Journal of Education and Teaching (JET)
Vol. 6 No. 2 2025

ELSA Speak as a Speaking Partner: A Mixed-Methods Study in an Indonesian Higher Education Context

Luh Indrayani[1], Ni Putu Astiti Pratiwi[2], I Ketut Resika Arthana[3]
English Language Education, Universitas Pendidikan Ganesha, Indonesia[1,2]
Computer Science, Universitas Pendidikan Ganesha, Indonesia[3]
Corresponding Email: luh.indrayani@undiksha.ac.id

Draft article history
Submitted: 08-05-2025;
Revised: 14-06-2025;
Accepted: 15-06-2025;

ABSTRACT: Developing effective speaking proficiency, particularly in pronunciation, accuracy, and fluency**,** remains a persistent challenge for EFL learners in Indonesian higher education. In response, this study investigates the impact of integrating ELSA Speak**,** a voice-based AI application, on enhancing these core aspects of speaking performance. Employing an explanatory sequential mixed-methods design, the research involved 23 Indonesian university students over an eight-week intervention. Quantitative data were collected through pre-tests and post-tests using CEFR-aligned speaking assessments, with paired-samples t-tests employed to examine the statistical significance of performance gains. Test validity was ensured through expert review and alignment with internationally recognized rubrics. Results indicated a statistically significant improvement in students' overall speaking scores, particularly in pronunciation and accuracy. To complement these findings, qualitative data were gathered through semi-structured interviews and analyzed thematically. Students reported that ELSA Speak provided constructive, personalized feedback**,** context-rich materials, and a flexible, motivating learning environment. However, limitations were noted, including restricted access to premium features**,** internet connectivity issues**,** and low self-regulation, especially in managing independent learning. These findings underscore both the pedagogical benefits and practical challenges of integrating AI-powered tools like ELSA Speak into EFL speaking instruction. The study offers practical insights into how AI-powered tools can be meaningfully integrated into speaking instruction. Theoretically, it contributes to the discourse on AI-mediated language learning**,** highlighting the need to balance technological affordances with learner readiness and contextual constraints.

Keywords: artificial intelligence, EFL, ELSA Speak, speaking skill.

ABSTRAK: *Mengembangkan kemampuan berbicara yang efektif, khususnya dalam aspek pelafalan, akurasi, dan kefasihan, tetap menjadi tantangan yang berkelanjutan bagi pembelajar Bahasa Inggris sebagai Bahasa Asing (EFL) di pendidikan tinggi di Indonesia. Sebagai respons, penelitian ini mengkaji dampak integrasi ELSA Speak, sebuah aplikasi berbasis kecerdasan buatan (AI) dengan teknologi pengenal suara, dalam meningkatkan aspek-aspek utama dari performa berbicara tersebut. Penelitian ini menggunakan desain metode campuran sequential explanatory, yang melibatkan 23 mahasiswa dari sebuah universitas di Indonesia selama delapan minggu intervensi. Data kuantitatif dikumpulkan melalui pre-test dan post-test menggunakan asesmen berbicara yang diselaraskan dengan standar CEFR, dan dianalisis menggunakan uji t berpasangan untuk menguji signifikansi statistik dari peningkatan performa. Validitas tes dijamin melalui tinjauan ahli dan kesesuaian dengan rubrik yang diakui secara internasional. Hasil penelitian menunjukkan adanya peningkatan yang signifikan secara statistik pada skor berbicara keseluruhan mahasiswa, terutama pada aspek pelafalan dan akurasi. Untuk melengkapi temuan kuantitatif, data kualitatif dikumpulkan melalui*

wawancara semi-terstruktur dan dianalisis secara tematik. Para mahasiswa melaporkan bahwa ELSA Speak memberikan umpan balik yang konstruktif dan dipersonalisasi, materi yang kaya konteks, serta lingkungan belajar yang fleksibel dan memotivasi. Namun, beberapa keterbatasan juga diidentifikasi, termasuk akses terbatas ke fitur premium, masalah konektivitas internet, dan rendahnya kemampuan pengaturan diri, khususnya dalam mengelola pembelajaran mandiri. Temuan ini menyoroti manfaat pedagogis sekaligus tantangan praktis dalam mengintegrasikan alat berbasis AI seperti ELSA Speak ke dalam pembelajaran berbicara Bahasa Inggris. Studi ini memberikan wawasan praktis tentang bagaimana alat berbasis AI dapat diintegrasikan secara bermakna dalam pengajaran berbicara. Secara teoretis, penelitian ini turut berkontribusi pada diskursus tentang pembelajaran bahasa berbasis AI, dengan menekankan pentingnya menyeimbangkan potensi teknologi dengan kesiapan pembelajar dan kendala kontekstual.

Kata kunci: *EFL, ELSA Speak, kecerdasan buatan, keterampilan berbicara.*

INTRODUCTION

The ability to communicate fluently and accurately in spoken English is essential for learners aiming to succeed in academic and professional environments (Dhimolea et al., 2022; Rao, 2019). Yet, many EFL learners, particularly in Indonesian university settings continue to face persistent difficulties in developing effective speaking skills (Hijra et al., 2024). These challenges often stem from limited English exposure outside the classroom, speaking anxiety, and insufficient corrective feedback (Mary & Paul, 2024). In Indonesia, structural issues such as large class sizes and exam-oriented instruction further reduce opportunities for meaningful speaking practice (Larson, 2014; Winnie et al., 2023). To fully grasp the complexity of speaking instruction, it is essential to view speaking not merely as a mechanical skill, but as an interactive process involving the production, reception, and processing of meaning (Burns & Joyce, 1997). According to Thornbury (2005) there are several criteria that must be mastered so that students can be said to be fluent in speaking English, such as: fluency, accuracy, vocabulary, grammar and pronunciation. Furthermore, the availability of contextual materials are also prominent to support students' speaking performance (Erniwati et al., 2024; Lukman, 2024; Tomlinson, 2011). For example, Lukman (2024) designed and implemented a contextual learning model for middle school students, and it was effective to improve students' speaking competence. Erniwati et al., (2024) strengthen the findings that contextual language teaching is highly effective in improving students' speaking ability by making learning more meaningful, connect to real life context and increase students' engagement.

However, research consistently shows that a majority of EFL learners still fall short in these areas and are unable to communicate orally with confidence and competence (Zhang et al., 2021). In response to these challenges, the integration of Artificial Intelligence (AI) offers new opportunities to support students' speaking development (Mingyan et al., 2025; Qassrawi et al., 2024). AI is when computers are designed to do things that usually need human thinking, like learning, problem-solving, or making decisions (Saleh, 2019). In particular, generative AI, a subfield that enables the creation of new content, namely: audio, code, images, text, simulations, and videos to facilitate learner interaction (Feuerriegel et al., 2024).

Research has increasingly demonstrated the potential of Generative AI

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

applications to enhance various aspects of language learning (Duong & Suppasetseree, 2024; Han, 2020; Kim & Kim, 2021; Wu et al., 2025). For instance, Phanwiriyarat et al., (2025) found that AI-powered tools like Duolingo can improve vocabulary, pronunciation, and listening skills, while its gamified features help lower anxiety and boost learner engagement. Similarly, Rad (2024) reported that using Speeko significantly enhanced learners' speaking proficiency and willingness to communicate (WTC). Supporting this, Fathi et al., (2024) highlighted that using AI to support interactice speaking activities helped EFL learners improve their speaking skills and willingness to communicate better than traditional teaching methods.

While previous studies have demonstrated the benefits of AI-powered tools such as: Speeko and Dulinggo in enhancing speaking skill, each focuses on specific areas, namely: public speaking delivery, text-based conversational practice or gamified vocabulary learning. These applications provide valuable support but often offer limited integration of multiple speaking components, namely: pronunciation, fluency, grammar and vocabulary within real time speech-based practice. In contrast, ELSA Speak is an AI speech recognition application designed to provide comprehensive, real-time feedback on multiple dimensions of spoken English, including pronunciation, intonation, and fluency (Akhmad & Munawir, 2022). What sets ELSA apart from similar tools is its integration with the Common European Framework of Reference (CEFR), offering standardized assessments across A1 to C2 levels (Council of Europe, 2001). Additionally, its mobile learning format enables students to practice anytime and anywhere, increasing exposure and autonomy (Zhang et al., 2021). ELSA's adaptive feedback and structured modules make it a promising tool for classroom and independent learning. However, previous research has also identified barriers such as limited access to premium features, technical issues, and the need for strong self-regulated learning skills to use such tools effectively (Baars et al., 2022; Garzón et al., 2024).

Despite its pedagogical potential, empirical research on ELSA Speak remains limited, especially in university-level EFL contexts in Indonesia. In particular, studies that employ robust mixed-methods approaches combining quantitative performance data with qualitative insights into learner experience are notably scarce. Addressing this gap, the present study investigates both the effectiveness of ELSA Speak in improving students' speaking skills and the perceptions of learners regarding its use within a structured classroom setting. Accordingly, the following research questions were posed to guide the investigation: 1) To what extent does the use of ELSA Speak improve students' speaking skills, as measured by pre-test and post-test scores? 2) How do students perceive their experiences and challenges in using ELSA Speak to improve their speaking performance?

RESEARCH METHOD

Research Design

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

This study employed a sequential explanatory mixed-methods design (J. Creswell, 2012) in which quantitative data collection and analysis were conducted first, followed by qualitative data collection to further explain or elaborate on the quantitative results (Creswell & Clark, 2007). The quantitative phase used a pre-experimental one-group pre-test–post-test design to evaluate students' speaking performance after using ELSA Speak. The qualitative phase consisted of semi-structured interviews aimed at exploring students' perceptions and experiences in greater depth. The independent variable was the use of ELSA Speak, and the dependent variable was the students' speaking performance.

Variables and Operational Definition

The independent variable in this study was the use of ELSA Speak, a mobile AI-powered application for English speaking practice. Conceptually, ELSA Speak supports learners in improving pronunciation, fluency, and vocabulary through AI-driven speech recognition (Akhmad & Munawir, 2022). Operationally, ELSA Speak use was defined by students' engagement with its core features: speech recognition, pronunciation scoring, vocabulary training, CEFR-aligned levels, and fluency evaluation. Usage intensity was monitored through structured in-app tasks and weekly class integration over an eight-week period.

The dependent variable was students' speaking performance. Conceptually, this refers to learners' ability to communicate effectively in spoken English, including fluency, accuracy, vocabulary, grammar, and pronunciation (Thornbury, 2005). Operationally, speaking performance was assessed through pre- and post-tests consisting of five-minute oral presentations, evaluated using a validated rubric based on Thornbury (2005). Each of the five dimensions was rated on a five-point scale by trained raters to ensure consistency and reliability.

Researcher's Role and Bias Control

The researcher acted as both the instructor and data collector. To minimize bias, peer observation was conducted by a research colleague who is also a co-author of this study. This peer observer attended several assessment sessions and provided independent notes and feedback to ensure scoring fairness and procedural consistency. Furthermore, peer debriefing was used to reduce interpretive bias during qualitative analysis, as proposed by (Lincoln et al., 1985). The researcher engaged in regular analytical discussions with the peer reviewer to cross-check emerging themes and interpretations.

Research Participants

This study involved 23 undergraduate EFL students who were selected to examine the impact of ELSA Speak on their speaking performance. The same participant group was involved in a complementary research project focusing on learning motivation; however, the present study specifically investigates speaking skill. Purposive sampling was used to select participants with relevant characteristics, namely, those who had access to smartphones, basic English proficiency, and were willing to participate in the full eight week learning period.

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

Prior to participation, students were informed about the research purpose and voluntarily provided consent. Students could withdraw at any time without penalty. Measures were taken to ensure confidentiality and anonymity, including assigning codes to participant data and securely storing all records. Therefore, it is expected that these participants would provide valuable insight (Kelly, 2010). Over eight weeks, participants engaged in structured ELSA-based speaking tasks such as business presentations and workplace simulations, designed to elicit students' speaking proficiency. The topics were about role play of conducting meetings, role play of conducting interviews, role play of talking on the telephone, role play of handling complaints and business presentation, like: presenting a company product. At the end of the trial week, the students were given a post-test where they will present a business company product by using all the vocabulary and knowledge that they got during the experimental periods.

Instruments

The data were collected through pre-tests, post-tests, and semi-structured interviews. The pre-tests and post-tests involved a five-minute oral presentation on topics related to Business English. To ensure comparability, both tests were designed with similar characteristics in terms of topic complexity, structure, and assessment criteria. Then a speaking scoring rubric adopted from Thornbury (2005) was implemented, which included criteria like: fluency, accuracy, vocabulary, grammar and pronunciation. The tests have been validated by two ELT experts. Revisions were made to clarify the performance indicators and ensure alignment with the learning objectives. Afterwards, the students were interviewed by using eight-question guide to elicit students' experiences with the ELSA Speak integration. The interview questions focused on how ELSA Speak improved students' speaking skills. Prior to conducting the interview, the interview guide underwent a content validity assessment by two ELT experts. Revisions were made based on clarity, construct relevance, and alignment with the research objectives. The interview results were written down as transcripts that included students' responses about their speaking skills while learning supported by ELSA Speak.

Data Analysis

The data collection procedure began with informing the participants that they would be involved as research participants. After getting their permission, the researcher told them about the learning procedures that will last for eight weeks and introduced the ELSA Speak. Then the students began the speaking exercises related to the topic accompanied with ELSA Speak as can be seen on the class activities in (Table 1).

Table 1. Activities with ELSA Speak

Meetings	Topic	Activities
1	Pre-test	Students' activities
2	Role play of conducting meeting	Reviewing material Watching video

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

		Reviewing relevant vocabularies Practicing with Elsa: having conversation with Elsa related to the topic of conducting meeting, this include reviewing vocabularies, pronunciation and content Students' got a clue for the conversation that should be made with their peers Practicing in front of the class Teacher gave feedback
3	Role play of conducting interview	Reviewing material Watching video Reviewing relevant vocabularies Practicing with Elsa: having conversation with Elsa related to the topic of conducting interview, this include reviewing vocabularies, pronunciation and content Students' got a clue for the conversation that should be made with their peers Practicing in front of the class Teacher gave feedback
4	Role play of talking on the telephone	Reviewing material Watching video Reviewing relevant vocabularies Practicing with Elsa: having conversation with Elsa related to the topic of talking on the telephone, this include reviewing vocabularies, pronunciation and content Students' got a clue for the conversation that should be made with their peers Practicing in front of the class Teacher gave feedback
5	Role play of handling complaints	Reviewing material Watching video Reviewing relevant vocabularies Practicing with Elsa: having conversation with Elsa related to the topic of handling compaints, this include reviewing vocabularies, pronunciation and content Students' got a clue for the conversation that should be made with their peers Practicing in front of the class Teacher gave feedback
6 & 7	Presenting a company product	Reviewing material Watching video

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

Journal of Education and Teaching (JET) Volume 6 No. 2, 2025
DOI: 10.51454/jet.v6i2.649

		Reviewing relevant vocabularies Practicing with Elsa: having conversation with Elsa related to the topic of business product presentation, this include reviewing vocabularies, pronunciation and content Students' got a clue for the conversation that should be made with their peers Practicing in front of the class Teacher gave feedback
8	Post-test	Students' activities

To visualize the research procedure, Figure 1 below presents the overall flow:

Figure 1. Research Procedure Flowchart

This flowchart outlines the main phases of the research: pre-test, daily ELSA Speak interventions, post-test, interviews, and data analysis. The students were asked to perform either a role play or presentation based on the topic. Before presenting, they should have a chance to practice relevant vocabularies, pronunciation and conversation with ELSA. At home, the students also engaged with the ELSA Speak and all of their progress was recorded on the application (Figure 2).

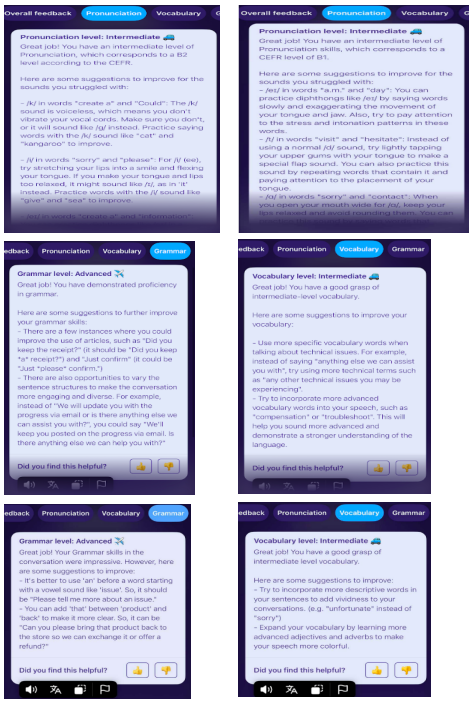

Figure 2. Student progress on ELSA Speak

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

Quantitative data were analyzed using descriptive and inferential statistics, including normality, homogeneity, paired-samples t-test, and effect size (eta squared). Moreover, SPSS 29 for Mac was used to examine the quantitative data. Meanwhile, the qualitative data from the interview guide were analyzed thematically, following the six-phase framework of Braun & Clarke (2006), which includes: (1) familiarization with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report.

RESULT AND DISCUSSION

Result: Students' Speaking Skill

The descriptive statistical analysis revealed a difference in students' speaking skills between the pre-test and post-test. The mean score on the pre-test was 72.87, categorized as good, while the post-test mean increased to 79.83, belongs to the very good category. This indicates an improvement in speaking performance following the integration of ELSA Speak. Supporting this, the pre-test scores ranged from 58.00 to 82.00, whereas the post-test scores ranged from 70.00 to 88.00, showing higher minimum and maximum values after the intervention. While these results suggest positive changes, further analysis using inferential statistics, specifically a paired-samples t-test, was conducted to determine the statistical significance of the differences observed.

Table 2. Descriptive Statistics

Descriptive Statistics								
	N	Range	Minimum	Maximum	Sum	Mean	Std. Deviation	Variance
Pretest	23	24	58	82	1676	72.87	7.430	55.209
Posttest	23	18	70	88	1836	79.83	5.458	29.787
Valid N (listwise)	23							

Table 3. The Theoretical Ideal Reference Assessment Criteria

No	Interval	Category
1	72 < X	Very Good (VP)
2	60 ≤ X < 72	Good (P)
3	48 ≤ X <60	Fair (SP)
4	36 ≤ X < 48	Poor (N)
5	X < 36	Very Poor (VN)

Note:
MI = 1/2 (ideal maximum score + ideal minimum score)
SDI = 1/6 (ideal maximum score - ideal minimum score)

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

Journal of Education and Teaching (JET) Volume 6 No. 2, 2025
DOI: 10.51454/jet.v6i2.649

Before conducting inferential statistical analysis, preliminary tests for normality and homogeneity were performed to confirm that the data met the required assumptions. The outcomes of these tests are presented in Table 4. The Shapiro-Wilk normality test produced significance values of 0.116 for the pre-test and 0.096 for the post-test, both of which exceed the α level of 0.05. These results indicate that the data for both tests are normally distributed.

Table 4. Normality Test

Tests of Normality							
	pre_post	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Nilai	Pre-test	.136	23	.200*	.931	23	.116
	Post-test	.153	23	.176	.927	23	.096
*. This is a lower bound of the true significance.							
a. Lilliefors Significance Correction							

Once the data was confirmed to be normally distributed, the next step was to check whether the data had consistent variance. The homogeneity test showed a significance value of 0.068, which is higher than the threshold of 0.05. This means the data met the requirement for equal variance, so it was appropriate to continue with the paired t-test analysis.

Table 5. Homogeneity Test

Test of Homogeneity of Variance					
		Levene Statistic	df1	df2	Sig.
Nilai	Based on Mean	3.491	1	44	.068
	Based on Median	2.730	1	44	.106
	Based on Median and with adjusted df	2.730	1	43.252	.106
	Based on trimmed mean	3.381	1	44	.073

After being tested using a paired t-test, it was found that the data significance value was 0.001. This data shows that the significance is smaller than α 0.05. So it can be concluded that the difference in pre-test and post-test scores is significant.

Table 6. Paired t test

Paired Samples Test										
		Paired Differences					t	df	Significance	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				One-Sided p	Two-Sided p
					Lower	Upper				
Pair 1	pretest - posttest	-6.957	3.037	.633	-8.270	-5.643	-10.984	22	<.001	<.001

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

Journal of Education and Teaching (JET) Volume 6 No. 2, 2025
DOI: 10.51454/jet.v6i2.649

Since this study used a paired t-test for analysis, the effect size was also calculated using eta squared to see how strong the impact of ELSA Speak was on students' speaking skills. The result was 0.846, which is considered a very large effect. This means ELSA Speak had a strong and meaningful influence on improving students' speaking abilities.

$$\text{eta-Squared} \quad \frac{(-10.984)^2}{(-10.984)^2 + (23 - 1)}$$

Students' perceived experiences and challenges

Apart from using tests to assess students' speaking skills, interviews were also conducted using a convenience sampling technique to confirm the data obtained from the tests regarding the effectiveness of ELSA Speak on students' speaking skills. Three participants agreed to be interviewed by researchers in this study. Below is the thematic analysis based on the interview results.

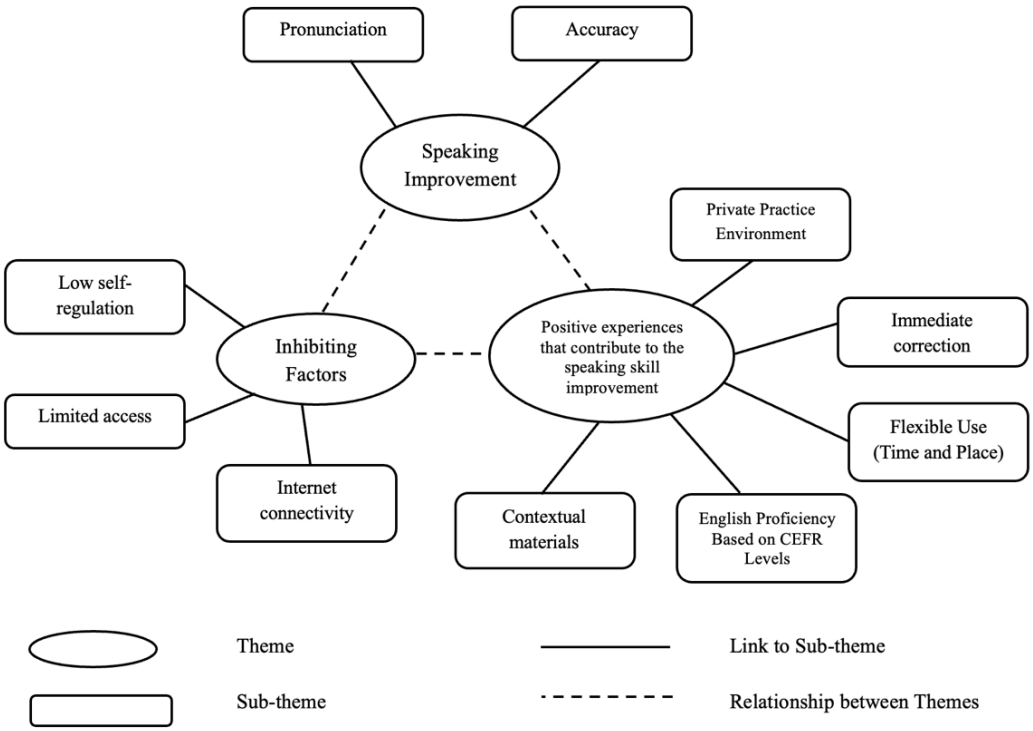

Figure 3. Thematic analysis

Based on the interview results, three participants said that their speaking skills had improved significantly in several aspects, such as: pronunciation and accuracy as quoted from the transcript below: "ELSA Speak has had an impact on my speaking skills, especially the pronunciation part, because I often practice pronunciation by topic" (Participant 1). While Participant 3 added that the accuracy is also improved:

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

> "Here are two aspects of speaking that show the most noticeable improvement, namely: pronunciation and accuracy because the app clearly tells us what is appropriate and what we are saying. It's very helpful. What makes it even more interesting is the complete feedback we receive" (Participant 2)

In addition to students' speaking development, they also addressed their experiences with ELSA which contribute to their pronunciation and accuracy development. Participant 2 mentioned that the private practice environment was the positive side from using AI: "I feel that the supportive and private learning environment make me feel comfortable in doing speaking practices with ELSA" (Participants 2). Participant 1 also mentioned similar thing "When I do speaking practices with ELSA, the environment makes me want to practice more and more because I can freely express my ideas without being afraid of being judged by other people" (Participant 1). All the three participants agreed that the immediate correction given by ELSA was the thing that help them a lot in improving their speaking skill. "The reason I enjoy using ELSA is because it tells me right away what I mispronounced. It gives feedback that feels personal and clear, so every time I practice, I learn exactly what to fix." (Participant 3). Another positive experience with ELSA was its flexibility. "I can do the speaking practices everywhere and every time, it is not limited by only the traditional classroom setting only" (Participant 2). The participants also addressed about their English proficiency level assessment. "Every time when I finished one practice, besides giving me immediate correction, it can assess my speaking proficiency level based on CEFL level which is good. So, I know my current level of English" (Participant 1). Lastly, all participants compliment the various contextual materials provided by ELSA Speak. "The materials provided by ELSA Speak is contextual which make me becoming even more enthusiastic in doing the practice because I know that this skill will be beneficial for me in the future" (Participant 3).

However, students also addressed the inhibiting factors about ELSA Speak for some reasons. "I felt limitations from this ELSA Speak application since the features that can be accessed per day are limited because the participants use the free version" (Participant 1)". Participant 2 mentioned different reasons related to self-regulated learning. "Since I use ELSA as my speaking partner at home, where there is no supervision from my lecturers, sometimes I skip using it, but if I use it every day, I believe that my speaking skills can be improved way better" (Participant 2)". Also, Participant 3 added about internet connectivity. "I think internet connectivity is one of the pitfalls of ELSA Speak because we cannot access it offline" (Participant 3)".

Discussion: Speaking Skill Improvement

This study investigated the effectiveness of ELSA Speak in enhancing the speaking skills of EFL learners. Guided by Creswell's (2012) mixed-methods framework, the findings revealed a statistically significant improvement in students' speaking performance, with an eta-squared coefficient of 0.846. An eta-

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

squared value of 0.846 suggests that approximately 84.6% of the variance in post-test scores can be associated with the intervention. This represents a very large effect size, indicating a substantial impact of ELSA Speak on students' speaking performance within the scope of this study. This finding further supports Han (2020), who found that AI chatbots enhanced EFL learners' speaking proficiency and affective engagement in Korea. Unlike Han's study, which focused more on chatbot interactions, this research emphasizes structured pronunciation, accuracy, fluency-based practice.

Similarly, Duong & Suppasetseree (2024) reported gains in grammar and vocabulary using Generative AI. This study extends their findings by demonstrating significant improvements in fluency, pronunciation, and overall speaking performance through real-time, individualized feedback provided by ELSA Speak. Qualitative data further confirmed these improvements. Students highlighted better accuracy and pronunciation after the intervention, which aligns with (Fathi et al., 2024), who emphasized AI-mediated speaking tasks' impact on both performance and willingness to communicate. These findings suggest that ELSA Speak's interactive features can meaningfully support learners' active engagement and speaking development.

Learning Experience and Pedagogical Insights

Students expressed positive emotional responses in using ELSA Speak. They appreciated the immediate, constructive feedback and the private learning environment that reduced their anxiety. These affective benefits align with Wu et al., (2025), who showed that AI-supported Think Pair Share lowered learners' Foreign Language Speaking Anxiety (FLSA) and increased enjoyment. This study confirms that mobile apps can also create low-anxiety environments, showing that such supportive conditions can be achieved even in self-paced and independent learning settings.

Participants also reported high engagement with contextualized materials in ELSA Speak, which mirrored real-life speaking situations. This study reinforces findings by Lukman (2024)and Erniwati et al., (2024), who showed that contextualized learning increases learner engagement. This study extends their findings by showing that mobile-assisted language learning tools like ELSA Speak can also support contextual engagement, even within blended or semi-autonomous learning formats.

A notable feature of ELSA Speak is its alignment with the CEFR framework. Beyond just design, some students reported perceivable progress through CEFR levels, with several noting improvement from A2 to B1 levels based on the app's internal diagnostics. This observation supports (Zaki & Darmi, 2021), who argued that CEFR helps learners set structured and attainable goals. This study confirms that such alignment can be operationalized in app-based learning environments, offering learners feedback that is both standardized and motivating.

Limitations

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

In terms of tool-related limitations, several challenges were noted with the use of ELSA Speak. These include restricted access to premium features that limited some students' ability to engage with full content, occasional internet connectivity issues that disrupted practice sessions, and low self-regulated learning among certain learners. In addition, this study was conducted at a single university in Bali, Indonesia, within a specific institutional and cultural context. The language exposure, digital readiness, and curriculum focus may differ significantly across regions. Thus, the findings offer context-specific insights and should not be generalized without careful consideration.

Implications and Future Research

The findings suggest that ELSA Speak can function as a semi-autonomous co-instructor, supporting CEFR-based progression while maintaining learner agency in a low-stress environment. Educators might consider integrating ELSA Speak as a supplementary self-practice platform aligned with in-class speaking targets. Training students in self-regulation strategies and providing institutional access to premium features may further enhance its effectiveness. Future research could explore longitudinal gains, multi-site implementations, or cross-national comparisons. Studies using larger, more diverse samples and measuring long-term language retention will offer more generalizable conclusions.

CONCLUSION

This study concludes that ELSA Speak positively contributes to enhancing EFL students' speaking proficiency. Quantitative results demonstrated a statistically significant improvement in post-test performance, while qualitative insights indicated gains in pronunciation accuracy and fluency. Participants reported improvements in articulating ideas and expressing themselves more clearly, attributing these advancements to key features such as speech recognition, real-time corrective feedback, contextualized practice, and CEFR-aligned assessments. Beyond evaluating a single tool, this research contributes to the broader discourse on AI-assisted language learning by illustrating how structured and personalized feedback delivered through mobile platforms can enhance speaking instruction. In particular, the CEFR integration in ELSA Speak not only facilitated learner progress tracking but also supported pedagogical alignment with internationally recognized proficiency standards. For practitioners, AI-powered tools like ELSA Speak can be effectively used as supplementary platforms to extend classroom practice. Educators are encouraged to incorporate these tools alongside in-class speaking targets and offer students guidance in goal setting and independent practice. Institutions, in turn, should consider investing in digital literacy programs and equitable access to educational technologies to maximize the benefits of AI-enhanced learning environments.

Future research should explore long-term effects of such tools on learner progress, investigate their adaptability across diverse contexts, and examine how different levels of CEFR gains are achieved in varied instructional models. This will

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

help shape more robust frameworks for integrating AI into EFL pedagogy across global settings.

ACKNOWLEDGMENT

This research was made possible with the generous support of Universitas Pendidikan Ganesha through a research grant. The authors are deeply grateful to all participants for their meaningful involvement and contributions to the study.

REFERENCES

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

Baars, M., Zafar, F., Hrehovcsik, M., de Jongh, E., & Paas, F. (2022). Ace Your Self-Study: A Mobile Application to Support Self-Regulated Learning. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.793042

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

Burns, A., & Joyce, H. (1997). *Focus on Speaking*. National Centre for English Language teaching and Research.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*. Cambridge University Press. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Creswell, J. (2012). *Planning, Conducting, and Evaluating Quantitative and Qualitative Research* . Pearson.

Creswell, J. W., & Clark, P. (2007). *Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage: Vol. 12(4)*. Sage.

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A Systematic Review of Research on High-Immersion Virtual Reality for Language Learning. *TechTrends*, *66*(5), 810–824. https://doi.org/10.1007/s11528-022-00717-w

Duong, T., & Suppasetseree, S. (2024). The Effects of an Artificial Intelligence Voice Chatbot on Improving Vietnamese Undergraduate Students' English Speaking Skills. *International Journal of Learning, Teaching and Educational Research*, *23*(3), 293–321. https://doi.org/10.26803/ijlter.23.3.15

Erniwati, E., Mertosono, R. S., Arid, M., & Gunawan, G. M. (2024). A Systematic Literature Review of Contextual Teaching and LearningApproach in EFL Speaking Skills. *Multicultural Education*, *10*(8), 4–8.

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, *121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Garzón, J., Kinshuk, Burgos, D., & Tlili, A. (2024). Advantages and challenges associated with mobile learning in education: a systematic literature review. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-024-00342-x

Han, D.-E. (2020). The Effects of Voice-based AI Chatbot on Korean EFL Students' Speaking Ability and Affective Factors. *International Journal of Computer Science and Information Technology for Education*, *5*(1), 25–30. https://doi.org/10.21742/IJCSITE.2020.5.1.03

Hijra, H., Rahim, T. R., & Syarif, A. R. (2024). Barriers to Oral Proficiency: Exploring Causes of Speaking Inhibition in EFL Learners. *KLASIKAL : JOURNAL OF EDUCATION, LANGUAGE TEACHING AND SCIENCE*, *6*(2), 409–418. https://doi.org/10.52208/klasikal.v6i2.1168

Kelly, S. E. (2010). Qualitative Interviewing Techniques and Styles. In *The SAGE Handbook of Qualitative Methods in Health Research* (pp. 307–326). SAGE Publications Ltd. https://doi.org/10.4135/9781446268247.n17

Kim, H.-S., & Kim, N. Y. (2021). Effects of AI Chatbots on EFL Students' Communication Skills*. *Korean Journal of English Language and Linguistics*, *21*, 712–734.

Larson, R. K. (2014). Critical pedagigy (ies) for ELT in Indonesia. *TEFLIN Journal*, *25*(1).

Lincoln, Y. S., Guba, E. G., & Pilotta, J. J. (1985). Naturalistic inquiry. *International Journal of Intercultural Relations*, *9*(4), 438–439. https://doi.org/10.1016/0147-1767(85)90062-8

Lukman. (2024). Development of Contextual Learning Models to Improve Student's Speaking Skill. *JOLLT Journal of Languages and Language Teaching*, *12*(1), 198–211.

Mary, C. C., & Paul, O. O. (2024). Challenges Faced by Students in Developing English Language Speaking Skills in Public Secondary Schools in Homa Bay County, Kenya. *Journal of Frontiers of Humanities and Social Sciences*, *2*(3), 80–98.

Mingyan, M., Noordin, N., & Razali, A. B. (2025). Improving EFL speaking performance among undergraduate students with an AI-powered mobile app in after-class assignments: an empirical investigation. *Humanities and Social Sciences Communications*, *12*(1), 370. https://doi.org/10.1057/s41599-025-04688-0

Phanwiriyarat, K., Anggoro, K. J., & Chaowanakritsanakul, T. (2025). Exploring AI-powered gamified flipped classroom in an English-speaking course: a case of Duolingo. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2488545

Qassrawi, R. M., ElMashharawi, A., Itmeizeh, M., & Tamimi, H. M. M. (2024). AI-Powered Applications for Improving EFL Students' Speaking Proficiency in

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)

Higher Education. *Forum for Linguistic Studies*, *6*(5), 535–549. https://doi.org/10.30564/fls.v6i5.6966

Rao, P. S. (2019). THE IMPORTANCE OF SPEAKING SKILLS IN ENGLISH CLASSROOMS. *Alford Council of International English & Literature Journal*, *2*(2).

Saleh, Z. (2019). Artificial Intelligence Definition, Ethics and Standards. *Electronics and Communications: Law, Standards and Practice* .

Rad, S. H. (2024). Revolutionizing L2 speaking proficiency, willingness to communicate, and perceptions through artificial intelligence: a case of Speeko application. *Innovation in Language Learning and Teaching*, *18*(4), 364–379. https://doi.org/10.1080/17501229.2024.2309539

Thornbury, S. (2005). *How to Teach Speaking*. Pearson Education.

Tomlinson, B. (2011). *Materials Development in Language Teaching* (Second Edition). Cambridge University Press.

Winnie, W., Leong, H. J., Badiozaman, I. F., & Yap, A. (2023). Negotiating the challenges in speaking English for Indonesian undergraduate students in an ESL university. *Studies in English Language and Education*, *10*(2), 822–840. https://doi.org/10.24815/siele.v10i2.26563

Wu, T.-T., Hapsari, I. P., & Huang, Y.-M. (2025). Effects of incorporating AI chatbots into think–pair–share activities on EFL speaking anxiety, language enjoyment, and speaking performance. *Computer Assisted Language Learning*, 1–39. https://doi.org/10.1080/09588221.2025.2478271

Zaki, W. A., & Darmi, R. (2021). CEFR: Education towards 21st Century of Learning. Why Matters? *Journal of Social Science and Humanities*, *4*(2), 14–20. https://doi.org/10.26666/rmp.jssh.2021.2.3

Zhang, Y., Li, B., & Tan, Y. (2021). Making AI available for everyone at anywhere: A Survey about Edge Intelligence. *Journal of Physics: Conference Series*, *1757*(1), 012076. https://doi.org/10.1088/1742-6596/1757/1/012076

Corresponding author: Luh Indrayani (luh.indrayani@undiksha.ac.id)