated measures ANOVA was then conducted for each CAFIS pronunciation dimension (accuracy, intonation, fluency) (see Table 2). Mauchly's Test of Sphericity was further used to verify the critical sphericity assumption for the within-subjects time factor and Group × Time interaction, and no violations of sphericity were detected, thus no corrective adjustments (e.g., Greenhouse-Geisser) were required for degrees of freedom or p-values. The results revealed a significant impact of time on all three pronunciation dimensions, which supports hypotheses 1, 2, and 3. In parallel, the interaction of groups with time was also found significant, which indicates that pronunciation improvement differed significantly between the experimental and control groups over time, which supports hypothesis 4. It shows that students in the experimental group experienced significantly greater pronunciation gains over time compared to those in the control group. Regarding strength, the effect of ELSA Speak app was strongest for Accuracy dimension $F(2, 116) = 36.58$, $p < 0.001$, $\eta^2 = 0.39$, with a mean difference of 2.67 (95% CI), whereas Intonation demonstrated comparatively smaller improvement then other, i.e., $F(2, 116) = 33.47$, $p = < 0.001$, $\eta^2 = 0.37$, with a mean difference 2.43.

Table 2. Repeated measures ANOVA results

Dimension	Effect	F	df	*p*-value	Partial η^2
Accuracy	Time	52.41	2, 116	< .001	.47
	Group	18.76	1, 58	< .001	.24
	Group × Time	36.58	2, 116	< .001	.39
Intonation	Time	48.29	2, 116	< .001	.45
	Group	16.94	1, 58	< .001	.23
	Group × Time	33.47	2, 116	< .001	.37
Fluency	Time	50.13	2, 116	< .001	.46
	Group	17.62	1, 58	< .001	.24
	Group × Time	35.89	2, 116	< .001	.38

4.2.3 Post Hoc Comparisons

Post hoc analyses with Bonferroni adjustments were conducted (see Table 3) to identify specific differences between time points. These revealed that the experimental group demonstrated significant improvements across all three pronunciation dimensions from pre-test to mid-test and from mid-test to post-test. In contrast, the control group did not show statistically significant changes between any of the measurement points.

Table 3. Post Hoc comparisons

Dimension	Group	Comparison	Mean Difference	SE	*p*-value
Accuracy	Experimental	Pre vs. Mid	1.31	0.19	< .001
		Mid vs. Post	1.36	0.18	< .001
		Pre vs. Post	2.67	0.21	< .001
	Control	Pre vs. Mid	0.26	0.17	.142
		Mid vs. Post	0.33	0.16	.081
		Pre vs. Post	0.59	0.18	.064
Intonation	Experimental	Pre vs. Mid	1.11	0.20	< .001
		Mid vs. Post	1.32	0.19	< .001
		Pre vs. Post	2.43	0.22	< .001
	Control	Pre vs. Mid	0.24	0.18	.188
		Mid vs. Post	0.28	0.17	.133
		Pre vs. Post	0.52	0.19	.098
Fluency	Experimental	Pre vs. Mid	1.10	0.19	< .001
		Mid vs. Post	1.39	0.18	< .001
		Pre vs. Post	2.49	0.21	< .001
	Control	Pre vs. Mid	0.28	0.17	.124
		Mid vs. Post	0.27	0.16	.141
		Pre vs. Post	0.55	0.18	.083

4.3 Inter-Rater Reliability

Inter-rater reliability for CAFIS ratings was calculated to ensure scoring consistency between the two independent raters (see Table 4). Intraclass Correlation Coefficient (ICC) values for all dimensions and time points exceeded 0.84, indicating excellent inter-rater agreement (Shieh, 2016).

Table 4. Inter-rater reliability matrix

Time Point	Accuracy ICC (95% CI)	Intonation ICC (95% CI)	Fluency ICC (95% CI)
Pre-test	0.86 (0.79–0.91)	0.84 (0.76–0.90)	0.88 (0.82–0.93)
Mid-test	0.87 (0.80–0.92)	0.85 (0.78–0.91)	0.89 (0.83–0.94)
Post-test	0.89 (0.83–0.94)	0.87 (0.80–0.92)	0.90 (0.85–0.95)

5. Discussion

Given the extensive application of AI-powered tools in educational settings, the present study examined the role of the ELSA Speak application in enhancing EFL learners' pronunciation via the CAFIS dimensions. The findings provide compelling evidence that technology-enhanced instruction, when grounded in a multidimensional framework, can lead to meaningful and sustained improvements in learners' pronunciation skills.

The results revealed that learners who used ELSA Speak achieved substantial improvements in pronunciation accuracy compared to their counterparts receiving traditional instruction. This finding aligns with Rusmawaty et al.'s (2024) identification, who emphasize the effectiveness of explicit, form-focused pronunciation practice supported by ELSA Speak. The AI-driven feedback provided by ELSA Speak appears to have facilitated learners' awareness of segmental errors, enabling repeated self-correction and refinement of sound production. In agreement with this, Abdelhalim and Alsehibany (2025) highlight that segmental accuracy is particularly responsive to technology-assisted instruction due to learners' ability to practice autonomously and receive individualized feedback. Linking with immediate feedback, our finding resonates with Xodabande et al. (2025) observation, who argued that accuracy improvements are most pronounced when feedback is immediate and tailored to learners' weaknesses. The present findings extend this line of research by demonstrating that such gains are not only immediate but also sustained across multiple time points, suggesting that AI-supported practice may contribute to longer-term stabilization of phonological representations.

Regarding intonation, the experimental group also showed significantly greater improvement than the control group, though the magnitude of gains was the most gradual. This pattern is consistent with Gordon and Darcy's (2022) findings, indicating that suprasegmental features tend to develop more slowly than segmental accuracy and require prolonged exposure and practice. The visual and auditory modeling features embedded in ELSA Speak likely supported learners' perception and production of pitch variation and stress placement, which are critical components of intelligible speech. This aligns with Dennis's (2024) reflection who found that feedback-driven prosodic training improves CAFIS intonation scores by increasing learners' awareness of rhythm and stress. Building on this, Al-Bogami and Alahmadi (2025) suggest that technology-mediated input can enhance learners' sensitivity to prosodic features that are often neglected in conventional classroom instruction. Importantly, the observed improvement in intonation reinforces the value of treating pronunciation as a multidimensional construct, as gains in this domain may not be captured through holistic assessment alone.

Regarding fluency, intervention yielded the gradual but consistent improvement across measurement intervals. This result aligns with theoretical accounts of fluency development that emphasize automatization and reduced cognitive load in speech production (Bóna & Bakti, 2020). The structured repetition and continuous speaking tasks encouraged by ELSA Speak likely contributed to smoother speech flow, fewer unnatural pauses, and increased speech rate among learners in the experimental group. Terzioğlu and Kurt (2022) further noted that fluency is often the most resistant aspect of pronunciation development, particularly in EFL contexts where opportunities for authentic spoken interaction are limited. The present findings suggest that AI-powered applications can partially compensate for this limitation by providing learners with frequent, low-anxiety speaking opportunities, thereby supporting gradual fluency development over time.

Taken together, the findings highlight the pedagogical value of integrating AI-powered pronunciation tools into EFL instruction, particularly in higher education contexts where individualized feedback is often difficult to provide. The alignment between ELSA Speak's instructional features and the CAFIS dimensions suggests that technology is most effective when it is theoretically grounded rather than used as a stand-alone supplement. Moreover, this study contributes to the limited body of research on AI-assisted pronunciation learning in the Saudi higher education context, addressing an important gap in the literature. The positive outcomes observed suggest that such tools can be culturally and pedagogically adaptable, offering scalable solutions for pronunciation instruction in EFL settings.

6. Conclusions

This study set out to explore the role of the ELSA Speak application in enhancing EFL learners' pronunciation development. Drawing on the CAFIS analytic framework, the findings provide robust evidence that AI-powered pronunciation instruction can significantly improve learners' pronunciation accuracy, intonation, and fluency when compared to traditional instructional approaches. The differential rates of improvement observed across CAFIS dimensions highlight that pronunciation development is neither uniform nor linear. Segmental accuracy responded most rapidly to AI-mediated feedback, while suprasegmental features such as intonation and fluency required sustained exposure and repeated practice. These findings reinforce the pedagogical value of analytic assessment frameworks that move beyond holistic scoring and allow instructors and researchers to track specific areas of pronunciation growth. The study also demonstrates the instructional potential of AI-driven applications such as ELSA Speak in addressing long-standing challenges in pronunciation teaching. By offering immediate, individualized feedback and opportunities for repeated practice, the application supported learners' autonomous

engagement with spoken English and compensated for limitations commonly associated with large EFL classrooms. Importantly, the positive outcomes observed in the Saudi higher education context suggest that AI-assisted pronunciation tools can be effectively adapted across diverse EFL settings.

Regardless of its significance, the study is not without limitations. The sample size was relatively modest, and pronunciation performance was assessed within a controlled instructional environment. Further research should include a wider diversity of students, long-term outcomes, and approaches to implementation of the tools in classrooms for better validation and generalization of artificial intelligence in language instruction. Additionally, examining the transfer of pronunciation gains to spontaneous communicative contexts would further strengthen understanding of the long-term impact of such interventions. In conclusion, the study provides practical verification for the incorporation of AI-powered pronunciation applications in EFL instruction and underscores the importance of theoretically grounded, multidimensional assessment approaches. By combining technological innovation with analytic pronunciation frameworks, educators and researchers can better support learners' oral development and advance pronunciation pedagogy in higher education.

Acknowledgments

The researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for their financial support (QU-APC-2026).

Authors' contributions

Dr. AHA and Dr. HA were responsible for study design, data collection, and revising. ATA, GFA, EHE, GTA drafted the manuscript and revised it. All authors read and approved the final manuscript. In this paragraph, also explain any special agreements concerning authorship, such as if authors contributed equally to the study.

Funding

Not applicable

Competing interests

The authors declare that they have no competing interests.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

AI Disclosure

During manuscript preparation, standard digital writing-support tools were used for language editing and improving textual. No AI tool was involved in the conception, design, data analysis, interpretation, or generation of scientific content or references. All authors have reviewed and approved the final manuscript and take full responsibility for its academic integrity, in accordance with the journal's policy on AI use.

References

Abdelhalim, S. M., & Alsehibany, R. A. (2025). Integrating AI-powered tools in EFL pronunciation instruction: effects on accuracy and L2 motivation. *Computer Assisted Language Learning*, 1-25. https://doi.org/10.1080/09588221.2025.2534015

Abdelrady, A. H., Ibrahim, D. O. O., & Akram, H. (2025). Unveiling the Role of Copilot in Enhancing EFL Learners' Writing Skills: A Content Analysis. *World Journal of English Language, 15*(8), 174-185. https://doi.org/10.5430/wjel.v15n8p174

Adams-Goertel, R. (2013). Prosodic elements to improve pronunciation in English language learners: A short report. *Applied Research on English Language*, *2*(2), 117-128. https://doi.org/10.22108/ARE.2013.15474

Akram, H., & Abdelrady, A. H. (2023). Application of ClassPoint tool in reducing EFL learners' test anxiety: an empirical evidence from Saudi Arabia. *Journal of Computers in Education*, 1-19. https://doi.org/10.1007/s40692-023-00265-z

Akram, H., & Abdelrady, A. H. (2025). Examining the role of ClassPoint tool in shaping EFL students' perceived E-learning experiences: A social cognitive theory perspective. *Acta Psychologica*, *254*(10477), 104775. https://doi.org/10.1016/j.actpsy.2025.104775

Akram, H., & Li, S. (2024). Understanding the Role of Teacher-Student Relationships in Students' Online learning Engagement: Mediating Role of Academic Motivation. Perceptual and Motor Skills, 00315125241248709. https://doi.org/10.1177/00315125241248709

Akram, H., Abdelrady, A. H., Al-Adwan, A. S., & Ramzan, M. (2022). Teachers' perceptions of technology integration in teaching-learning practices: A systematic review. *Frontiers in psychology*, *13*, 920317. https://doi.org/10.3389/fpsyg.2022.920317

Akram, H., Yingxiu, Y., Al-Adwan, A. S., & Alkhalifah, A. (2021). Technology Integration in Higher Education During COVID-19: An Assessment of Online Teaching Competencies Through Technological Pedagogical Content Knowledge Model. *Frontiers in Psychology*, 12, 736522-736522. https://doi.org/10.3389/fpsyg.2021.736522

Al-Adwan, A. S., Nofal, M., Akram, H., Albelbisi, N. A., & Al-Okaily, M. (2022). Towards a sustainable adoption of e-learning systems: The role of self-directed learning. *Journal of Information Technology Education: Re-search*, *21*, 245-267. https://doi.org/10.28945/4980

Al-Bogami, R. M., & Alahmadi, N. A. (2025). Effects of an AI-based reading progress tool on third-grade EFL learners' oral reading fluency. *Computers and Education Open*, 100283. https://doi.org/10.1016/j.caeo.2025.100283

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

Arbain, A., Sari, I. P., & Rahmawati, D. (2023). The effectiveness of Elsa Speak and Google Translate in improving students' pronunciation. *Journal of English Language Teaching and Linguistics*, *8*(1), 45-58. https://doi.org/10.21462/jeltl.v8i1.1234

Aryanti, T., & Santosa, R. (2024). Artificial intelligence in language learning: A systematic review of pronunciation tools. *International Journal of Educational Technology*, *11*(2), 101-115. https://doi.org/10.31098/ijet.v11i2.567

Aulia, R., & Santosa, R. (2025). Enhancing pronunciation through AI: A PRISMA review of Elsa Speak in EFL and ESL contexts. *TESOL International Journal*, *20*(1), 88-102. https://doi.org/10.31098/tesol.v20i1.789

Bóna, J., & Bakti, M. (2020). The effect of cognitive load on temporal and disfluency patterns of speech: evidence from consecutive interpreting and sight translation. *Target*, *32*(3), 482-506. https://doi.org/10.1075/target.19041.bon

Cole, R. (2024). Inter-rater reliability methods in qualitative case study research. *Sociological Methods & Research*, *53*(4), 1944-1975. https://doi.org/10.1177/00491241231156971

Dahmen, S., Grice, M., & Roessig, S. (2023). Prosodic and segmental aspects of pronunciation training and their effects on L2. *Languages*, *8*(1), 74. https://doi.org/10.3390/languages8010074

Dennis, N. K. (2024). Using AI-Powered Speech Recognition Technology to Improve English Pronunciation and Speaking Skills. *IAFOR Journal of Education*, *12*(2), 107-126. https://doi.org/10.22492/ije.12.2.05

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL quarterly*, *39*(3), 379-397. https://doi.org/10.2307/3588486

Derwing, T. M., & Munro, M. J. (2022). Pronunciation learning and teaching. In *The Routledge handbook of second language acquisition and speaking* (pp. 147-159). Routledge. https://doi.org/10.4324/9781003022497-14

Engel de Abreu, P. M., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology*, *104*(4), 974. https://doi.org/10.1037/a0028390

Fetzer, A. (2008). Communicative intentions in context. In *Rethinking sequentiality: Linguistics meets conversational interaction* (pp. 37-69). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.103.03fet

Godwin-Jones, R. (2018). Using mobile technology to develop language skills and cultural understanding. *Language Learning & Technology*, *22*(3), 3-11. https://doi.org/10.64152/10125/44150

Gordon, J., & Darcy, I. (2022). Teaching segmentals and suprasegmentals: Effects of explicit pronunciation instruction on comprehensibility, fluency, and accentedness. *Journal of Second Language Pronunciation*, *8*(2), 168-195. https://doi.org/10.1075/jslp.21042.gor

Hammond, M. (2020). Prosodic phonology. *The handbook of English linguistics*, 365-384. https://doi.org/10.1002/9781119540618.ch19

Jalalzai, N. N., Akram, H., Khan, M., & Kakar, A. K. (2025). Technology Readiness in Education: An Analysis of ICT Facilities in High Schools of Loralai, Balochistan. *Contemporary Journal of Social Science Review*, *3*(3), 2835-2842. https://doi.org/10.63878/cjssr.v3i3.1321

Jenkins, J. (2004). 5. research in teaching pronunciation and intonation. *Annual review of applied linguistics*, *24*, 109-125.

https://doi.org/10.1017/S0267190504000054

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/10.32332/JOELT.V9I1.2723

Li, S., & Akram, H. (2023). Do emotional regulation behaviors matter in EFL teachers' professional development? A process model approach. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, *9*, 273-291. https://doi.org/10.30827/portalin.vi2023c.29654

Li, S., & Akram, H. (2024). Navigating Pronoun-Antecedent Challenges: A Study of ESL Academic Writing Errors. *SAGE Open*, *14*(4), 21582440241296607.

Lin, H., & Chen, Q. (2024). Artificial intelligence (AI)-integrated educational applications and college students' creativity and academic emotions: students and teachers' perceptions and attitudes. *BMC psychology*, *12*(1), 487. https://doi.org/10.1186/s40359-024-01979-0

Liu, D., McGregor, A., Zielinski, B., Reed, M., & Meyers, C. (2025). ESL teachers' metalanguage as evidence of their metalinguistic knowledge of the English intonation system. *Language awareness*, *34*(2), 324-344. https://doi.org/10.1080/09658416.2024.2370886

Liunokas, Y. (2025). The Effectiveness of the Elsa Speak Application in Enhancing Speaking Skills among Students at University. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, *13*(2), 8221-8235. https://doi.org/10.24256/ideas.v13i2.6955

Ma, D., Akram, H., & Chen, I. H. (2024). Artificial Intelligence in Higher Education: A Cross-Cultural Examination of Students' Behavioral Intentions and Attitudes. *International Review of Research in Open and Distributed Learning*, *25*(3), 134-157. https://doi.org/10.19173/irrodl.v25i3.7703

Metruk, R. (2024). Mobile-assisted language learning and pronunciation instruction: A systematic literature review. *Education and Information Technologies*, *29*(13), 16255-16282. https://doi.org/10.1007/s10639-024-12453-0

Mora-Plaza, I., Mora, J. C., Ortega, M., & Aliaga-Garcia, C. (2024). Is L2 pronunciation affected by increased task complexity in pronunciation-unfocused speaking tasks?. *Studies in Second Language Acquisition*, *46*(4), 1117-1149. https://doi.org/10.1017/s0272263124000470

Ou, A. W., Stöhr, C., & Malmström, H. (2024). Academic communication with AI-powered language tools in higher education: From a post-humanist perspective. *System*, *121*, 103225. https://doi.org/10.1016/j.system.2024.103225

Pennington, M. C. (2021). Teaching pronunciation: The state of the art 2021. *Relc Journal*, *52*(1), 3-21. https://doi.org/10.1177/00336882211002283

Pham, V. T. T., & Pham, A. T. (2025). English major students' satisfaction with ELSA Speak in English pronunciation courses. *PloS one*, *20*(1), e0317378. https://doi.org/10.1371/journal.pone.0317378

Rusmawaty, D., Limbong, E., Ahada, I., Hafizh, M., & Rahmatullah, Achmad. N. (2024). Unlocking Phonological Proficiency: Exploring Allophonic Variation Using ELSA Speak App in Early Semester EFL Students at Mulawarman University. *Indonesian Journal of EFL and Linguistics*, 337-348. https://doi.org/10.21462/ijefl.v9i2.828

Shieh, G. (2016). Choosing the best index for the average score intraclass correlation coefficient. *Behavior research methods*, *48*(3), 994-1003. https://doi.org/10.3758/s13428-015-0623-y

Sholekhah, M. F., & Fakhrurriana, R. (2023). The use of ELSA Speak as a mobile-assisted language learning (MALL) towards EFL students pronunciation. *Journal of Education, Language Innovation, and Applied Linguistics*, *2*(2), 93-100. https://doi.org/10.37058/jelita.v2i2.7596

Sohail, A., & Akram, H. (2025). The Role of Self-Awareness and Reflection in Academic Achievement: A Psychological and Bayesian Analysis. *Pedagogical Research*, *10*(1). https://doi.org/10.29333/pr/15682

Suzuki, Y., Maie, R., & Hui, B. (2025). Research timeline: Automatization in second language learning. *Language Teaching*, 1-20. https://doi.org/10.1017/S026144482500059X

Terzioğlu, Y., & Kurt, M. (2022). Elevating English language learners' speaking fluency and listening skill through a learning management system. *Sage Open*, *12*(2), 21582440221099937. https://doi.org/10.1177/21582440221099937

Twisk, J., & Proper, K. (2004). Evaluation of the results of a randomized controlled trial: how to define changes between baseline and follow-up. *Journal of clinical epidemiology*, *57*(3), 223-228. https://doi.org/10.1016/j.jclinepi.2003.07.009

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama*, *310*(20), 2191-2194. https://doi.org/10.1001/jama.2013.281053

Xodabande, I., Shiri, S., & Zohrabi, M. (2025). Exploring the impacts of an AI-driven instructional intervention on Iranian EFL learners' pronunciation skill development. *Discover Education*, *4*(1), 307. https://doi.org/10.1007/s44217-025-00782-2

Zuhairya, N., Maharani, P., Ridwan, A., Khairunnisa, K., Ahwani, S., & Lubis, Y. (2024). The Role of Suprasegmental Features in English Phonology: Prosodic Hierarchy and Intonation Patterns. *Fonologi: Jurnal Ilmuan Bahasa dan Sastra Inggris*, *2*(2), 154-161. https://doi.org/10.61132/fonologi.v2i2.670