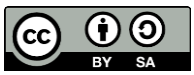
IDEAS

Journal of Language Teaching and Learning, Linguistics and Literature

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Volume 13, Number 1, June 2025
pp.1073 - 1088

Copyright © 2025 The Author
IDEAS is licensed under CC-BY-SA 4.0 License



Issued by English study program of IAIN Palopo

Leveraging ELSA Speak for Enhanced Speaking Proficiency: AI-Education (AIEd) Tool at Private University

Retno Rahayuningsih[1], Rudi Hartono[2], Sri Wahyuni[3]
[1]Sastra Inggris, Universitas Bina Sarana Informatika
[2,3]Pendidikan Bahasa Inggris, Univeritas Negeri Semarang
E-mail: retno.rrg@bsi.ac.id

Received: 2025-05-09 Accepted: 2024-03-09
DOI: 10.24256/ideas. v13i1.6573

Abstract

In response to the growing demand for effective English-speaking instruction, especially in EFL contexts, this study addresses the limited integration of AI tools that target all key components of speaking proficiency. This research investigates the impact of ELSA Speak, an AI-powered language learning application, on students' speaking proficiency. Utilizing a quasi-experimental mixed-method design with convenience sampling at Universitas Bina Sarana Informatika. Two available classes were assigned as the experimental (n=35) and control (n=36) groups. Oral pre-and post-tests were conducted, and the results were analyzed using t-tests in SPSS version 25. Results showed that the experimental group's scores increased from 61.43 to 75.03, surpassing gains in the control group from 58.78 to 65.81, with $p < 0.05$. Thematic analysis of student interviews supported these findings, highlighting enhanced pronunciation, grammar awareness, vocabulary acquisition, fluency, and content organization. The findings indicate that ELSA Speak significantly enhanced students' speaking proficiency in grammar, vocabulary, pronunciation, fluency, and content. The study demonstrates that integrating AI-based tools like ELSA Speak enhances higher speaking proficiency and offers a practical, scalable solution for improving language instruction aligned with Quality Education.

Keywords: *ELSA Speak, Speaking Proficiency, Technology, AI-Education (AIEd)*

Introduction

Many EFL students struggle with speaking proficiency despite years of formal English education. A recent study by (Nation ISP, 2022; Ngoc et al., 2023) highlights that pronunciation inaccuracy, lack of fluency, limited vocabulary, and weak idea organization remain the most common challenges for Southeast Asian learners. These persistent gaps are particularly evident in Indonesian university settings, where large class sizes and limited access to individualized instruction hinder oral language development (Hismanoglu & Hismanoglu, 2010). The result is a generation of learners who lack the confidence and competence to communicate effectively in English, especially in real-world contexts.

In response to these issues, educators and researchers have increasingly explored how technology can leverage software applications to create interactive and realistic lessons (Ghory & Ghafory, 2021; Pikhart, 2020), particularly Artificial Intelligence in Education (AIEd), which can transform speaking instruction. AIEd integrates AI innovations with learning theory to support personalized, data-driven, and adaptive instruction(Churi et al., 2022; Crompton & Burke, 2023; Doroudi, 2023; Ou et al., 2024; Zawacki-Richter et al., 2019). Applications range from intelligent tutors and learning partners to real-time assessment systems (Hwang et al., 2020; Roshanaei et al., 2023; Valverde et al., 2019). These tools aim to increase efficiency, engagement, and learner autonomy while responding to individual learning needs (de Souza Zanirato Maia et al., 2023; Kamalov et al., 2023). However, ethical, pedagogical, and implementation challenges remain.

One prominent AIEd tool is ELSA Speak, an AI-powered mobile application that offers structured pronunciation practice and real-time corrective feedback through advanced speech recognition (Anggraini, 2022; Becker & Edalatishams, 2019; Kholis, 2021a). ELSA's pedagogical design is based on phonetic and fluency development, enabling learners to record their speech, compare it to native models, and receive individualized scoring. The app promotes learner autonomy through gamified features and personalized learning paths (Saragih, N., Silalahi, R., & Pardede, 2014; Silaen & Rangkuti2, n.d.), making it increasingly relevant for both self-directed and classroom-based language learning.

Previous research has explored the impact of ELSA Speak on specific aspects of speaking skills. For instance, (Masekan et al., 2024; Rismawati et al., 2021) reported improved pronunciation and fluency among secondary learners. (Kholis, 2021b) found gains in pronunciation accuracy among university students, while (Ridhon & Daulay, 2023) highlighted its potential for learner motivation and articulation through gamification. Despite these positive outcomes, existing studies tend to isolate one or two speaking components, primarily pronunciation, leaving a gap in understanding ELSA Speak's comprehensive effects.

This study addresses that gap by examining the impact of ELSA Speak on five key components of speaking proficiency: grammar, vocabulary, pronunciation, fluency, and content. It employs a quasi-experimental mixed-method design to

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

evaluate whether the app can significantly enhance overall speaking performance among EFL university students. Additionally, this unique study explores students' perceptions and learning experiences through thematic analysis, offering more profound insights into how AIEd tools influence engagement and skill development.

By situating the research in a private university context, this study expands the scope of prior investigations beyond secondary education and highlights the growing relevance of AIEd for higher education. Therefore, the objectives of this study are:

1. To examine the effect of ELSA, speak on students' speaking proficiency across five components.
2. To explore student perceptions and experiences using ELSA Speak as an AI-assisted learning tool.

Method

Research Design

This study employed a quasi-experimental design with a mixed-methods approach, integrating quantitative and qualitative data to investigate the effect of ELSA Speak on students' speaking proficiency. Specifically, a nonequivalent control group design was used, suitable for real-world classroom settings where random assignment is often impractical (Cohen, 2018; Maciejewski, 2020). Pre-test and post-test assessments were administered to measure changes in student speaking performance. The research design is represented as follows:

O_1	X	O_2
O_3		O_4

Notes:

O_1 = represents the Pre-test of the Experimental group
O_2 = represents the Post-test of the Experimental group
X = Intervention using ELSA Speak
O_3 = represents the Pre-test of the control group
O_4 = represents the Post-test of the control group

This design enabled comparisons between the experimental and control groups to evaluate the impact of the intervention

Participant and Sampling

Participants were first-year English major students at Universitas Bina Sarana Informatika, comprising two existing classes with 71 students (35 in the experimental group and 36 in the control group). Due to institutional constraints and class structure, random assignment was not feasible. Therefore, convenience sampling was applied, allowing the researcher to work with intact classes that were readily accessible and willing to participate (Etikan, 2016). While this method introduces a risk of selection bias, baseline comparability was established via pre-

test equivalence and homogeneity tests.

Intervention Procedure

The experimental group participated in a structured six-week intervention program using the ELSA Speak mobile application, while the control group received traditional instruction without any AI integration. The intervention was designed to develop five key aspects of speaking proficiency: content, fluency, grammar, vocabulary, and pronunciation. These components are embedded in ELSA Speak's evaluation system and were adopted as the assessment criteria for the pre-test and post-test in this study.

During the first session (Monday), the teacher introduced the ELSA Speak application to students and explained how it enhances the five targeted speaking skills. The "New Kinds of Food" theme was selected for the pre-test task to align with the class syllabus. Each student was asked to perform a speaking activity using the corresponding ELSA module, which included an animated video and real-time scoring based on the five components. These automated scores were converted using a standardized analytic rubric to determine the students' pre-test performance.

This session functioned as both a diagnostic assessment and a teacher-guided orientation on how to use the AI-based tool effectively. After the pre-test, students were guided through ELSA Speak's initial placement test, which generated personalized learning paths tailored to each student's specific pronunciation profile. They were instructed to follow these custom lesson plans consistently and to engage in at least 30 minutes of speaking practice per day, five days per week, both at home and during class sessions.

The intervention emphasized active engagement with ELSA's feedback features, including, listening to model audio carefully, repeating speaking exercises until a high score was achieved, Reviewing targeted corrections for mispronounced words. In addition, features such as "Daily Coach" and topic-based speaking exercises were integrated to support vocabulary development and fluency in practical, real-world contexts. Speaking tasks were embedded within daily lessons, encouraging students to practice in both classroom and home environments. To support accurate speech recognition, students were advised to use headphones and practice in quiet environments. The integration of ELSA feedback into classroom activities allowed students to reinforce learning outcomes collaboratively while remaining focused on individualized improvement.

Student progress was monitored weekly through the ELSA dashboard, which allowed the teacher to track scoring trends, identify problem areas, and provide supplementary guidance. The researcher also offered initial orientation, ongoing support, and regular check-ins throughout the six-week period to ensure consistent implementation. To sustain engagement, motivational tools such as leader boards, streak tracking, and performance milestones were introduced.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

These gamified elements were instrumental in fostering consistent app usage and maintaining learner motivation.

By the following sessions, the teacher observed increased enthusiasm and engagement, as demonstrated by students' frequent questions, collaborative behaviours, and proactive use of the application.

The post-test was conducted during the sixth week using the same procedure and assessment criteria as the pre-test. Students completed the "New Kinds of Food" task again via ELSA Speak, and the AI-generated scores were converted into rubric-based scores for performance analysis. This ensured consistency in measurement and allowed for an objective comparison of students' progress before and after the intervention.

Data Collection Methods

a. Quantitative Component

Students' speaking proficiency was evaluated using oral proficiency tests administered before and after the intervention. The test assessed five key components: grammar, vocabulary, pronunciation, fluency, and content, using a 5-point analytic rubric adapted from (Ahmadi et al., 2020). An example item included prompts such as: "*Describe your favorite local food and how it is prepared."* Each component was rated from 1 (very limited) to 5 (excellent) with clearly defined descriptors to enhance scoring objectivity. Inter-rater reliability was maintained by a trained rater who evaluated pre-and post-tests for consistency.

b. Qualitative Component

To complement quantitative findings, semi-structured interviews were conducted with five purposefully selected participants from the experimental group. Selection was based on variation in pre-test and post-test performance (high, moderate, low gains) to ensure diversity of experience. The interviews explored: Students' perceptions of ELSA Speak features, Daily usage routines, Perceived benefits and challenges, Confidence in speaking after using the app. A standardized protocol was used to guide interviews and reduce bias. All interviews were audio-recorded, transcribed verbatim, and analyzed using thematic analysis to identify patterns in learner responses.

Data Analysis

Quantitative data were analyzed using paired-sample t-tests and independent-sample t-tests via SPSS Version 25 to examine intra-group improvements and inter-group differences. Normality and homogeneity tests were also conducted to meet statistical assumptions.

Qualitative data were manually coded using Microsoft Excel to identify recurring themes such as accuracy in speaking, phonological awareness, fluency, vocabulary retention, and coherent content delivery.

Validity of the study

Validity refers to how accurately a test or its parts measure what it is intended to assess (Li, 2011). The process of validation involves collecting evidence to support the accuracy and relevance of the conclusions drawn from students' responses to specific assessments (Moskal & Leydens, 2000). The validity of the oral test used in this study is supported through careful alignment with established speaking assessment frameworks, ensuring it accurately measures students' speaking proficiency. Using a structured analytic rubric based on recent guidelines, the test assessed five key components (grammar, vocabulary, pronunciation, fluency, and content).

This ensures strong content validity, as the test tasks reflect essential aspects of spoken communication. To support construct validity, each criterion was clearly defined with performance descriptors on a 5-point scale, minimizing subjectivity in scoring and focusing on the intended construct of speaking ability (Ahmadi et al., 2020). Furthermore, by administering both pre-tests and post-tests under controlled conditions and ensuring similar difficulty levels across test versions (Pranata et al., 2024), the study enhances its internal validity, allowing observed improvements to be reliably attributed to the ELSA Speak intervention. These measures contribute to the overall validity of the instrument in evaluating the impact of AI-assisted tools on English-speaking proficiency in EFL learners.

Reliability of the study

Reliability refers to the ability of a test to consistently produce the same results when administered under similar conditions at different times (Sürücü & Maslakci, 2020); a reliable test yields stable and dependable outcomes, which is essential for any effective assessment tool. The reliability of the oral speaking test in this study was ensured through several key strategies designed to promote consistent and accurate measurement. First, a structured analytic rubric evaluated five core speaking components (grammar, vocabulary, pronunciation, fluency, and content), rated on a clearly defined 5-point scale.

This rubric helped standardize the scoring process and reduce subjectivity, essential for inter-rater reliability (Ahmadi et al., 2020). To further enhance reliability, the same ratter assessed all student performances, and scoring procedures were applied uniformly across both the pre-test and post-test sessions. Consistency in test administration, such as timing, instructions, and testing conditions, also contributed to test-retest reliability, ensuring that scores were due to changes in student ability rather than inconsistencies in testing procedures. Well-designed rubrics and consistent administration protocols are vital for obtaining reliable results in performance-based language assessments. These practices ensured that the oral test provided dependable results, allowing the

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

researcher to draw valid conclusions about the impact of the ELSA Speak application on students' speaking proficiency.

Ethical Considerations

Before data collection, ethical clearance was obtained from the university's institutional review board. Participants were informed about the research purpose, assured confidentiality, and signed informed consent forms. Participation was voluntary, and students could withdraw without academic consequences.

Pilot Testing and Timeline

A pilot test was conducted with five non-participant students to ensure clarity of the speaking prompts and functionality of the ELSA app in the school's digital infrastructure. Minor adjustments were made based on feedback. The timeline of the study was as follows: week 1: Pre-test and orientation, weeks 2–7: Intervention (6 weeks), and week 8: Post-test and interviews.

Results

Pre-Test Results: Baseline Proficiency

Descriptive analysis was conducted to assess students' speaking proficiency before the intervention. As shown in Table 1, most students in the experimental and control groups performed fairly to reasonably, with only a few achieving excellent ratings.

Table 1. Speaking Proficiency Pre-Test Scores Distribution

Grade	Score	Experimental	Control
"Excellent"	76-100	6	6
"Good"	60-75	10	11
"Fair"	50-59	11	13
"Poor"	30-49	8	5
"Very Poor"	10-29	-	1
Total		35	36

Assumption Testing

Normality was assessed using both Kolmogorov-Smirnov and Shapiro-Wilk tests. As shown in Table 2, all values were above 0.05 in the Shapiro-Wilk test, supporting the normality assumption.

Table 2. Normality Test for Pre-Test Scores

Group	Kolmogorov-Smirnov Sig.	Shapiro-Wilk Sig.
Experimental	0.041	0.168
Control	0.200*	0.372

Homogeneity of variance was confirmed via Levene's test (p = 0.905), allowing for parametric testing. The two groups' variances are homogeneous because all importance values exceed 0.05. Given that the assumption of equal variance is satisfied, other parametric statistical tests, such as the t-test, can be performed confidently. Since H_0 was accepted, it can be concluded that the sample data met the homogeneity requirement.

Independent Samples T-Test (Pre-Test)

Following confirmation of the homogeneity and normality, the researcher examined the t-test of the pre-test data in more detail. A t-test confirmed no statistically significant difference in pre-test scores between the two groups (t (69) = -0.887, p = 0.378), as presented in Table 3.

Table 3. Independent Samples T-Test – Pre-Test

	Group	N	Mean	Std. Deviation	t	df	p-value	Cohen's d
Result	Experiment	35	61.43	12.45	-0.89	69	0.378	0.21
	Control	36	58.78	12.72				

Interpretation: The effect size (Cohen's d = 0.21) indicates a small and non-significant difference, suggesting both groups had comparable proficiency at the outset.

Post-Test Results: Impact of ELSA Speak

After six weeks of intervention, significant improvement was observed in the experimental group's speaking proficiency, as shown in Table 4.

Table 4. Speaking Proficiency Post-Test Scores Distribution

Grade	Score	Experimental	Control
"Excellent"	76-100	19	9
"Good"	60-75	16	12
"Fair"	50-59	-	12
"Poor"	30-49	-	3
"Very Poor"	10-29	-	-
Total		35	36

T-Test Results: Post-Test Comparison

As shown in Table 5, the post-test t-test revealed a statistically significant difference between groups (t (69) = -3.64, p = 0.001), with a Cohen's d effect size of 0.85, indicating an important practical impact.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Table 5. Independent Samples T-Test – Post-Test

	Group	N	Mean	SD	t	df	p-value	Cohen's d
Result	Experiment	35	75.03	10.28	-3.64	69	0.001	0.85
	Control	36	65.81	11.07				

Interpretation: The significant difference confirms the positive impact of the ELSA Speak intervention.

Qualitative Results: Thematic Insights from Student Feedback

To gain a deeper understanding of learners' experiences with the ELSA Speak application, the researcher conducted semi-structured interviews with five participants from the experimental group. These participants were purposefully selected to represent a range of proficiency levels (high, moderate, and low improvement), ensuring diverse perspectives.

The interview data were analyzed using thematic analysis, which revealed five major themes aligned with the five speaking components assessed quantitatively. Each theme was supported by specific excerpts from student responses, providing rich insight into how the app contributed to their speaking development.

Table 6. Thematic Analysis of Student Interview

Theme	Excerpts & Codes	Speaking Aspect
Accuracy in Speaking	"I understand sentence structure better"; "I can spot grammar mistakes when I talk" (S1)	Grammar
Lexical Development	"I learn new words every time I practice"; "Repetition in ELSA helps me remember vocabulary" (S2)	Vocabulary
Phonological Awareness	"ELSA corrects my pronunciation immediately"; "Now I pronounce words more clearly and correctly" (S3)	Pronunciation
Fluency and Expression	"Now I speak more fluently and confidently"; "My speech sounds more natural when I practice daily" (S4)	Fluency
Coherence in Speaking	"I can organize my ideas better when speaking"; "ELSA helps me express my ideas more clearly" (S5)	Content

Students reported increased grammatical awareness. S1 shared, "I understand sentence structure better," and "I can spot grammar mistakes when I talk now." This shows that ELSA's structured, grammar-integrated lessons helped learners monitor and self-correct their sentence construction.

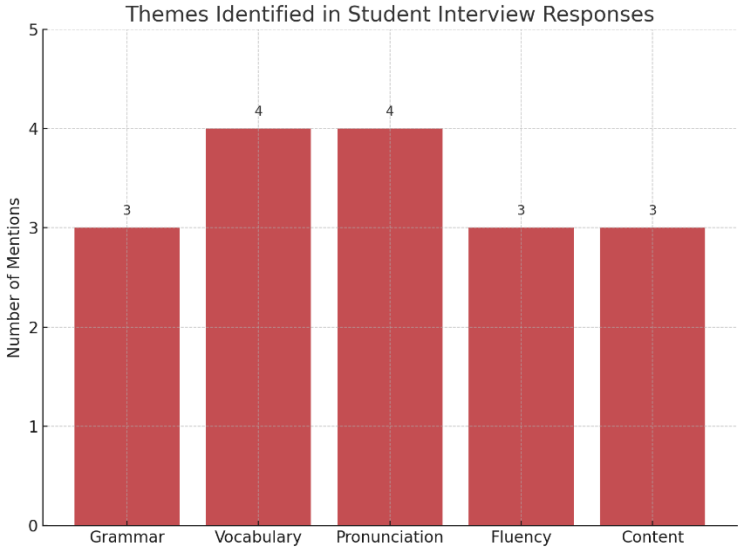
S2 expressed enthusiasm about vocabulary acquisition, stating, "I learn new words every time I practice," and highlighted the app's repetitive method: "Repetition in ELSA helps me remember vocabulary." These reflections support the idea that consistent exposure to contextual vocabulary builds retention and active usage.

Pronunciation improvement was one of the most cited benefits. S3 remarked, "ELSA corrects my pronunciation immediately," and added, "Now I pronounce words more clearly and correctly." The app's ASR (Automatic Speech Recognition) feature offered real-time feedback that students found valuable and corrective.

Participants like S4 noted smoother and more natural speech. "Now I speak more fluently and confidently," and "My speech sounds more natural when I practice daily." These insights demonstrate how consistent use of speaking prompts improves spontaneous verbal output.

Finally, improvements in idea structuring were mentioned by S5: "I can organize my ideas better when speaking," and "ELSA helps me express my ideas more clearly." This indicates the app's effectiveness in promoting clarity and logical flow in spoken discourse.

Here is a graph illustrating the main qualitative themes derived from interviews with five participants in the experimental group. The themes of vocabulary and pronunciation were mentioned most frequently, indicating the highest perceived impact of ELSA Speak according to the students.



Graph 1. Theme Identified in Student Interview Responses

Overall, the statistically significant improvement in post-test scores among the experimental group aligns with student feedback about increased confidence, fluency, and grammatical control. These parallel findings strengthen the validity of the conclusion that ELSA Speak effectively enhances EFL learners' speaking proficiency, both cognitively and affectively.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Discussion

The findings demonstrate that ELSA Speak, an AI-powered learning application, significantly improved students' speaking proficiency across five key components: grammar, vocabulary, pronunciation, fluency, and content. This is evidenced by the statistically significant gain in the experimental group's mean score ($M = 75.03$) compared to the control group ($M = 65.81$, $p < 0.01$), alongside a substantial increase in the number of students scoring in the "excellent" and "good" ranges.

These improvements align with prior studies on AI-mediated feedback in language learning. (Dai & and Wu, 2023) When combined with personalized feedback mechanisms, automatic speech recognition (ASR) tools can support pronunciation and fluency development. Similarly, (Sun, 2023) found that consistent use of ASR boosts speech automaticity and learner confidence. ELSA Speak's real-time correction and adaptive curriculum, tailored through placement testing, allowed learners to engage in repeated, self-paced practice, directly addressing individual weaknesses.

The qualitative themes reinforce these quantitative gains: students noted improved grammar accuracy, lexical retention, phonological awareness, fluency, and content organization. These reflections highlight how AI feedback provides error correction and supports deeper metacognitive engagement with language performance.

A notable contribution of this study lies in its emphasis on learner autonomy. ELSA Speak encourages students to control their learning through gamified features such as leader boards, streaks, and progress tracking. These elements enhance motivation and consistency—essential for developing oral skills (Ridhon & Daulay, 2023).

Theoretically, the study supports Second Language Acquisition (SLA) principles, particularly in fostering noticing, output production, and self-monitoring, all of which are central to communicative competence. It also extends current AIEd frameworks by demonstrating how mobile AI applications can function as supplemental tools and interactive learning environments that combine feedback, scaffolding, and learner control. This aligns with (Doroudi, 2023) assertion that AIEd applications should enhance both instructional quality and learner agency.

The results underscore the potential for pedagogical integration of AI tools in EFL settings. ELSA Speak can serve as a supplemental resource that provides individualized feedback, a task often impractical in large, resource-limited classrooms. Teachers can leverage the app to assign structured pronunciation tasks, monitor learner progress, and differentiate instruction based on learners' profiles.

For institutions, adopting AI-driven tools like ELSA Speak can enhance curriculum personalization, support remote learning strategies, and provide cost-effective solutions for improving speaking skills. Policymakers may promote digital

literacy initiatives and AI-based tool adoption in language programs to support SDG 4 (Quality Education) by increasing access to English instruction, equity, and quality.

While AI-driven platforms like Google's Read Along, Speechling, or Duolingo offer general language support, ELSA Speak stands out for its ASR precision, phonetic focus, and real-time pronunciation feedback. Its structure, rooted in fluency and accent training, directly targets the needs of EFL learners struggling with spoken intelligibility, especially in non-native academic environments.

Although the study demonstrated strong internal validity, as evidenced by normality, homogeneity tests, and pre-test equivalence between groups, several limitations should be noted. First, convenience sampling and intact classes limit the generalizability of the findings; future research should consider employing randomized or stratified sampling across multiple institutions. Second, the six-week intervention period may be too short to evaluate long-term learning outcomes, suggesting a need for longitudinal studies to assess the durability of speaking proficiency gains.

Third, as the study was conducted within a single institutional context, its external validity is limited; future research across diverse educational and cultural settings would provide a broader perspective. Lastly, although scoring was consistent, it relied on a single rater, which may reduce objectivity. Incorporating multiple raters and implementing blind scoring procedures in future studies would enhance the reliability of the assessment.

Integrating ELSA Speak as an AI-education (AIEd) tool demonstrates measurable benefits for speaking proficiency development in EFL learners. The study contributes to AIEd and SLA theory by validating that AI feedback, learner autonomy, and structured, personalized practice can lead to substantive linguistic gains. Practically, it offers a scalable model for technology integration in English instruction, particularly in under-resourced or large-class contexts. Future innovations in AI language tools should continue to prioritize learner agency, feedback precision, and curriculum alignment to optimize impact.

Conclusion

This study proves that ELSA Speak is an effective tool for improving speaking proficiency among English students at Universitas Bina Sarana Informatika, as both quantitative and qualitative findings demonstrate. The quantitative results, derived from an independent samples t-test conducted using SPSS version 25, revealed a statistically significant difference between the control and experimental groups. The experimental group, which utilized ELSA Speak, achieved a higher average post-test score ($M = 75.03$) compared to the control group ($M = 65.81$), with a p-value of 0.00 (< 0.05). These results support rejecting the null hypothesis and confirm the app's positive impact on speaking performance.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

However, it is essential to acknowledge that the observed improvement may not be solely attributable to ELSA Speak. Additional factors, such as the teacher's active involvement and the novelty of AI integration, likely contributed to the learning outcomes. The teacher played a crucial role by guiding students in using the app effectively, monitoring their progress, and encouraging consistent practice, which may have independently influenced student achievement.

Qualitative data further supported these findings. Thematic analysis revealed five key themes that align with the study's evaluation components: speaking accuracy, vocabulary development, phonological awareness, oral expression, and coherence. Students reported that the app enhanced their understanding of grammar and sentence structure (accuracy), supported vocabulary retention through repetition (lexical development), and raised awareness of pronunciation errors through instant corrective feedback (phonological awareness). They also experienced improved fluency, more natural speech (oral expression), and better organization of ideas (coherence).

These findings suggest that ELSA Speak can be a valuable supplemental resource, especially in educational environments with limited time and resources. Its flexible, gamified format promotes student engagement and supports self-directed learning. While this study highlights the app's potential to enhance speaking proficiency and scale language instruction, it also acknowledges limitations, including the small qualitative sample size and the specific institutional context.

Further research is needed to generalize these findings. Comparative studies involving other AI-powered educational tools, such as Google's Read Along, Speechling, or Duolingo, could provide more comprehensive insights into the effectiveness of personalized learning and adaptive feedback mechanisms. As digital education grows, leveraging AI-driven platforms like ELSA Speak will be crucial in developing more accessible, personalized, and engaging language learning experiences across diverse educational settings.

References

Ahmadi, B., Namaziandost, E., & Etemadfar, P. (2020). A Comparative Study of Using Analytic and Holistic Scoring Methods in Measuring Speaking Skill Among Iranian Second-Year University EFL Students. In *Journal of English Language Studies* (Vol. 5). http://jurnal.untirta.ac.id/index.php/JELS

Anggraini, A. (2022). Improving Students' Pronunciation Skill Using Elsa Speak Application. *Journey*, *5*(1), 135–141. https://doi.org/10.33503/journey.v5i1.1840

Becker, K., & Edalatishams, I. (2019). ELSA Speak-Accent Reduction [Review. In *Iowa State University. Pronunciation in Second Language Learning and Teaching* (Vol. 10).

Churi, P. P., Joshi, S., Elhoseny, M., & Omrane, A. (2022). *Artificial intelligence in higher education: A practical approach*. CRC Press.

Cohen, L., M. L., & M. K. (2018). *Research Methods in Education (8th ed.)*. Routledge.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00392-8

Dai, Y., & and Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: a mixed-methods study. *Computer Assisted Language Learning*, *36*(5–6), 861–884. https://doi.org/10.1080/09588221.2021.1952272

de Souza Zanirato Maia, J., Bueno, A. P. A., & Sato, J. R. (2023). Applications of Artificial Intelligence Models in Educational Analytics and Decision Making: A Systematic Review. *World*, *4*(2), 288–313. https://doi.org/10.3390/world4020019

Doroudi, S. (2023). The Intertwined Histories of Artificial Intelligence and Education. *International Journal of Artificial Intelligence in Education*, *33*(4), 885–928. https://doi.org/10.1007/s40593-022-00313-2

Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1. https://doi.org/10.11648/j.ajtas.20160501.11

Ghory, S., & Ghafory, H. (2021). The impact of modern technology in the teaching and learning process. *International Journal of Innovative Research and Scientific Studies*, *4*(3), 168–173. https://doi.org/10.53894/ijirss.v4i3.73

Hismanoglu, M., & Hismanoglu, S. (2010). Language teachers' preferences of pronunciation teaching techniques: Traditional or modern? *Procedia - Social and Behavioral Sciences*, *2*(2), 983–989. https://doi.org/10.1016/j.sbspro.2010.03.138

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. In *Computers and Education: Artificial Intelligence* (Vol. 1). Elsevier B.V. https://doi.org/10.1016/j.caeai.2020.100001

Husnaini, H., Yahya, A., & Putri, N. I. W. (2023). The Efficacy of the Presentation, Practice, and Production (PPP) Method on the Speaking Skill of the English Learners Community (ELC) Students. EDULANGUE, 6(1), 45-61.

Husnaini, H., Syam, A. T., Niken, N., & Soares, A. (2024). Teachers' Strategies in Teaching Speaking at High Schools: Obstacles and Challenges. Linguistics Initiative, 4(2), 243-262.

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution. *Sustainability (Switzerland)*, *15*(16). https://doi.org/10.3390/su151612451

Kholis, A. (2021a). Elsa Speak App: Automatic Speech Recognition (ASR) for

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Supplementing English Pronunciation Skills. *Pedagogy: Journal of English Language Teaching*, *9*(1), 01. https://doi.org/10.32332/joelt.v9i1.2723

Kholis, A. (2021b). Elsa Speak App: Automatic Speech Recognition (ASR) for Supplementing English Pronunciation Skills. *Pedagogy: Journal of English Language Teaching*, *9*(1), 01. https://doi.org/10.32332/joelt.v9i1.2723

Li, W. (2011). Validity Considerations in Designing an Oral Test. *Journal of Language Teaching and Research*, *2*(1). https://doi.org/10.4304/jltr.2.1.267-269

Maciejewski, M. L. (2020). Quasi-experimental design. *Biostatistics & Epidemiology*, *4*(1), 38–47. https://doi.org/10.1080/24709360.2018.1477468

Masekan, S., Liando, N. V. F., & Rorintulus, O. A. (2024). THE USE OF "ELSA" APPLICATION TO IMPROVE STUDENTS SPEAKING SKILLS AT EIGTH GRADE STUDENTS AT SMP N 8 SATAP TONDANO. *JoTELL Journal of Teaching English*, *3*(1), 32–46.

Moskal, B. M., & Leydens, J. A. (2000). *Scoring Rubric Development: Validity and Reliability. - Practical Assessment, Research & Evaluation* (Vol. 7, Issue 10).

Masruddin, Hartina, S., Arifin, M. A., & Langaji, A. (2024). Flipped learning: facilitating student engagement through repeated instruction and direct feedback. Cogent Education, 11(1), 2412500.

Nation ISP. (2022). Learning Vocabulary in Another Language. In I. S. P. Nation (Ed.), *Learning Vocabulary in Another Language* (3rd ed., pp. i–i). Cambridge University Press. https://www.cambridge.org/core/product/C21184E780F20F5342C2E425BE5B166B

Nasriandi, N., & Masruddin, M. (2021). The Use of British Parliamentary Debate Style in Teaching Speaking Skill. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 9(1).

Ngoc, N. K., Thanh, N. T. M., & Thanh, N. T. M. (2023). TERTIARY STUDENTS' PERCEPTIONS ON ELSA SPEAK APPLICATION FOR PRONUNCIATION LEARNING. *European Journal of Applied Linguistics Studies*, *7*(1). https://doi.org/10.46827/ejals.v7i1.483

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

Pikhart, M. (2020). Intelligent information processing for language education: The use of artificial intelligence in language learning apps. *Procedia Computer Science*, *176*, 1412–1419. https://doi.org/10.1016/j.procs.2020.09.151

Pranata, S. P., Mahkota, U., & Unggul, T. (2024). *From Clicks to Trust: Relationships on TikTok Insights into Customer*. *3*(11), 4043–4058.

Ridhon, M., & Daulay, S. H. (2023). ENHANCING STUDENTS' ORAL ARTICULATION

IN ENGLISH BY USING GAMIFICATION: STUDENTS' PERCEPTION. *Ethical Lingua: Journal of Language Teaching and Literature*, *10*(2). https://doi.org/10.30605/25409190.556

Rismawati, dina, Suryana, yayan, & Agustiana, V. (2021). THE EFFECTIVENESS OF ELSA SPEAKING APPLICATION IN IMPROVING ENGLISH PRONUNCIATION. *The Proceedings of English Language Teaching, Literature, and Translation (ELTLT)*, *10*, 177–184. https://proceeding.unnes.ac.id/eltlt/article/view/1329

Roshanaei, M., Olivares, H., & Lopez, R. R. (2023). Harnessing AI to Foster Equity in Education: Opportunities, Challenges, and Emerging Strategies. *Journal of Intelligent Learning Systems and Applications*, *15*(04), 123–143. https://doi.org/10.4236/jilsa.2023.154009

Saragih, N., Silalahi, R., & Pardede, H. (2014). The Effectiveness of Using Recount Text to Improve Writing Skill For Grade III Students of Kalam Kudus Elementary School 2 Pematangsiantar. *IOSR Journal Of Humanities And Social Science*, 56–64.

Silaen, T. N., & Rangkuti[2], R. (n.d.). *Elsa Speak App Usage in Blended Learning During The Covid-19 Pandemic: Students' Perspectives*.

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1210187

Sürücü, L., & Maslakci, A. (2020). Validity and Reliability in Quantitative Research. *Business & Management Studies: An International Journal*, *8*(3), 2694–2726. https://doi.org/10.15295/bmij.v8i3.1540

Valverde, P., Rivas, A., Subosa, M., & Pedró, F. (2019). *Artificial intelligence in education: challenges and opportunities for sustainable development*.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? In *International Journal of Educational Technology in Higher Education* (Vol. 16, Issue 1). Springer Netherlands. https://doi.org/10.1186/s41239-019-0171-0