

# The Effectiveness of the English Language Speech Assistant Speak Application in Improving Students' Pronunciation

 <https://doi.org/10.31004/jele.v11i2.2073>

\* Aurelia Nova Ramadyanti, Esti Kurniasih<sup>ab</sup>

<sup>12</sup>English Education, State University of Surabaya, Surabaya, Indonesia

<sup>†</sup>Corresponding Author: [aurelia.22024@mhs.unesa.ac.id](mailto:aurelia.22024@mhs.unesa.ac.id)

## ABSTRACT

Pronunciation remains one of the most challenging skills for EFL students in Indonesia, particularly at the elementary school level where teachers often lack English expertise and class time is limited. This study examined the effectiveness of the ELSA Speak application as an AI-powered pronunciation learning tool in addressing these challenges. Using a true experimental design with pre-test and post-test measures, the study was conducted at UPT SD Negeri 79 Gresik involving nine sixth-grade students randomly assigned to experimental (n=5) and control (n=4) groups. The experimental group received three treatment sessions using ELSA Speak, while the control group followed conventional instruction. Results revealed a substantial and statistically significant improvement in pronunciation scores in the experimental group compared to the control group ( $p=0.016$ ), with the experimental group demonstrating a dramatic gain from pre- to post-test ( $p<0.001$ ). These findings indicate that ELSA Speak's features – particularly its accurate pronunciation modeling and real-time automated feedback – effectively compensate for the constraints of conventional EFL instruction in elementary settings. This study contributes novel evidence for the application of mobile-assisted language learning at the elementary school level in Indonesia, a context previously underexplored in the literature. Practically, integrating ELSA Speak into classroom instruction is recommended as a feasible and impactful strategy to enhance pronunciation learning, provided that device access and internet connectivity are adequately supported.

**Keywords:** EFL Pronunciation, ELSA Speak, Experimental Design, Mobile Learning, True Experiment

### Article History:

Received 21<sup>st</sup> January 2026

Accepted 03<sup>rd</sup> March 2026

Published 04<sup>th</sup> March 2026



## INTRODUCTION

English has become the dominant global language in various fields such as international communication, trade, diplomacy, and education. Mastery of English, particularly speaking skills, is increasingly important in the era of the industrial revolution 4.0, where national and multinational companies require English language skills for job opportunities. In Indonesia, English is taught as a foreign language (EFL) with an emphasis on the oral aspect to support students' overall competency.

Pronunciation is one of the essential aspects of speaking, along with fluency, vocabulary richness, grammatical correctness, speech logic, word stress, facial expressions, intonation, and eye contact (Róji, 2024). Without proper pronunciation, listeners have difficulty understanding spoken words, potentially leading to misinterpretation of meaning. This makes pronunciation a foundational skill that must be developed from the earliest stages of English language learning.

Pronunciation requires constant practice and error correction along with useful feedback until students achieve satisfactory improvement and can use the target language appropriately (Karafil & Uysal İlbay, 2024; Ngoc & Dung, 2020). However, EFL environments like those in Indonesia are less supportive of pronunciation practice because the use of the mother tongue dominates daily, so students rarely practice repetitively (Si, 2019). This challenge is

particularly acute at the elementary school level, where students are at a critical period of language acquisition – a window that research has shown to be especially influential in shaping long-term linguistic competence (Hartshorne et al., 2018; Xu, 2023).

Teacher feedback alone is insufficient given the large number of students per class compared to the limited time allocated for English lessons. This leads to insufficient time for pronunciation practice and inaccurate pronunciation demonstrations by non-expert teachers. In elementary schools, this challenge is compounded by teachers often having a background in basic education rather than English, necessitating the use of additional, state-of-the-art technology for pronunciation assessment and feedback.

Artificial Intelligence (AI), specifically automatic speech recognition (ASR), offers a potential solution through applications such as ELSA Speak that provide real-time feedback. A growing body of research has demonstrated the effectiveness of ELSA Speak in improving pronunciation among EFL learners, particularly at the tertiary level (Al-Shallakh, 2024; Fauziah et al., 2024). These studies consistently highlight the application's ability to deliver instant, automated feedback that supplements traditional teacher instruction. However, two critical gaps remain in the existing literature. First, the majority of studies on ELSA Speak have been conducted with university or secondary school students, leaving its effectiveness at the elementary school level largely unexplored – despite this stage being recognized as the golden age of language learning (Hartshorne et al., 2018; Xu, 2023). Second, most prior studies in the Indonesian EFL context have relied on pre-experimental designs, such as a single-group pre-test post-test design, which offer limited control over extraneous variables and weaken the internal validity of their findings (Cohen et al., 2018). A true experimental design with random assignment to experimental and control groups has yet to be widely applied in this area of research within Indonesia.

This study therefore aims to address these gaps by testing the effectiveness of ELSA Speak in improving the pronunciation of elementary school EFL students using a true experimental design. By targeting the previously underexplored elementary school context and employing a more rigorous research design, this study contributes both theoretical and practical insights to the field of mobile-assisted language learning in Indonesia.

## METHOD

This study used a quantitative approach with a true experimental design to test the effect of the independent variable, namely the use of the ELSA Speak application, on the dependent variable, namely the pronunciation scores of elementary school EFL students. The true experimental design was chosen because it involves the random assignment of participants to experimental and control groups, where both groups receive a pre-test and post-test while only the experimental group receives the intervention treatment (Ary et al., 2010; Creswell, 2014). This approach ensures tighter control of extraneous variables than quasi-experimental designs or single-group pre-test post-test designs, thereby increasing the internal validity of the research results (Emzir, 2021; Sugiyono, 2019).

The research instrument consisted of a pre-test and post-test in the form of 30 word and phrase pronunciation tests that measured baseline abilities and post-intervention changes, with a Cronbach's Alpha reliability of 0.612 indicating good consistency and validity confirmed by English Language Education expert. Data analysis techniques included the Shapiro-Wilk normality test, Levene's homogeneity test, independent sample t-test to compare post-tests between groups, and paired sample t-test for intra-experimental group changes, processed using SPSS version 23 (Aulia et al., 2024). This analysis complies with parametric standards with a significance threshold of 0.05 to determine the effectiveness of the intervention (Lim, 2024; Sudaryono, 2022).

The study population was all 9 sixth-grade students of UPT SD Negeri 79 Gresik, with a total sample of 9 students randomly divided using odd-even numbers into experimental (5 students) and control (4 students) groups. It is acknowledged that this small sample size (n=9) constitutes a limitation of the study, as it restricts the generalizability of the findings to broader

populations. However, the use of a true experimental design remained feasible and methodologically justified for two reasons. First, because the entire population of sixth-grade students at the school comprised only 9 individuals, the sample represents a complete population study rather than a partial selection, which inherently minimizes sampling bias (Creswell, 2014). Second, the random assignment of participants to groups – a defining feature of true experimental design – was fully achievable at this scale, preserving the design's core requirement for controlling extraneous variables regardless of sample size (Ary et al., 2010). This small sample size is also acceptable in quantitative EFL research because it focuses on the normal distribution of data rather than absolute size, as supported by the Central Limit Theorem for small samples (Gunawan et al., 2021; Memon et al., 2020).

The research procedure began with instrument validation by experts, followed by a tryout on 10 non-sample students for reliability testing, and then a 70-minute pre-test on both groups. The experimental group received ELSA Speak treatment for three meetings on the topic of past verbs and adverbs according to the My Next Words Grade 6 textbook (Team, 2022), while the control group used conventional methods. A post-test was administered after the treatment to measure improvement, with pronunciation assessment adopted from Al-Shallakh (2024) on a score range of 1-100. All procedures systematically followed experimental research ethics to ensure replicability (Emzir, 2021; Sugiyono, 2019).

## FINDINGS AND DISCUSSION

To evaluate the data, the researchers evaluated students' test scores. These scores were subsequently inputted into SPSS IBM Statistics 23. The researcher utilized normality test and homogeneity within SPSS test to fulfill the assumption required to conduct t-tests in which the data should be normal and homogeneous. T-tests were utilized in SPSS as well to analyze differences in both pre-test and post-test assessments between the control and the experimental groups. The analytical findings are presented below.

First, the researcher checked whether the data was normal or not. Since the sample size of this research is less than 30, it is categorized as "small sample" (Wahyudi et al., 2024). There is a suggestion which states that a sample size for quantitative study should be more than 30. However, that is not always the case for all data distribution (Gunawan et al., 2021; Koh & Ahad, 2020). Moreover, Randahl (2024) suggested that 30 as the minimum threshold of sample size rule is overly conservative in some cases and insufficient in others. Therefore, it needs to be emphasized that there is no absolute rule for the minimum threshold in sample size, just a suggestion (Memon et al., 2020). To understand this suggestion, we must first understand why this suggestion was made in the first place. (Memon et al., 2020) and Krithikadatta (2014) stated that the logic behind the suggestion of 30 as the minimum threshold is based on the Central Limit Theorem (CLT) which states that the more the sample size increases, the more the variability of sampling distribution decreases, leading to an increased chance that the data might be normally distributed. Therefore, the purpose of the suggestion that the sample size in quantitative study should be at least 30 is to achieve greater opportunities for the data to be normally distributed. This means that the suggestion was not given because 30 is the absolute number required for quantitative study, but for the data to be normal.

The Shapiro-Wilk test, which is especially appropriate for small to intermediate sample sizes (less than 100), as is the case in this study, was employed by the researcher to assess for normality. For each data set, the Shapiro-wilk test yields a significance value (Sig or p-value).

Table 1. Normality Test Score

	Group	Kolmogorov-Smirnov*			Shapiro-Wilk*		
		Statistics	df	Sig.	Statistics	df	Sig.
Result	Pre-Test A (Experimental)	.186	5	.200*	.943	5	.687
	Post-Test A (Experimental)	.221	5	.200*	.953	5	.758
	Pre-Test B (Control)	.214	4		.963	4	.798
	Post-Test B (Control)	.277			.944	4	.679

Based on the results of the table above, the calculation using Shapiro-Wilk for the pre-test of the experimental group is 0.687, the post-test of the experimental group is 0.758, the pre-

test of the control group is 0.798, and the post-test of the control group is 0.679. All of these values are above the 0.05 threshold. This means that for each group and each test (both before and after the treatment), the data are normally distributed.

After confirming the normality, the researcher checked whether the data was homogeneous or not. Homogeneity means that the variance within each group being compared is similar. The researcher used Lavene's Test which provides a significance value (often referred to as "Sig" or p-value). If this value is greater than 0.05, it indicates that there is no significant difference in variance between groups, meaning the data are considered homogeneous.

Table 2. Homogeneity Test

		Lavene Statistics	df1	df2	Sig.
Result	Based on Mean	1,867	3	14	.181
	Based on Median	1,401	3	14	.284
	Based on Median and with Adjusted df	1,401	3	5,313	.340
	Based on Trimmed Mean	1,782	3	14	.197

The table provided shows several ways Lavene's test can be calculated (based on mean, median, trimmed mean, etc.), but the most common and straightforward is the one based on the mean. In this study, the significance value of Lavene's test based on the mean is 0.181. This value is higher than the 0.05 threshold, which indicates that the variances between the post-test scores of the control and experimental groups are indeed equal. Then the researcher compared the pre and post test scores in both groups to determine how much their language mastery had improved. This comparison was conducted to determine the distinction between those who were taught using the *ELSA Speak* application and those who did not. As with the initial assessment, the researcher employed an independent sample t-test to evaluate the disparity in abilities between students from different groups. Detailed results are presented in the accompanying table.

Table 3. Mean Score Pre-Test and Post-Test from Experimental Group and Control Group

Mean Score	Experimental Group		Control Group	
	Pre-Test	Post-Test	Pre-Test	Post-Test
	33.7356	94	31,169	53,292

The mean score from the post-test of the experimental group and the control group was then determined by the researcher using an independent t-test. There is a significant difference between the experimental and control groups if the mean Sig. score is less than 0.05. Additionally, the treatment is considered successful if the experimental group's mean score is higher than that of the control group. The independent sample t-test results are displayed in the table below.

Table 4. Independent Sample T-Test Mean Score

Groups		N	Mean	Standard Deviation	Std. Error Mean
Results	Post-Test Control	4	53.04200	12.968864	2.484432
	Post-Test Experimental	5	94.00000	4.873397	2.179449

Table 5. Independent Sample T-Test

		Lavene's Test for Equality of Variances		T-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Standard Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Results	Equal variances assumed	3,449	.106	-3,160	7	.016	-40.958000	3.959626	-71.602646	10.313354
	Equal variances not assumed			-2,796	36	.065	-40.958000	4.647483	-86.448428	428

According to the data presented in the table, there is a significant difference in results between the experimental group ( $M = 94$ ,  $SD = 4.873397$ ) and the control group ( $M = 53.04200$ ,  $SD = 12.968864$ ), with  $t(7) = 3.160$ ,  $p = 0.016$ .

Next, the researcher compared the pre-test and post-test results of the experimental group. The purpose of this section was to determine the group's improvement following the treatment.

Table 6. Paired Sample Test Mean Score

Groups		Mean	N	Standard Deviation	Std. Error Mean
Pair 1	Pre-Test Experimental Group	33.73560	5	13.572098	6.069627
	Post-Test Experimental Group	94,00000	5	4.873397	2.179449

Table 7. Paired Sample Test

		Paired Differences				t	d	Sig. (2-tailed)	
		Mean	Standard Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Pretest-Posttest	60.26440	10.698381	2.784461	73.54819	46.98060	12,59	.000	

Based on the tables provided above, a paired sample t-test was conducted to assess the impact of *ELSA Speak* application on students' performance in pronunciation tests. There was a remarkable improvement in scores from the pre-test ( $M = 33.73560$ ,  $SD = 13.572098$ ) to the post-test ( $M = 94$ ,  $SD = 4.873397$ ),  $t(4) = 12.596$ ,  $p = 0.000$ .

Based on the results of independent sample t-test, it shows that sig. (2-tailed) score is 0.016, which is lower than 0.05. This indicates that the data is correlated. The utilization of *ELSA Speak* application is effective for improving pronunciation compared with traditional methods; If students' pronunciation increases, it is in line with their achievement scores. In accordance with these findings, which show how the utilization of *ELSA Speak* application as a treatment to assist students in having an effect on their pronunciation compared with the traditional method, their learning results show an influence on the mastery of pronunciation. According to a study done by Ngoc & Thanh (2023), the results of the study indicate a clear preference among students for *ELSA Speak* application. The students agreed that *ELSA Speak* has many impressive features such as: 1) Content design which addresses many aspects of pronunciation; 2) Good pedagogical design with detailed instructions; 3) Amazing flexibility for learning; 4) Great multimedia design such as the installation, animation, images and sounds and; 5) Practical automatic speech recognition system.

The positive outcomes observed in this study corroborate existing literature on the effectiveness of *ELSA Speak* application in educational settings. The accurate pronunciation demonstration feature of *ELSA Speak* application helps students in imitating the sounds more accurately. This helps a lot especially for elementary school settings where there are no guarantees that the English teachers graduated from English major, thus having a chance to make pronunciation errors themselves during pronunciation demonstration (Tiwari, 2023). In this case, the English teacher of grade 6 at UPT SD Negeri 79 Gresik is majoring in elementary school education (PGSD), not from English major. *ELSA Speak* application can be a good role model for the students, especially in pronunciation practice. In this study, the students in the experimental group could listen and repeat the pronunciation demonstration limitless times. If they find the pronunciation demonstration too fast, they can press the 'snail' button to listen to the lowered-speed version to listen better.

Moreover, this study highlighted the role of instant and real-time feedback and error correction in pronunciation practice. Students in the experimental group were able to show

better improvement in pronouncing the same words and phrases. This can be achieved due to the increased efficiency in learning, where the students in the experimental group can practice their pronunciation more compared to the students in the control group during the same limited time. Within 15 minutes of pronunciation practice, the students in the experimental group can attempt pronouncing the words, then reattempt them unlimited times until they are satisfied with the results. After they pressed the microphone button and pronounced the word, the accuracy percentage popped out immediately. Their voices were also transcribed by the application and colored differently: green for correctly pronounced letters, and red for mispronounced letters. They can also reattempt practicing each word and keep on getting instant feedback and error correction without waiting at all.

The students in the control group, on the other hand, had to wait a long time for the teacher to switch from one student to the others. The teacher must first give pronunciation demonstration to the students, and if there are students who forgot how to pronounce the words or cannot hear the teacher, they will keep asking the teacher to repeat the demonstration on and on, which consumes a lot of times. The same goes with giving feedback and error correction on students' pronunciation. The teacher must listen attentively to each student one at a time which makes the other students wait, and the time might not be enough. Therefore, the existence of *ELSA Speak* application makes up for the lack of time allocation for pronunciation practice by providing a wholesome learning experience such as receiving pronunciation demonstration as well as real-time feedback and error correction limitlessly without depending and waiting for the teacher to come to them.

The findings of this research suggest that *ELSA Speak* application significantly enhances pronunciation among sixth grade EFL students. This conclusion aligns with previous studies, which consistently demonstrate that *ELSA Speak* application is an effective instructional medium for improving students' pronunciation. The accurate pronunciation demonstrations feature provided by *ELSA Speak* application acts as a role-model in pronunciation practice for the students. The real-time feedback and correction also saved a lot of time for students to learn more despite limited time without waiting for the teacher to come to them to give feedback and error correction on their pronunciation.

The effectiveness of *ELSA Speak* application is influenced by several factors, including facilitation of devices and internet connection and varying levels of students' motivation. Addressing these factors is essential for optimizing the impact of *ELSA Speak* application. Future research should explore these variables further to validate and refine the utilization of *ELSA Speak* application in diverse educational settings. Overall, the *ELSA Speak* application not only fosters a better pronunciation, but also enhances students' academic performance, making it valuable instructional media in contemporary education.

## CONCLUSION

This study concluded that the *ELSA Speak* application was proven effective in improving the pronunciation skills of sixth-grade EFL students. The experimental group demonstrated a substantial and consistent improvement following the intervention, markedly outperforming the control group in post-test results. These findings confirm that *ELSA Speak*'s accurate pronunciation demonstration and real-time automated feedback features effectively address the limitations of conventional instruction, particularly in elementary school EFL environments where English teachers may lack specialized linguistic training. This finding is in line with previous literature showing the effectiveness of *ELSA Speak* at higher levels (Al-Shallakh, 2024; Fauziah et al., 2024), while extending its evidence base to the previously underexplored elementary school context in Indonesia. However, this study has limitations in the form of a small sample size (n=9), which limits the generalizability of the results, and a short intervention duration (3 sessions), which did not measure long-term retention. For future research, it is recommended to use a longitudinal design with a larger sample, a mixed-method approach to explore student perceptions, and a comparison with similar applications such as BoldVoice or Chatterfox. Practically, elementary school teachers are advised to integrate *ELSA Speak* by ensuring stable device access and technical training, while school administrators can

provide Wi-Fi infrastructure and backup devices to optimize pronunciation learning in the digital era.

## REFERENCES

- Al-Shallakh, M. A. I. (2024). Embedding artificial intelligent applications in higher educational institutions to improve students' pronunciation performance. *Theory and Practice in Language Studies*, 14(6), 1897–1906. <https://doi.org/10.17507/tpls.1406.31>
- Ary, D., Jacobs, L. C., Sorensen, C., & Razavieh, A. (2010). Introduction to research in education. In *Wadsworth*.
- Aulia, U., Junaidi, H., Setiawan, D. I., Kurnia, P. S. D., Meidatuzzahra, D., Nirmala, I. R., Nadjib, H. R., Sari, M. P., Syilfi, R., Rakanita, Y., Astuti, N. H., & Larasati, R. D. (2024). Parametric statistics: Theory and action with SPSS applications. In *CV. Media Sains Indonesia*.
- Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education. In *Routledge*.
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). SAGE Publications, Inc.
- Emzir. (2021). Qualitative research methodology: Theory and application. In *Prenada Media Group*.
- Fauziah, N., Aororah, Z. Z., Ramadhany, R. A., & Hamid, S. M. (2024). The impact of The English Language Speech Assistant (ELSA) application on students' pronunciation skills. *INTERACTION: Journal of Language Education*, 11(2), 193–203. <https://doi.org/10.36232/interactionjournal.v11i2.26>
- Gunawan, J., Marzilli, C., & Aunguroch, Y. (2021). Establishing appropriate sample size for developing and validating a questionnaire in nursing research. *Belitung Nursing Journal*, 7(5), 356–365. <https://doi.org/10.33546/bnj.v7i5>
- Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263–277. <https://doi.org/10.1016/j.cognition.2018.04.007>
- Karafil, B., & Uysal İlbay, Ö. (2024). EFL instructors' satisfaction with teaching English online at tertiary level. *Participatory Educational Research*, 11(2), 76–96. <https://doi.org/10.17275/per.24.20.11.2>
- Koh, K. L., & Ahad, N. A. (2020). Normality for non-normal distributions. *Journal of Science and Mathematics Letters*, 8(2), 51–60. <https://doi.org/10.37134/jsml.vol8.2.7.2020>
- Lim, W. M. (2024). What is quantitative research? An overview and guidelines. *Australasian Marketing Journal*. <https://doi.org/10.1177/14413582241264622>
- Memon, M. A., Ting, H., Cheah, J.-H., Thurasamy, R., Chuah, F., & Cham, T. H. (2020). Sample size for survey research: Review and recommendations. *Journal of Applied Structural Equation Modeling*, 4(2), i–xx.
- Ngoc, D. T. B., & Dung, T. T. (2020). Key factors influencing learners' oral fluency in English speaking classes: A case at a public university in Viet Nam. *VNU Journal of Foreign Studies*, 36(6), 93–108. <https://doi.org/10.25073/2525-2445/vnufs.46.31>
- Rójs, A. (2024). Different aspects of evaluating the speaking skill in secondary classroom. *Shanlax International Journal of English*, 12(2), 9–17. <https://doi.org/10.34293/english.v12i2.7206>
- Si, P. (2019). A study of the differences between EFL and ESL for English classroom teaching in China. *IRA International Journal of Education and Multidisciplinary Studies*, 15(1), 32–35. <https://doi.org/10.21013/jems.v15.n1.p4>
- Sudaryono. (2022). Educational research methods. In *Student Library*.
- Sugiyono. (2019). Quantitative, qualitative, and R&D research methods. In *Student Library*.
- Team, E. (2022). My next words grade 6: Student's book for elementary school. In *Book Center, Education Standards, Curriculum, and Assessment Agency, Ministry of Education, Culture, Research, and Technology*.
- Wahyudi, S., Resci, M., Putri, D., Nadilah, S. P., & Marbun, N. (2024). Analysis of small sample

*The Effectiveness of ELSA Speak Application in Improving EFL Students' Pronunciation*  
hypothesis testing. *ResearchGate*.

Xu, L. (2023). The age effect in second language acquisition and its study design method. *Journal of Education, Humanities and Social Sciences*, 8, 1215–1222.  
<https://doi.org/10.54097/ehss.v8i.4454>