How AI-Powered Voice Recognition Has Supported Pronunciation Competence among EFL University Learners

Nguyen Truong Sa[1*], Nguyen Thi Diem Thi[1], Hoang Ngoc Quynh Nhu[1], Do Thi Kim Hieu[1]

[1]Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Vietnam
[*]Corresponding author's email: nguyentruongsa@iuh.edu.vn
[*]https://orcid.org/0000-0002-4025-2144
https://doi.org/10.54855/callej.252634

Received: 07/10/2024 Revision: 05/03/2025 Accepted: 11/03/2025 Online: 12/05/2025

ABSTRACT

This study investigates the extent to which AI-powered voice recognition technology supports and enhances pronunciation competence among EFL learners and scaffolds instructional practices. A 4-month experiment was conducted with 37 first-year English majors, integrating this technology developed in the ELSA Speak App into classroom activities and self-study sessions during a pronunciation course. Students' progress was evaluated through pretest and posttest scores, as well as daily practice records. Findings from the study revealed a strong relationship between the frequency and intensity of practice with the AI tool and improvements in students' pronunciation competence. Repeated and targeted drills contributed to noticeable enhancements in pronunciation accuracy, while teacher guidance was essential in facilitating learners' progress. The students reported that they had employed the tool to accommodate their diverse learning needs and strategies. Despite technical issues and variations in accent recognition, the participants demonstrated positive attitudes toward the technology, recognizing its value in pronunciation instruction. Future research should explore the long-term impact of AI-powered tools on pronunciation improvement and their applicability across diverse linguistic and cultural contexts.

Keywords: AI, pronunciation competence, voice recognition

Introduction

The most common and direct method of communication among humans is through speech. In the era of integration and internationalization, communicating in spoken English has become increasingly important for work and study. As pronunciation instruction is a key foundation for successful oral communication and an important part of improving learners' communicative competence (Do, 2021; Hismanoglu, 2006; Hamada, 2016; Sarmita Samad & Aminullah, 2019), the role of pronunciation teaching has received much attention from researchers and language educators (Foote et al., 2016; Pourhosein, 2016; Bogach et al., 2021; Anggraini, 2022). Researchers have also highlighted that the acquisition of native English pronunciation

CITATION | Nguyen, T. S., Nguyen, T. D. T., Hoang, N. Q. N., & Do, T. K. H. (2025). How AI-Powered Voice Recognition Has Supported Pronunciation Competence among EFL University Learners. *Computer-Assisted Language Learning Electronic Journal (CALL-EJ),* 26(3), 64-83. DOI: https://doi.org/10.54855/callej.252634

is not an easy goal and that most learners need much practice to improve their pronunciation (Fraser, 2000; Pourhosein, 2016). Teachers' beliefs and practices also significantly influence pronunciation teaching in EFL classrooms (Do, 2021). According to Fraser (2000), to improve the quality of pronunciation instruction, a teacher should be provided with both courses and materials that are supportive of pronunciation practice.

The fact that artificial intelligence (AI) is becoming increasingly prevalent in educational settings is not a new matter; AI developed for educational purposes, with the intelligence to mimic human brains and skills, is significantly transforming practices in various majors, including language teaching. In an attempt to investigate how AI technologies are reshaping language teaching and learning practices, many researchers, such as Popenici and Kerr (2017), Wang and Liu (2019) and Zou et al. (2023), have highlighted that AI helps personalize teaching styles and learning experiences and enhances engagement and learning outcomes. Particularly in teaching pronunciation to learners of English as a second or foreign language, there has been significant growth in interest and attention given to technology employment in the last decade (Popenici & Kerr, 2017; Wang & Liu, 2019; Sariani et al., 2022; Kazu & Kuvvetli, 2023); Zou et al., 2023; Nicolau, 2024), with expectations to provide rigorous evidence on the influences and outputs. With the development of educational technology, some recent studies have even examined how voice recognition is employed for pronunciation instruction, and some evidence of its effectiveness has been confirmed (Sujatha et al., 2020; Akhmad & Munawir, 2022; Nhu et al., 2024). However, these results were evaluated only by teachers (Akhmad & Munawir, 2022), from meta-analyses of collected data from the literature (Nhu et al., 2024), or from only the segmental level of pronunciation, as in the study of Sujatha et al. (2020). Consequently, a more systematic and triangulated study to describe the extent to which voice recognition powered by AI technology supports and enhances pronunciation competence is imperative to fill the gap in theory and scaffold instructional practices.

Literature review

Pronunciation refers to the capacity to accurately perceive and produce vowel and consonant sounds, which are classified as segmental aspects of speech. These sounds, unlike English alphabet letters, are phonetic symbols that learners need to be aware of and distinguish to master their pronunciation (Anggraini, 2022). The scope of teaching pronunciation also revolves around suprasegmental elements that make learners more like native speakers, namely, rhythm and intonation.

Challenges in teaching and learning pronunciation

When practicing their pronunciation, learners frequently grapple with formidable difficulties. First, the influence of their mother tongue impairs learners' mastery of pronunciation (Sarmita Samad & Aminullah, 2019). Furthermore, a lack of interaction with native English speakers inhibits their pronunciation learning. Minimal exposure to native speech negatively impacts EFL learners' pronunciation skills, as they are not accustomed to the nuances of native speech (Almalki & Algethami, 2022). In addition, time constraints in classes prevent teachers from taking care of every individual and giving them constructive feedback on their pronunciation errors, while they are also demotivated from self-learning pronunciation at home. Apparently, their insufficient practice exacerbates their struggle to speak English fluently and precisely. Teachers' corrective feedback strategies, additionally, are not effective enough, as teachers usually do not point out the error but only ask for mimicry, which is hence ineffective in promoting learners' pronunciation uptake (Saito, 2013). In this sense, an investigation into

instructional approaches that provide detailed feedback is essential. As demand continues to grow, educators and scholars are exploring innovative approaches to enhance pronunciation competency. The conventional teaching techniques and methods applied in classes cannot greatly improve pronunciation. Employing technology, particularly mobile-based artificial intelligence, to address this situation would be plausible. The incorporation of efficient technology, exemplified by AI-powered pronunciation tools, can be instrumental in enhancing students' pronunciation proficiency and enriching their language learning journey.

AI for pronunciation instruction and practice

AI technologies integrate human intelligence into computer systems, allowing them to think, work, and make judgments like humans do while also using data for learning, self-correction, and self-improvement (Kok et al., 2009; Popenici & Kerr, 2017). In the field of language education, technologies and their far-reaching impact have gradually been recognized and proven by scholars; they have revolutionized language learning by automating administrative tasks, providing personalized and interactive experiences, and differentiating feedback to learners. Teachers in Nicolau's research (2024) perceived AI as a beneficial tool to streamline their workload, allowing them to focus more on facilitating learning rather than performing repetitive tasks. The researcher also emphasized that collaboration among teachers could be facilitated by providing platforms for sharing resources, strategies, and best practices in integrating technology into their teaching. For students, in Zou et al.'s (2023) quantitative research, AI apps were proven to enhance fluency in speaking skills, grammar, and idea organization. This finding aligns with the study conducted by Wang and Liu (2019), which revealed that students' oral proficiency and confidence were increased when AI-driven applications were harnessed. Despite the benefits, AI tools have limitations, such as reliance on specific accents (e.g., US English in ELSA), emphasis mainly on accurate production and comprehension of oral language with little or no emphasis on the context, and fee payment required for better versions (Nushi & Sadeghi, 2021), etc.

In addition to fostering the four fundamental skills, AI technologies also excel in enhancing pronunciation. Noviyanti (2020) conducted a study on improving students' pronunciation via an AI-based pronunciation checker. The findings of the study highlighted that it significantly improved students' pronunciation skills and that it was effective in promoting independent learning of English pronunciation. Besides, while traditional teaching methods often fall short of providing immediate, personalized feedback, AI offers real-time, precise evaluations (Moxon, 2024). Automatic speech recognition (ASR) in Sariani et al.'s study in 2022 was observed to support learners by allowing them to practice frequently and receive real-time assessments. In addition, learner autonomy was enhanced through self-directed learning, flexible practice, prompt feedback, intrinsic motivation, goal setting, and collaboration with peers. ASR can analyze various aspects of speech, such as intonation, stress, and phonetic accuracy, offering personalized suggestions for improvement (Zou et al., 2023). Kazu and Kuvvetli (2023) employed an experimental design with high school students and reported that, compared with traditional phonetic alphabet methods, AI-supported speech recognition significantly improved vocabulary retention and pronunciation skills. Web-based applications with ASR and Automatic Pronunciation Assessment (APA) also offer unbiased, comprehensive feedback and facilitate autonomous practice (Moxon, 2024).

To describe a more detailed improvement and examine deeper into the learning process, this study aims to employ AI-powered voice recognition of the ELSA Speak to enhance students' pronunciation skills. The ELSA (English Language Speech Assistant) Speak describes itself as a world-leading mobile solution designed to help non-native English learners improve speaking

skills through proprietary AI-powered voice and speech recognition technology. The developers also claim in their user manual that the ELSA Speak can provide instant feedback on pronunciation, fluency, intonation, word stress, listening, grammar, and vocabulary use and suggest personalized learning plans for each user. Accordingly, some of their latest features with AI technology to facilitate voice recognition include detecting pronunciation errors at various levels of speech and providing real-time feedback and assessment on the basis of the American accent.

To evaluate the efficacy of the ELSA Speak for improving English, Samad and Aminullah (2019) highlighted students' positive perceptions of the ELSA Speak software for pronunciation training, with favorable scores in automatic speech recognition and software design. The study indicated that AI-powered pronunciation tools, such as the ELSA Speak, enhanced learner engagement and motivation and tripled the chance to practice. Darsih et al. (2021) reported that students positively perceived the ELSA Speak app and highly rated its content, pedagogical design, and assessment features, which were deemed beneficial for improving pronunciation and speaking skills. According to Anggraini (2022), the app's interactive features and engaging content also create an immersive learning experience that motivates students to have a good command of their pronunciation. In the same year, Akhmad and Munawir employed a pre-test and post-test design in which twenty students read a text prepared by teachers aloud to assess pronunciation before and after using the ELSA Speak app. The increase in scores from 52.30 to 65.60 and the positive reporting of 82% of the students in their research showed significant improvement, underscoring the app's effectiveness in enhancing their skills. Another study in 2022 by Sariani et al. revealed that the use of the ELSA Speak application not only effectively enhanced students' pronunciation skills but also increased self-confidence, reduced anxiety, and increased learner autonomy when learning language. To substantiate these claims, Ngoc et al. (2023) scrutinized the perspectives of tertiary nonspecialized students through a structured questionnaire-based survey and explored the predominantly positive reception of AI in language teaching. Therefore, an app like ELSA can empower language learners to independently improve their pronunciation and develop greater autonomy in mastering the skill while also being a valuable supplementary tool for teachers (Nushi & Sadeghi, 2021).

Generally, studies on the ELSA Speak app have highlighted its effectiveness in enhancing pronunciation accuracy and motivating extensive practice among EFL learners. Students from diverse backgrounds reported positive experiences, with significant improvements in their pronunciation skills. The ability of the app to provide instant and AI-driven feedback has been particularly valuable, making it a favorable tool for language learners seeking to refine their spoken English. Overall, the ELSA Speak has been evaluated as a beneficial resource in the realm of computer-assisted language learning.

This study has the same agenda as the above works in that it focuses on the integration of the ELSA Speak in educational practices. Nevertheless, previous studies highlighted students' or teachers' perceptions and examined improvements but lacked a detailed description of how the improvements were assured. This research is distinctive since its emphasis was placed not only on the efficacy of the ELSA Speak but also on key drivers leading to students' pronunciation enhancement.

Research Questions

This study investigated the impacts of the AI-powered voice recognition provided by the ELSA Speak on the development of learners' pronunciation competence and their evaluation of such tool employment. The research questions are as follows:

1. How does voice recognition powered by AI technology impact the development of learners' pronunciation proficiency?
2. How do the learners evaluate the implementation of AI-powered voice recognition in their pronunciation practice?

Methods

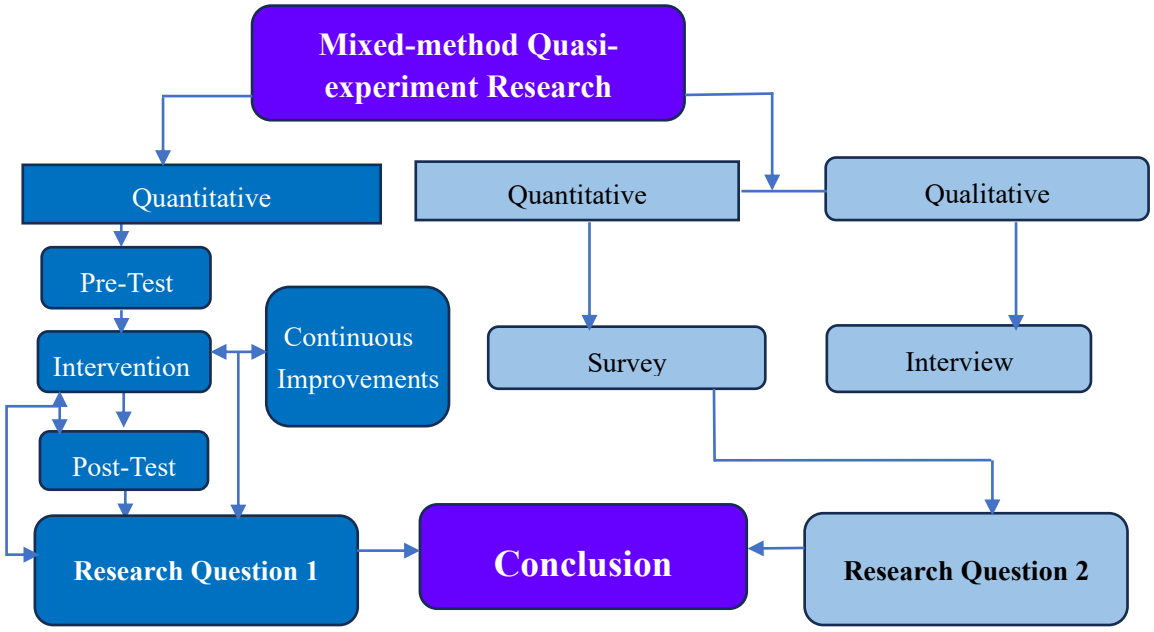
Pedagogic setting and participants

The participants were 38 first-year students in a Bachelor in English language program. However, as one of them misused the given ELSA account by sharing it with other students, her data became invalid for research. Pronunciation Practice, a compulsory course in the program, was 45 hours (3 credits) long and was scheduled 3 hours for each class per week. The students registered for the course by themselves and were assigned randomly to the same class. In a regular class, the teacher and students used a coursebook describing sounds and sound articulation as reference material. Most of the class time was given for practice and correction at individual sounds or phonemes, words, and connected speech levels. Most homework was self-practice without any systematic tool to support the teacher in evaluation, and she did not have much time for such work. The final score was a combination of formative and summative assessments; the assessment process was based on rubrics marked by a subjective evaluation of the teacher on the student's ongoing performance of English sounds in the pronunciation test. In this course, the ELSA Speak was employed by the teacher as a tool to provide additional assignments and to encourage the students to use it for self-study.

Design of the study

Figure 1

Design of the study



The design is summarized in Figure 1. The research employed mixed methods as its methodological framework and was conducted as a one-group pretest/posttest quasi-experiment

study. Quasi experiments have been widely adopted by scholars and noted recently by Gopalan, Rosinger, and Ahn (2020) as the "gold standard" for establishing causal evidence in education. This recognition stems from their advantages, such as lower costs, the ability to avoid challenges associated with place-based or policy-driven interventions, and fewer ethical concerns than other research methods. One group pretest-posttest design was suggested by scholars such as Johnson and Christensen (2019) and Privitera and Delzell (2019) as "a quasiexperiment research design in which the same dependent variable is measured in one group of participants before and after a treatment is administered" (Privitera & Delzell, 2019, p.337). By employing this design, Marsden and Torgerson (2012) realized that the observed improvements in the results may be ascribed to the intervention in a causal relationship. Consequently, to minimize problems related to validity, such as history, maturation, and test effects, to make conclusions about the effectiveness of voice recognition in enhancing the quality of teaching and learning, researchers have reached conclusions on the basis of triangulations of both qualitative and quantitative data: pretest, posttest, continuous improvements in scores, correlations between the scores and recorded learning behaviors, and evaluations from the students.

Data collection and analysis

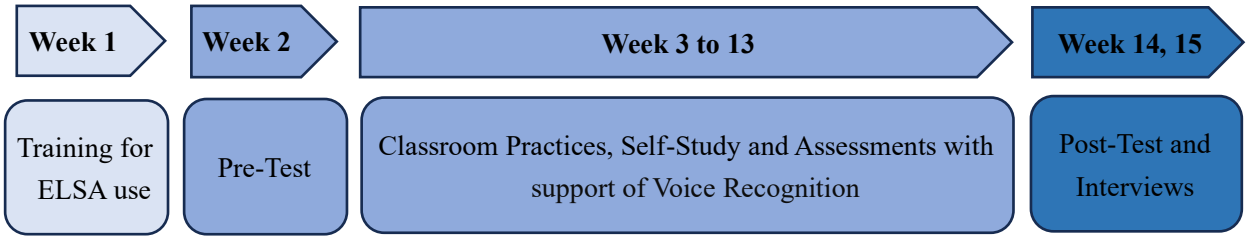
Training on using the ELSA Speak

To help the teachers and students become familiar with the functions and use of the ELSA Speak, an orientation was organized at the beginning of the course by an ELSA technician. Additionally, as the provided drills in the ELSA Speak did not match all the content of the syllabus, the teacher was guided to use the customized function wired in the system to tailor some of the teaching lessons and drills for her students to practice.

Data collection

Figure 2

Timeline of data collection

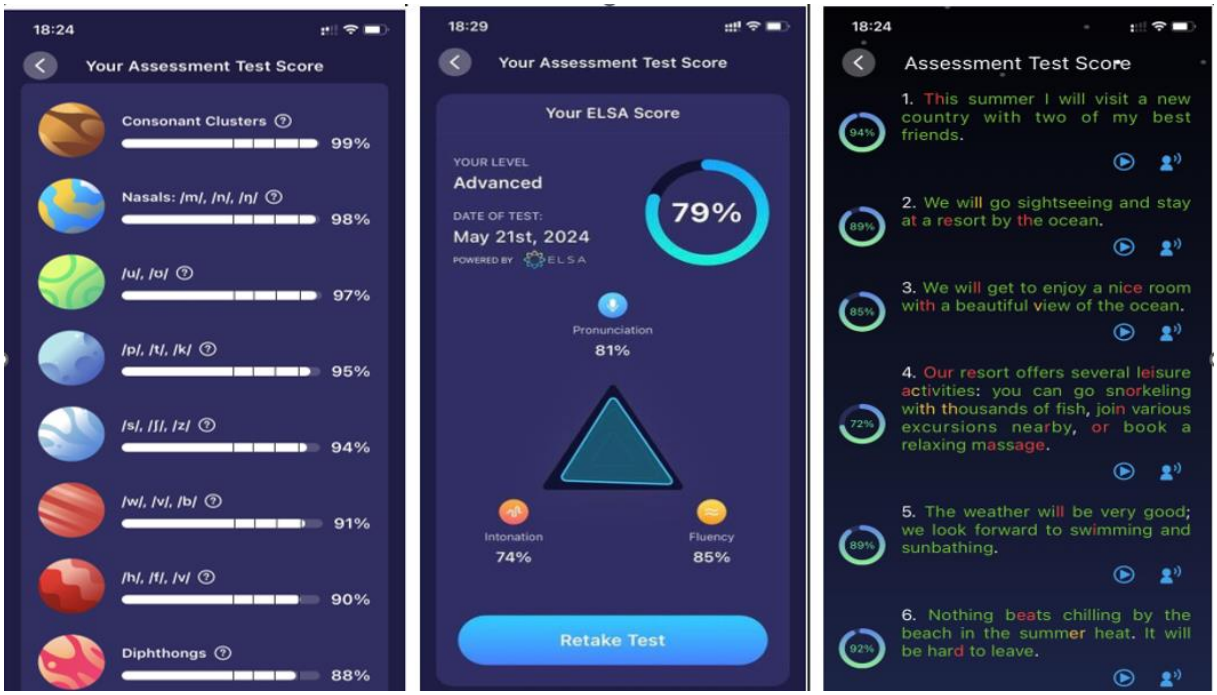


After orienting the teachers and learners on how to make use of the ELSA Speak to support the training process in week 1, the students were encouraged to take the pretest in week two of the syllabus. The test required users to read aloud sixteen sentences, covering all English vowels and consonants. The participants completed the test individually with their given account to interact with the system. Once finished, the score was updated to evaluate proficiency in terms of pronunciation, fluency, and intonation (Figure 3). The students could see detailed reports showing the scores and analysis for each group of pronunciation skills in each sentence they had recorded. From the reports, the teacher could adjust the classroom content and methods to suit most of the students' levels from weeks 3 to 13, and the ELSA Speak then accordingly created a personalized learning path in Daily Lessons (Coach) for individuals. As students completed more lessons, the ELSA Speak scores were also automatically updated to reflect

their latest progress. In week 14, the students were encouraged to take the test again (posttest) to determine how their performance had changed. They were also asked to complete a survey reporting their evaluation of the learning methods. All the interaction behaviors and history with the system were recorded during the course. On the basis of such records, the researchers chose the most and the least frequent ESLA users for focus group interviews to determine how they evaluated the learning progress with the support of the ELSA Speak in greater depth.

Figure 3

Screenshot examples of pretest visibility to users



Data analysis

Figure 4

A teacher and admin dashboard screenshots showing final reports

The management system automatically evaluated individual students' progress by tracking and calculating their amount of practice time, system usage frequency, and level of accuracy and fluency (Figure 4). Additionally, student progress was also evaluated by comparing the results from the pretest and posttest assessments. When all of the data had been collected, correlations among factors such as the engagement rate, number of active days, total practice time, and assessment scores were examined via SPSS. To determine how the participants evaluate

teaching and learning pronunciation with the support of AI, a survey with all of the participants and focus group interviews with the most and the least active students in terms of total practice time were conducted.

Results

The impacts of AI-powered voice recognition on learners' pronunciation proficiency

Table 1

Final records reported by the ELSA Speak

1. Stu. Num -ber	2. Time Studied (second)	4. Lessons Comple -ted (Total)	5. Days Active (Total)	6. Learner Engage- ment Rate (%)	7. Avg. Study Time (Per Day)	8. Avg. Study Time (Per Active Day)	9. Avg. Lessons (Per Day)	10. Avg. Lessons (Per Active Day)	11. First Assess -ment Score (%)	12. Last Assess -ment Score (%)	13. Assess- ment Score Change (%)	14. EPS Improve- ment (%)
1	**3260**	24	3	2	24	1087	0	8	62	62	**0**	**01**
2	**6609**	88	7	5	48	944	1	13	56	56	**0**	**01**
3	**8098**	106	17	12	59	476	1	6	55	54	**-1**	**06**
4	**15040**	240	15	11	109	1003	2	16	68	76	**08**	**11**
5	**15555**	228	13	9	113	1197	2	18	68	70	**02**	**03**
6	**16334**	203	6	4	118	2722	1	34	54	50	**-4**	**17**
7	**18823**	286	16	12	136	1176	2	18	76	81	**05**	**12**
8	**20286**	276	19	14	147	1068	2	15	53	55	**02**	**12**
9	**20324**	272	13	9	147	1563	2	21	57	62	**05**	**11**
10	**21620**	290	9	7	157	2402	2	32	54	68	**14**	**19**
11	**21655**	162	13	9	157	1666	1	12	56	60	**04**	**16**
12	**21675**	222	26	19	157	834	2	9	66	68	**02**	**19**
13	**22733**	279	22	16	165	1033	2	13	47	45	**-2**	**14**
14	**22927**	259	18	13	166	1274	2	14	48	55	**07**	**24**
15	**23378**	266	17	12	169	1375	2	16	55	58	**06**	**17**
16	**24554**	303	17	12	178	1444	2	18	29	52	**23**	**13**
17	**26236**	223	21	15	190	1249	2	11	49	55	**16**	**24**
18	**26497**	288	19	14	192	1395	2	15	55	58	**28**	**06**

1. Stu. Num -ber	2. Time Studied (second)	4. Lessons Comple -ted (Total)	5. Days Active (Total)	6. Learner Engage- ment Rate (%)	7. Avg. Study Time (Per Day)	8. Avg. Study Time (Per Active Day)	9. Avg. Lessons (Per Day)	10. Avg. Lessons (Per Active Day)	11. First Assess -ment Score (%)	12. Last Assess -ment Score (%)	13. Assess- ment Score Change (%)	14. EPS Improve- ment (%)
19	**26610**	286	14	10	193	1901	2	20	22	61	**39**	**53**
20	**27619**	284	26	19	200	1062	2	11	50	78	**28**	**19**
21	**28906**	242	24	17	209	1204	2	10	42	58	**16**	**24**
22	**29493**	312	18	13	214	1638	2	17	56	54	**14**	**20**
23	**32197**	312	23	17	233	1400	2	14	47	63	**16**	**24**
24	**33129**	213	31	22	240	1069	2	7	37	61	**24**	**33**
25	**34657**	347	26	19	251	1333	3	13	51	72	**32**	**17**
26	**37336**	329	27	20	271	1383	2	12	47	54	**13**	**23**
27	**37382**	297	22	16	271	1699	2	14	47	48	**02**	**18**
28	**39360**	453	40	29	285	984	3	11	37	65	**28**	**44**
29	**41096**	347	23	17	298	1787	3	15	33	87	**54**	**39**
30	**42267**	343	27	20	306	1565	2	13	55	87	**32**	**30**
31	**43947**	341	27	20	318	1628	2	13	53	64	**11**	**23**
32	**48083**	349	29	21	348	1658	3	12	63	71	**08**	**23**
33	**49672**	390	34	25	360	1461	3	11	31	44	**33**	**32**
34	**50042**	288	24	17	363	2085	2	12	49	49	**10**	**16**
35	**53089**	427	65	47	385	817	3	7	59	63	**27**	**16**
36	**55895**	468	36	26	405	1553	3	13	76	74	**17**	**11**
37	**69807**	631	65	47	506	1074	5	10	55	61	**23**	**12**

Table 1 shows the general records of 37 students in the class; the records are visible and exportable to the teacher and admin of the system. The report showed each student's name (column 1), their interaction history with the system via indicators such as total interaction time (column 2), the number of completed lessons (column 4), their engagement rate (column 6), the outcome of the pretest (column 11), the posttest (column 12), the difference between the pretest and posttest (column 13), and their continuous improvement while working on the assignments (column 14). The data were arranged in ascending order of total time (in seconds) for the students working with the ELSA system. The results showed that the time students used to work with the ELSA Speak was quite varied. The shortest total interaction time was 3260 seconds

(nearly 1 hour), whereas the longest interaction time was 69807 seconds (more than 19 hours). With the help of the ELSA Speak, more than half of the class spent 7--10 hours in the course to practice their pronunciation. As a result, the exported records also show a variation in the number of completed lessons, engagement rates, number of active days, average amount of time and number of lessons, and continuous improvement among the students. As mentioned, continuous improvement was calculated by evaluating each student's ability to produce sounds accurately and fluently at the phoneme, word, and sentence levels. This calculation was carried out automatically by the voice recognition technology of the ELSA Speak during every completed lesson made available for the student to practice. The improvement level, which was assessed by percentage (column 14), was very low for some students (1 to 3 percent) but very high for others (more than 40 percent). In terms of posttest scores in comparison with pretest scores (column 13), interestingly, some students performed even worse than they did just a couple of months before (student numbers 3, 6, and 13). Two of the students who had the least interaction time (students 1 and 2) did not show any change in score after completing the posttest. However, most of them achieved significant changes in their scores, resulting in improvements. In particular, the score of student number 29 even decreased from 33% to 87%, making him the lead in the table regarding test core improvement.

Table 2

Pearson correlation between improvement scores and recorded learning behaviors

		EPS Improvement (%)	Assessment Score Change (%)	Time Studies (by second)	Lessons Completed (Total)	Days Active (Total)	Learner Engagement Rate (%)	Avg. Study Time (Per Day)	Avg. Study Time (Per Active Day)	Avg. Lessons (Per Day)	Avg. Lessons (Per Active Day)
EPS Improvement (%)	Pearson Correlation	1	.644**	.372*	.338*	.226	.231	.372*	.259	.307	.001
	Sig. (2-tailed)		.000	.023	.041	.179	.169	.023	.121	.064	.994
	N	37	37	37	37	37	37	37	37	37	37
Assessment Score Change (%)	Pearson Correlation	.644**	1	.525**	.522**	.429**	.437**	.525**	.093	.537**	-.130
	Sig. (2-tailed)	.000		.001	.001	.008	.007	.001	.584	.001	.441
	N	37	37	37	37	37	37	37	37	37	37

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Pearson correlation analysis (2-tailed) at the 0.05 significance level revealed the relationships among student continuous improvement scores (ESP improvement), posttest scores compared with pretest scores (assessment score change), and different variables of interaction with the AI technology employed to support their learning. The results revealed that the ESP score was positively related to the time spent studying with the ELSA Speak (r = .372, p = .023 < .05), the number of completed lessons (r = 338, p = .041), and the average study time per day (r =. 372, p = .023). Moreover, correlations with other variables seemed to be inconcludable, as the

p values were found to be greater than 0.05. These results implied that the accumulated number of drills or intensive and regular practices clearly affected the pronunciation of more fluent and accurate English sounds, words, and sentences.

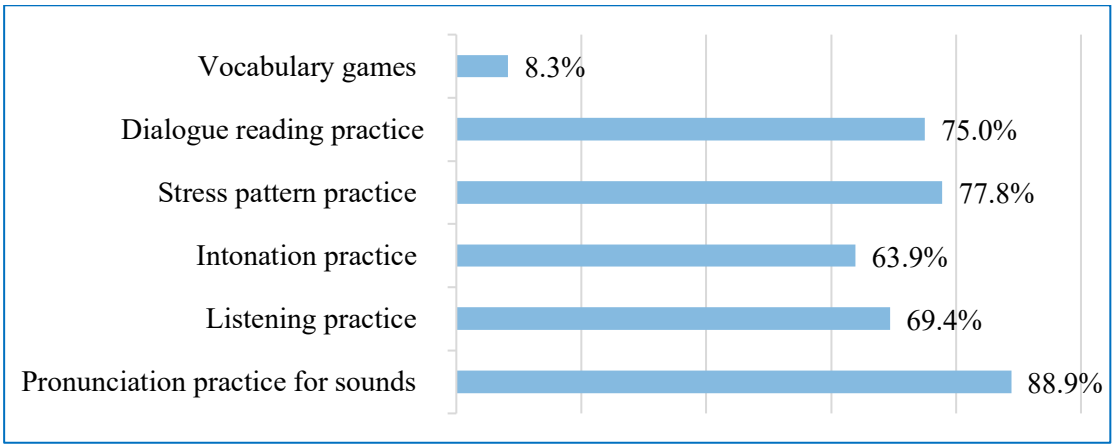
The correlations were found to be much more positively significant (with a *p value* of 0.01) when the changes in posttest scores were compared with the pretest scores (assessment score change). As shown in Table 2, the correlated values were as follows: r = .525 (*p* = .001) with total study time, *r* = .522 (*p* = .001) with the number of completed lessons, *r* = .429 (p = .008) with the total number of active days, *r* = .437 (*p* = .007) with the engagement rate, and *r* = .537 (*p* = .001) with the average number of lessons per day. The correlation values with the average study time and average number of lessons per active day were not statistically meaningful, as the probability values were higher than expected (P>0.05). These statistics strongly suggest that the intensive and regular practice with the ELSA Speak, in addition to the earlier-mentioned improvements in the ongoing drills, also helped students enhance their competence in pronunciation.

Learners' employment of the ELSA Speak with AI-powered voice recognition

When the ELSA Speak app was used, the students demonstrated various usage patterns to support their learning. The majority (69.4%) were focused on medium-level lessons, whereas 13.9% and 16.7% engaged with difficult and easy lessons, respectively. In addition, most students (66.7%) selected materials that integrated both personal interests and class content, which indicated a preference for a balance of relevance and personalization in their learning. Moreover, 27.8% of the students concentrated exclusively on class-related content, and a smaller group (5.5%) opted for content solely on the basis of personal interests. While the ELSA Speak app offers a range of features, the students focused primarily on improving their pronunciation, speaking, and listening skills. In particular, pronunciation practice for sounds was the most frequently used feature, with 88.9% of the students engaging in it, followed closely by stress pattern practice (77.8%) and dialog reading practice (75.0%). Listening practices (69.4%) and intonation practices (63.9%) were also widely implemented, reflecting a strong emphasis on oral proficiency. In contrast, vocabulary games were the least utilized feature, with only 8.3% of the students using them. These findings revealed that students used the ELSA Speak app primarily for practical and engaging learning strategies to increase their language proficiency. Consequently, this tool was found to effectively cater to diverse learning needs and strategies among EFL students.

Figure 4

Aspects that students practiced the most when using the ELSA Speak



Learners' evaluation of the benefits of AI-powered voice recognition via the ELSA Speak

The findings from the survey on students' evaluations of the usefulness and convenience of the ELSA Speak app are presented in Tables 3 and 4. The students were invited to score their opinions against a scale of 1 to 7, where 1 represented "strongly disagree" and 7 signified "strongly agree". With the range (7-1=6) and the difference (6/7=0.85), the intervals the authors used to interpret the results are: 1.00-1.85 (strongly disagree), 1.86-2.71 (disagree), 2.72-3.57 (somewhat disagree), 3.58-4.43 (neutral), 4.44-5.29 (somewhat agree), 5.30-6.15 (agree), and 6.16-7.00 (strongly agree).

Table 3

Mean-score of students' satisfaction with the usefulness of the ELSA Speak

	N	Mean	Std. Deviation	Std. Error Mean
Useful	37	5.30	1.47	.241

The analysis of the usefulness of the ELSA Speak app for students' pronunciation practice revealed a positive assessment. With a mean score of 5.30, the app appeared to be rated as moderately to highly beneficial by the users. This mean score suggested that students generally found the ELSA Speak app to be a valuable tool for enhancing their pronunciation skills. The relatively low standard deviation of 1.47 indicated consistent agreement among users regarding its effectiveness, although there was still some variability in individual experiences. Overall, the data support the notion that the ELSA Speak app is a useful resource for students looking to improve their pronunciation, aligning well with its intended purpose of providing effective practice and feedback.

Table 4

Meanscore of students' evaluations of the convenience of the ELSA Speak

	N	Mean	Std. Deviation	Std. Error Mean
Convenient	37	5.54	1.26	.207

On the basis of the students' evaluations, the convenience of using the ELSA Speak app was rated positively, with a mean score of 5.54 on a scale where higher values indicate greater convenience. The standard deviation of 1.26 suggests some variability in the students' responses, whereas the standard error mean (.207) indicates a reliable estimate of the mean. The result indicates that the students found the app highly convenient for their learning experiences.

Table 5 below presents the outcomes of the thematic analysis of the theme of positive experiences.

Table 5

Thematic analysis of students' positive experiences when using the ELSA Speak

Themes	Codes
Customization	- Individualized learning path and workload - Adjustability of listening speed - Flexibility of time-management - Thematic learning feature of the app
Self-improvement	- Opportunities for self-practice - The accessible pre and post-test scores for self-refection - Drills to practice pronunciation and intonation - Model conversations and dialogs for out-loud reading practice
AI for evaluation	- Realtime and automatic feedback - Visualized and quantified evaluation reports - Correction to users' pronunciation mistakes - Detailed check and feedback at individual sound and utterance level

After the experiment, several notable benefits of this app were acknowledged by the students. Among the several compelling benefits aimed at enhancing students' pronunciation, there has been a strong focus on the features of customization and self-improvement. The standout feature recognized by many students was the ability to create a customized and flexible learning path tailored to individual study conditions, such as available time and specific proficiency levels. They noted that this tool helps assess a student's ability to create a suitable learning path for them (S9, Student 34). The system was also highly evaluated for the adjustability of listening speed, allowing the students to choose an appropriate tempo of the voices. It was convenient, as they could pause, replay, and rewind when listening to the recording or talking in the app. Moreover, the ELSA Speak was found to be useful for providing topic-based learning, which allows students to focus on relevant vocabulary and phrases of the same theme and facilitates their memory.

The voice recognition technology of the app supported repeated listening and recording of one's speech to track improvements over time, and it also included detailed video lessons on phonetics and a structured learning path to help users progress systematically. The students commented that they could "have multiple repetitions and recordings" (Student 36) during their practice and "listen to correct pronunciation" (Students 8, S18, S24). This personalization ensured that the learners received practice and feedback relevant to their unique needs. In addition, with feedback provided by its unique voice recognition, the app also included robust self-reflection tools, such as recording and repeatedly listening to one's speech, which helped users identify and correct their mistakes effectively, and "*there were pre- and post-tests to see how much one has improved*" (Student 28). At both the sound and discourse levels, voice recognition by the ELSA Speak provided comprehensive practice opportunities, allowing users to focus on individual phonemes, sentence stress and intonation, and pronunciation within dialogs, which are vital for students' natural-sounding speech. Users could practice reading dialogs and receive evaluations on their pronunciation proficiency, and it also offered phonetic transcriptions, allowing learners to see correct pronunciations and gauge their accuracy. Student 29 stated:

> *There was assessment and instant feedback to read again correctly. There were full meanings and IPAs of vocabulary to learn, and I knew more about the meaning and pronunciation of the words. In addition to the exercises in*

> *class, the lessons from the ELSA Speak were also quite good. There were video lectures with detailed instructions on pronunciation (vowel, consonant, diphthong) along with the learning path. It was also OK.*

Additionally, the app leverages AI for evaluation, offering precise and immediate feedback on pronunciation, which helps learners track their progress and refine their skills over time. In particular, the app was confirmed to be capable of providing feedback on mispronunciations by highlighting specific errors and offering corrections, which was crucial for students' targeted learning. The students commented that the voice recognition technology helped "show me where I pronounced incorrectly" and "correct pronunciation mistakes" (Students 2, 10, 12, 27), as they could "see the transcription of the word and see how many percent of the word you pronounce correctly" (Student 13). The students also appreciated the app for its ability to check and provide detailed feedback at both the segmental and suprasegmental levels of speech.

Overall, with the distinct features of providing immediate feedback, phonetic details, and personalized learning paths, the app proved to be a valuable tool for effectively enhancing pronunciation skills. The ELSA Speak was realized by students to be a highly effective tool for improving pronunciation through personalized practice, self-assessment, and advanced technological support. Student 29 commented that there were "*full pronunciation lessons to learn. In general, the content of the app is quite perfect*."

Challenges for learners when using the ELSA Speak to practice pronunciation

The ELSA Speak app was acknowledged to provide valuable support for learners' pronunciation practices. Most of the students reported that they faced minimal difficulties or only minor issues with the app and found the app beneficial overall. However, several reported some limitations that could impact their experience, ranging from personal to technical effectiveness. Table 5 below presents some students' feedback on the drawbacks of their practice on the ELSA Speak app with its AI-powered voice recognition.

Table 6

Students' feedback on the drawbacks of practice with the ELSA Speak app

Themes	Codes
No drawbacks	Ideas expressing no or very little difficulties when using the app
Application's shortcomings	- Occasional incorrect transcription and missing words - Unstable sensibility to all voice differentiation - Navigating to users' voice quality regardless of other causes - To be evaluated, an error-identified speech needs to be completely recorded over - Occasional glitching and lagging of voice recognition features - Clearer instruction for some in-app activities
Personal self-efficacy and proficiency	- Lack of autonomy in learning - My pronunciation competence - Interaccent (accent shaped by mother tongue) - Need a lot of reminders from the teacher
Contextual factors	- Limit time for practice - Busy class schedule - Private and supportive space for self-practice - Personal device and equipment quality

Among the challenges mentioned in Table 6, a particular issue was related to recording quality. Users occasionally encounter technical problems where the app fails to capture their pronunciation accurately, leading to frustration and inconsistent feedback (students 1, 3, 4, 5, 9, 10, 14, 15, 19, 25, 29, 30, 31, 36). Student 12 shared that "*sometimes the app had an error. I pronounced the same thing repeatedly, but the app said it couldn't hear it though I was in a quiet space*." Other students (students 18, 29, 30, 31, and 36) also had the same issue of having to repeat multiple times for voice recognition to recognize their pronunciation, even when they read the same word as student 30 stated: "*I read the same word but sometimes ELSA could hear it and sometimes she could not*." This problem resulted in unnecessary delays and frustration, as they had to repeat their attempts multiple times and more additional practice to achieve a satisfying evaluation. Similarly, student 14 claimed that the ELSA Speak sometimes "*could not hear all the sounds clearly when I said, so I had to repeat many times*". This was also the reason why student 5 felt frustrated. From the participants' viewpoint, to make the application more effective, several features, such as microphone sensitivity and recording quality, need to be improved. Next, lag and technical glitches were also frequently noted, such as issues with recording and playback. Users (students 3, 19, and 28) reported problems with the app lagging or freezing, which disrupted their practice sessions and required them to restart the exercises. Additionally, some students believed that the app could be better able to support their understanding. They suggested incorporating clearer explanations, even in their mother tongue, to help users better grasp new vocabulary and pronunciation rules (students 11, 34).

Another significant limitation was related to self-efficacy and proficiency. Although the students found the app useful for their pronunciation practice, they found themselves not engaged enough in this process to progress and become more advanced (students 31 and 34). Owing to some students' low proficiency levels, the voice recognition of the app could not identify their pronunciation correctly. Because correct pronunciation could be difficult for students to achieve, student 16 noted the following:

> *Regarding my pronunciation ability, it was still too weak, and I couldn't pronounce it completely correctly. During the practice, I kept making the same mistakes, which made me very frustrated. In addition, sometimes I was a bit bored. I just kept repeating, but it was still incorrect, then I just quit.*

This could be explained by student 29's sharing that "*because it was also the first time I had learned in detail about sounds and pronunciation, I found pronunciation practice quite difficult. Some sounds were exceedingly difficult to read and required a lot of practice*". Students 7 and 18 considered their ability to be the main reason why voice recognition by the ELSA Speak could not realize their pronunciation. Similarly, student 3 believed that "*due to my ability, there were some words that took forever to pronounce*". Notably, many students highly evaluate the role of the teacher's regular reminders of tasks and self-practice. They admitted that their active hours with the app would not have been that high without frequent reminders from the teacher.

In addition, some contextual obstacles mentioned by the students were time constraints, class schedules, self-study space, and personal devices. Some students claimed to have so demanding schedules that allocating time for consistent practice with the ELSA Speak seemed challenging for them (students 1 and 6). In addition, some of them could not practice comfortably and effectively in the limited space of their dorms or with their own learning equipment and recording devices, which they found were not good enough.

In short, despite these hurdles, the benefits of the app were found to be significant – the ELSA Speak, with its voice recognition technology and personalized learning paths, catered to

individual needs. Its ability to provide repeated recordings and self-reflection opportunities was also highly appreciated. The app was found to support practice at both the sound and discourse levels, enhancing users' ability to improve pronunciation across various contexts. In addition, AI-driven evaluations offer precise feedback, further supporting students' progress. While there were occasional technical and personal challenges, the app remained a valuable tool for improving students' English pronunciation. By addressing these areas—enhancing microphone sensitivity, improving sound-recognition accuracy, minimizing lag, and adding more comprehensive language support—the usability and effectiveness of the ELSA Speak App could be significantly improved and aligned more closely with students' needs and expectations.

Discussion

The first research goal of the study was to measure the effects of employing voice recognition to increase the pronunciation proficiency of learners. The results from both the SPSS Pearson correlation analysis and the pretest–posttest evaluation revealed that students' progress in their pronunciation competence was closely related to the degree of practice with the AI tool. This positive correlation confirms the results of recent studies on applying technology to help enhance pronunciation instruction, such as Popenici and Kerr (2017), Wang and Liu (2019), Sujatha et al. (2020), Akhmad and Munawir (2022), Sariani et al. (2022), Kazu and Kuvvetli (2023), Zou et al. (2023), and Nicolau (2024). Interestingly, despite the teacher's classroom instructions and considerable amount of self-practice, some students' pronunciation performance did not improve. This demonstrates that, as noted by Fraser (2000) and Pourhosein (2016), acquiring native-like pronunciation is difficult. However, the results revealed considerable evidence of differences in the number of practices and engagement rates among different students in the class; these variations caused diversities in postcontinuous performance and final scores of the pronunciation learners. Based on the correlation analysis, this study also revealed that intensive and repeated practices on any drill led to a clear improvement in the pronunciation output of sounds, words, and sentences in those drills. In addition to the frequency of practice, the teacher's supervision appeared to play a crucial role in the recorded progress. These results imply that the quantity and attitude of practice improve fluency and accuracy in both segmental and suprasegmental levels of pronunciation.

The second research question assessed how the learners evaluate the implementation of AI-powered voice recognition in their pronunciation practice. In general, despite some challenges related to the learners, the application, and the learning context, most of the participants recommended employing such AI-powered technology to support pronunciation instruction and practice, which are aligned with learners' learning styles and progress. The standout features of the technology to create a customized and flexible learning path tailored to individual study conditions, with their different available times and specific proficiency levels, and to provide precise and immediate feedback on pronunciation performance to promote immediate self-correction are highly valued. These results are consistent with those of previous studies (Popenici & Kerr, 2017; Kazu & Kuvvetli, 2023; Zou et al., 2023; Nicolau, 2024), which indicate that instructors and learners favor the adoption of AI-supported technology for teaching and learning pronunciation. Furthermore, it was suggested that, by employing AI-powered voice recognition for pronunciation practice and instruction, the outputs of learning would be more fruitful and practical if technical issues, such as lagging, glitching, and quality of voice recording, could be improved and evaluation standards could be more flexible, recognizing the pronunciation with different English accents. In addition, what has not been reported in past studies, such as Sarmita Samad and Aminullah's (2019) and Anggraini's (2022) studies in the

ELSA Speak technology, is the influence of learners' confidence and motivation on the effectiveness of adopting AI technology. This study realized that students who struggled with self-discipline or felt less confident in their pronunciation found it harder to benefit fully from any of the technological features provided.

Conclusion

The study revealed several key findings regarding the use of AI-powered voice recognition tools to enhance pronunciation competence among EFL learners. The results demonstrated that students' progress in pronunciation was closely tied to the frequency and intensity of their practice with the AI tool. Intensive and repeated drills were found to significantly improve pronunciation accuracy at both the segmental and suprasegmental levels. However, individual differences in the number of practices and engagement levels contributed to variations in learners' performance. In addition, teacher supervision also played a critical role in guiding students' progress and ensuring consistent improvements. These findings highlight the significance of both the quantity of practice and the learner's attitude and active engagement in achieving fluency and accuracy in pronunciation.

The learners' feedback provided further insights into the practicality and potential of AI-powered tools in language learning. Most participants recommended the use of technologies such as tailored learning paths and precise, real-time feedback for self-correction for pronunciation practice. However, challenges such as technical issues, variability in learners' confidence and motivation, and the need for more flexible standards for different English accents were noted. These findings emphasize the potential of AI-powered tools to revolutionize pronunciation instructions, provided that their technical and contextual limitations are effectively addressed.

The findings of this study offer valuable implications for both future research and practical applications. Longitudinal studies are needed to investigate the long-term impacts of AI-powered tools on pronunciation retention and their effectiveness across diverse linguistic and cultural contexts. Additionally, further exploration into how learners' confidence, motivation, and self-discipline influence the outcomes of AI-supported learning would provide deeper insights into optimizing its implementation. For practice, educators and developers should focus on improving the technical aspects of AI tools, such as reducing lag, enhancing voice recording quality, and accommodating diverse English accents. Teachers should also be trained to integrate these technologies effectively into their instructional practices and provide adequate supervision to guide students' engagement and progress. By addressing these challenges, AI-powered tools have the potential to become a transformative resource for pronunciation instruction, fostering greater learner autonomy and improved language competence.

References

Anggraini, A. (2022). Improving students' pronunciation skill using ELSA Speak application. *Journey: Journal of English Language and Pedagogy*, 5(1), 135-141. https://doi.org/10.33503/journey.v5i1.1840

Akhmad, N. W., & Munawir, A. (2022). Improving the students' pronunciation ability by using ELSA Speak app. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, *10*(1), 846-857. https://doi.org/10.24256/ideas.v10i1.2868

Almalki, N., & Algethami, G. (2022). An exploration of the potential benefit of extensive listening along with orthography for improving EFL learners' pronunciation. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), 3. https://doi.org/10.1186/s40862-022-00132-x

Darsih, E., Wihadi, M., & Hanggara, A. (2021). *Using ELSA app in speaking classes: Students' voices*. Conference Paper. https://doi.org/10.4108/eai.12-12-2020.2304993

Do, L. M. (2021). The Role Pronunciation Teaching in EFL Classrooms: Beliefs and Practices of Teachers and Students. *International Journal of TESOL & Education*, *1*(1), 47-53. https://i-jte.org/index.php/journal/article/view/7

Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. S. (2016). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal, 44*(2), 181-196.

Fraser, H. (2000). *Coordinating Improvements in Pronunciation Teaching for Adult Learners of English as a Second Language*. Department of Education, Training and Youth Affairs, Canberra.

Gopalan, M., Rosinger, K., & Ahn, J. B. (2020). Use of Quasi-Experimental Research Designs in Education Research: Growth, Promise, and Challenges. *Review of Research in Education, 44*(1), 218-243. https://doi.org/10.3102/0091732X20903302

Hamada, Y. (2016). Shadowing: Who benefits and how? Uncovering a booming EFL teaching technique for listening comprehension. *Language Teaching Research, 20*(1), 35-42.

Hismanoglu, M. (2006). Current Perspectives on Pronunciation Learning and Teaching. *Journal of Language and Linguistic Studies, 2*(1), 101-110.

Johnson, B & Christensen, L. B. (2019). *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. Sage Publication.

Kok, J. N., Boers, E. J., Kosters, W. A., Van der Putten, P., & Poel, M. (2009). Artificial intelligence: definition, trends, techniques, and cases. *Knowledge for sustainable development: an insight into the Encyclopedia of life support systems, 1*, 1095-1107.

Kazu, İ. Y., & Kuvvetli, M. (2023). The influence of pronunciation education via artificial intelligence technology on vocabulary acquisition in learning English. *International Journal of Psychology and Educational Studies*, *10*(2), 480-493. https://doi.org/10.52380/ijpes.2023.10.2.1044

Marsden, E., & Torgerson, C. J. (2012). Single group, pre- and posttest research designs: Some methodological concerns. *Oxford Review of Education, 38*(5), 583-616. https://doi.org/10.1080/03054985.2012.731208

Moxon, S. (2024). A Review of the Website Pronounce. com. *Computer-Assisted Language Learning Electronic Journal*, *25*(3), 195-212. https://callej.org/index.php/journal/article/view/22/397

Ngoc, N. K., Thanh, N. T., & Thanh, N. T. (2023). Tertiary students' perceptions on ELSA Speak application for pronunciation learning. *European Journal of Applied Linguistics Studies*, *7*(1), 1. https://doi.org/10.46827/ejals.v7i1.483

Nhu, N. T. T., Hao-Jan C. H, Kuo-Wei L. K. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL, 36*(1), 4-21. https://doi.org/10.1017/S0958344023000113

Nicolau, M. (2024). Artificial intelligence – Friend or foe. *IPI Letters*, (2) 1, 34-41. https://doi.org/10.59973/ipil.54

Noviyanti, S. D. (2020). Artificial intelligence (ai)-based pronunciation checker: An alternative for independent learning in pandemic situation. *ELT Echo : The Journal of English Language Teaching in Foreign Language Context*, *5*(2), 162. https://doi.org/10.24235/eltecho.v5i2.7246

Nushi, M., & Sadeghi, M. (2021). A critical review of ELSA: A pronunciation app. *Computer-Assisted Language Learning Electronic Journal*, *22*(3), 287-302. https://callej.org/index.php/journal/article/view/365

Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1). https://doi.org/10.1186/s41039-017-0062-8

Pourhosein Gilakjani, A. (2016). What Factors Influence the English Pronunciation of EFL Learners? *Modern Journal of Language Teaching Methods (MJLTM)*, 6(2), 314-326.

Privitera, G& Delzell. (2019). *Quasi-Experimental and Single-Case Experimental Design Chapter Thirteen*. California: Sage Publication Inc.

Sarmita Samad, I., & Aminullah, A. (2019). Applying ELSA speak software in the pronunciation class: Students' perception. *Edumaspul - Jurnal Pendidikan*, 3(1), 56-63. https://doi.org/10.33487/edumaspul.v3i1.85

Sariani, S., Miladiyenti, F., Rozi, F., Haslina, W., & Marzuki, D. (2022). Incorporating mobile-based artificial intelligence to English pronunciation learning in tertiary-level students: Developing autonomous learning. *International Journal of Advanced Science Computing and Engineering*, *4*(3), 220-232. https://doi.org/10.30630/ijasce.4.3.92

Sujatha, B.; Vanajakshi, B. and Nirmala, K. N. (2020). SALA-An Integrated Framework for Speech Recognition Using Lexical Analyzer, *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Bengaluru, India, 386-390. https://doi.org/10.1109/ICSTCEE49637.2020.9276996

Wang, Y., & Liu, H. (2019). The application of AI in language learning: A case study of an AI-powered language learning application. *International Journal of Distance Education Technologies, 17*(1)*,* 1-14.

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (ai)-assisted language learning. *Sustainability*, *15*(4), 2872. https://doi.org/10.3390/su15042872

Biodata

Dr. Nguyen Truong Sa is the Dean of Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Vietnam. His teaching and research favorites and publication are related to quality assurance in language teacher education, English as a medium of instruction, and applied technology in language teaching.

Ms. Thi Nguyen has been lecturing at Industrial University of Ho Chi Minh City for 14 years for courses related to English pronunciation, English speaking skills, and foreign language teaching methodology. Currently, she is conducting her PhD training at Hue University, and her

research interests are mobile-assisted/ technology-assisted language learning, autonomous and self-directed/ self-regulated learning, and mobile learning resources.

Ms. Hoang Ngoc Quynh Nhu is an M.A. in language teaching; she is currently teaching courses in Translation, Research Writing Skills, and pronunciation in a Bachelor in English language program.

Ms. Do Thi Kim Hieu is an M.A. in language teaching. Ms. Hieu has spent over 15 years teaching courses in English phonetics and phonology, and pronunciation to undergraduate students in English language.