Canadian Journal of Language and Literature Studies
Vol. 4(5), 2024, pp. 22-33

AI and Non-AI Tools in Teaching English Pronunciation to EFL Learners

M. Zaky Mubarok[1] & Feisal Aziez[2]

[1,2] Universitas Muhammadiyah Purwokerto, Banyumas, Indonesia
Correspondence: Feisal Aziez, Universitas Muhammadiyah Purwokerto, Indonesia
Email: m.zaky.mubarok@gmail.com; feisalaziez@ump.ac.id

DOI: 10.53103/cjlls.v4i5.178

Abstract

This study investigated the effectiveness of AI-based and non-AI-based apps in enhancing students' pronunciation skills. A quasi-experimental design was employed with two groups: an experimental group utilizing an AI-based app (ELSA Speak) and a control group using a non-AI-based app (Google Translate). Participants were tenth-grade students from a senior high school in Brebes, Indonesia. Data were collected through pre- and post-tests assessing pronunciation accuracy. Results indicated no significant difference in pronunciation scores between the two groups, with both groups demonstrating superior improvement. These findings suggest that AI-based and non-AI-based apps can be valuable tools for enhancing students' pronunciation skills.

Keywords: Pronunciation, AI-Based App, Non-AI-Based App, Comparative Study

Introduction

Language is the most crucial and fundamental aspect of communication since language represents a significant foundation of communication (Hrehova, 2010). Without adequate language skills, communication will be difficult. In oral communication, pronunciation plays a very important role (Suwartono, 2014b). Every foreign language student aspires to achieve accurate and efficient communication. To effectively interact with others, an individual must possess not just a robust vocabulary and grammar, but also proficient pronunciation. Furthermore, effective communication is impossible without good pronunciation.

Pronunciation might be the most marginalized in the teaching and learning of English (Suwartono, 2014b). Many experts regard pronunciation as an "extraneous" element of the language (Brown & Yule, 1983). But now, pronunciation is no longer an aside in language classes and everybody who works in this area recognizes that pronunciation plays "crucial functions in verbal communication include the listener's

interpretation and the speaker's sense of identity." (Liu, 2008). It appears unusual, but if we examine the responsibilities of pronunciation in detail, we can realize how significant pronunciation in many facets of our lives. It plays a significant influence in both our personal and societal lives. Individuals develop their personalities through their speech and demonstrate their reliance on a social group in society. In the EFL contexts where English is not the first language, inadequate pronunciation abilities can lead to difficulties in oral communication and can hinder success (Arslan, 2013).

Some variables, such as the students' listening comprehension, influence the students' ability to pronounce English. Pronunciation and listening skills are two sides of a coin that cannot be separated. Listening is beneficial for improving students' pronunciation. The more students are exposed to and comprehend spoken English, the more they internalize correct pitch, intonation, stress, and the sounds of individual words as well as those that merge in connected speech. The fundamental aspect of improving students' pronunciation lies in their ability to listen attentively (Bennett, 2013). However, it is crucial that the listening material is both understandable to the learners and relevant to their life outside the classroom. This is because learners should be exposed to many voices and varied modes of communication. It indicates that the more students listen, the better they pronounce. It is important to remember that good verbal communication is dependent not just on the ability to communicate, but also on how well the students listen.

Based on initial observations, the students had difficulty pronouncing words correctly in one of the senior high schools in Brebes, Central Java, Indonesia. Many students in senior high schools in Brebes often encounter significant challenges in acquiring accurate pronunciation. A primary obstacle is the inability to produce spoken English words correctly due to a lack of familiarity with phonetic patterns. Furthermore, unlike in many other languages, the inconsistent grapheme-phoneme correspondence in English presents additional difficulties in associating written forms with their corresponding sounds. As a result, this research offered a plan to address the issues, which included the use of AI (Artificial Intelligence) based apps and non-AI-based apps in the conversation class. The technique was adopted to allow students to learn on their own with the help of modern technologies.

This study is based on numerous studies in language acquisition stated that the more students are exposed to and comprehend spoken English, the more they internalize correct pitch, intonation, stress, and the sounds of individual words and those that merge in connected speech. In this case AI-based apps can provide personalized feedback, adaptive learning, and engaging interactive experiences. AI-based apps can analyze students' pronunciation and provide tailored feedback on their errors, helping them identify areas for improvement. AI-based apps can adjust the students' difficulty level based on students' progress, ensuring that they are constantly challenged. AI-based apps can offer fun and engaging activities, such as games or simulations, that can make learning

pronunciation more enjoyable. Based on these former explanations, the basic assumption of this research is the students who are taught with AI-based apps would do better than those who are taught with non-AI-based apps. Based on the background of the study, the research question of this study is, "Is there any significant difference in students' pronunciation between those who are taught with AI (Artificial Intelligence) based apps and non-AI based apps?"

Research Methodology

This study is experimental research. Experimental research is divided into three based on the degree of tightness of control, i.e., pre-experimental and quasi-experimental, and the most reliable is fundamental experimental research (Suwartono, 2014a). Furthermore, this study employed a quasi-experimental approach to conduct the study, consisting of two groups: experimental and control groups. Both groups will be given a pretest and a post-test to determine the outcome. After experimentation, the final conditions of the two groups were evaluated using statistical tools. Pre-test as a measurement of specific attributes or characteristics that will be examined on experiment participants before they receive a treatment (Creswell, 2008). A post-test is a method used to assess specific attributes or qualities in participants after therapy.

The students in the experimental group received treatment that involved the use of an AI-based app. The control class, on the other hand, received treatment that involved the use of non-AI-based apps. The research design can be seen in Table 1.

Table 1: Research design

Class	**Measurement**	**Treatment**	**Measurement**
Experimental Group	Pre-test	AI-based app	Post-test
Control Group	Pre-test	Non-AI based app	Post-test

The population of this research is two classes of tenth grade, which amounts to 70 students in one of the senior high schools in Brebes, Central Java, Indonesia, as seen in Table 2.

Table 2: Population in tenth grade

Class	**Male**	**Female**	**Total**
X MIPA	6	29	35
X IPS	6	29	35

The study used saturation sampling. Saturation sampling is a sampling strategy that involves collecting several samples equal to the population's size (Sugiyono, 2016). Thus, in this study, two classes were divided into the experimental group and the control group. This study employed simple random sampling to determine the experimental and

controlled groups. Simple random sampling, also called lottery sampling, is a probability-based method for selecting research participants. This technique ensures that every member of the population has an equal chance of being chosen for the sample. Simple random sampling offers several benefits. Firstly, it ensures equal selection opportunities for all population members, promoting fairness. Secondly, it helps to minimize the influence of personal judgment, thereby reducing potential bias in the sample.

A test was employed to collect the data and information required for this study. A test is a set of inquiries or tasks, along with other instruments, employed to assess the aptitude, expertise, intellectual capability, or talent held by a person or collective (Arikunto, 2006). This study employed oral tests to determine the students' pronunciation skills. The oral test emphasized the analysis of segmental characteristics, encompassing vowels, diphthongs, and consonants. The tests consisted of two components: a pre-test and a post-test. The pre-test and post-test were employed to assess the students' pronunciation proficiency and measure the enhancement of students' pronunciation skills before and after the treatment. A system of scores was utilized to examine each student's test results and assess students' English pronunciation skills. Following data collection from the pre-test and post-test, the results of the experimental and control groups were compared. The data collection was conducted on Monday (10th June 2024) – Wednesday (12th June 2024) in the language lab at the school where the study was conducted. The first data collection procedure is, in the experimental group, the teacher asks the students to download ELSA (English Language Speech Assistant) in the Google Play store. In the control group, the teacher asks the students to open Google Translate web. Then, the teacher asks the students to click pronunciation in the ELSA app and train within the app in an experimental group. In the control group, the teacher has prepared some words to be trained with Google Translate web. The researcher got the data in the form of student pronunciation scores in the pre-test and post-test. Then, the data collected was analysed.

A study's findings can be declared valid if there is concordance between the data collected from the subjects and the actual data that transpired in the study's subject (Sugiyono, 2004). A valid instrument as a measuring tool used to obtain accurate results (Sugiyono, 2004). "Valid" means that the instrument may be used to measure what is intended to be measured. Face validity pertains to how an exam appears valid to teachers, other testers, and indications. The test utilized in this study may be suitable for students in senior high school who are at a comparable proficiency level. The test was suitable for tenth-grader students at the senior high school level. The test blueprint can be seen in Table 3.

Table 3. Test blueprint

Number of Items	Indicators	Type of Tests	Content	Score
Item 1 (consonant)	The students can pronounce some minimal pairs of words with correct pronunciation. The item consists of a set of words with similar sounds; the students have to distinguish the words with similar consonants and sounds, and they have to pronounce them correctly. The consonants are: /tʃ/, /dʒ/, /f/, /v/, /θ/, /ð/, /s/, /z/, /ʃ/, /ʒ/, /h/, /m/, /n/, /ŋ/, /l/, /r/, /j/, /w/.	Oral test consonants	chain, jam, fall, van, thin, this, see, zoo, shoe, vision, hat, man, now, sing, leg, red, yes, wet.	18
Item 2 (Vowels)	The students can pronounce some homophone words correctly. The item consists of a set of words with the same sound; the students have to distinguish the words with similar vowels and sounds, and they have to pronounce them correctly. The vowels are: /iː/, /i/, /ɪ/, /e/, /æ/, /ɑ/, /ɒ/, /ɔː/, /ʊ/, /u/, /uː/, /ʌ/, /ɜː/, /ə/	Oral test vowels	see happy, sit, ten, cat, father, got, saw, put, actual, cup, fur, about.	14
Item 3 (diphthong)	The students can pronounce words correctly. The diphthongs are: /eɪ/, /əʊ/, /aɪ/, /ɔɪ/, /aʊ/, /ɪə/, /eə/, /ʊə/	Oral test diphthong	Say, go, my, boy, now, near, hair, pure	8

(https://www.oxfordlearnersdictionaries.com/, n.d.)

To address the research question, the data was gathered from the pre-test and post-test pronunciation scores of students in both two groups. The data was examined using the Social Science Statistics Package (SPSS). The results were organized into a score table and assessed for normality and homogeneity based on the Saphiro-Wilk test. Data was analyzed using the Mann-Whitney test, with the SPPS program utilized. Data analysis is a rigorous endeavour aimed at ensuring data reliability, understanding the nuances of various valuation approaches, and effectively organizing the data. Analyzing this data will enable us to ascertain the disparity in outcomes between the pronunciation achievement scores of students who were taught using AI-based apps and the students who were taught using non-

AI-based apps. The results were interpreted, conclusions were drawn, and a summary was provided.

Results and Discussion

Based on the results of the data obtained from the control group test, it was stated that the difference score between the pre-test and post-test of 35 students produced a minimum score of 10 and a maximum score of 30 with an average of 19.42 and a standard deviation of 4.37. Meanwhile, the difference in the experimental group is that it has a minimum score of 10 and a maximum value of 27.50, with an average of 18.50 and a standard deviation value of 3.89.

Table 4 below presents the mean scores and standard deviations of the students' pronunciation skills for both experimental and control groups in the pre-test and post-test.

Table 4: The mean scores and standard deviations

Groups	Pre-test			Post-test		
	N	X	SD	N	X	SD
Experimental	35	47.00	7.61	35	61.28	6,78
Control	35	45.57	7.55	35	54.57	7.98

A test of normality was employed to examine the normal distribution of the difference scores of both the experimental and control groups. It is observed from the Shapiro-Wilk table that the difference score in the control group is normally distributed because Sig. (2-tailed) <0.05, i.e. 0.024. On the other hand, the difference score in the experimental group is not normally distributed because Sig. (2-tailed) >0.05, i.e. 0.27. See Table 5

Table 5: The result of the normality test

	Kolmogorov-Smirnov			Saphiro-Wilk		
	statistic	df.	sig	statistic	df.	sig
Difference score of the control group	0.16	35	0.02	0.93	35	0.02
Difference score of the experimental group	0.17	35	0.01	0.96	35	0.27

Consequently, a non-parametric test, i.e., the Mann-Whitney U test, has to be utilized to determine the degree of progress in pronunciation between these two groups by comparing the different score results of the experiment class and control class. See Tables

6 and 7.

Table 6: Table of rank

	Group	N	Mean rank	Sum of rank
The difference score	Experiment	35	33.69	1179.00
	Control	35	37.31	1309.00
	Total	70		

Table 7: Table of statistics test

	Difference score
Mann-Whitney	549.00
Wilcoxon W	1179.00
Z	-0.77
Asymp. Sig. (2 tailed)	0.45

The table shows the difference in scores between the experimental and control groups. Ho is not rejected based on the table's conclusion, Asymp. Sig. (2-tailed) =0.45>0.05. It indicates that the difference in score between the control group and the experimental group is not significant. In other words, the difference score between the control group and the experimental group is homogeneous or equal.

The results of this study may make a substantial contribution to our knowledge of how AI-based apps and non-AI-based apps affect the pronunciation of English language learners. AI-based apps, based on several types of research, improved communication in various domains, like lowering speaking anxiety (Dizone, 2017; Hsu et al., 2021). However, several studies have indicated that AI-based apps are comparatively fewer used for pronunciation practice than for other language domains. (Huang et al., 2023; Mousalli and Cardoso, 2020; Zhai and Wibowo, 2023).

Based on the results from the pre-post score test analysis, the score test shows that both groups have improved after one day of conversation class. The result is in line with the study stated students' strong inclination towards pronunciation learning combined with technology (Benzie, 2017). However, comparing the mean scores of the pronunciation pre-test and post-test, the control group and experimental group have made progress in different degrees, although they were relatively at the same level of pronunciation before the treatment.

Multiple studies conducted on ELSA Speak, an AI-based app, such as Pangastuti (2021), align with this research. The findings of Pangastuti's study suggest that the post-

test results exhibit a higher value than the pre-test results. This demonstrates that the utilization of the ELSA Speak App positively impacts enhancing students' proficiency in English pronunciation. The research conducted by Gelu (2020) revealed that ELSA Speak had a good influence on the students. Specifically, it improved students' pronunciation skills, enhanced students' enthusiasm to speak English, and boosted students' confidence in speaking the language.

The finding shows that the difference score in the control group is normally distributed because Sig. (2-tailed) <0.05, i.e. 0.024 and in the control group, is not normally distributed because Sig. (2-tailed) >0.05, i.e. 0.27 based on Shapiro-Wilk. Thus, the Mann-Whitney U test should be employed to determine the degree of progress in pronunciation between these two groups by comparing the different score results of the experiment class and control class.

Mann-Whitney U Test, sometimes called the Wilcoxon Rank Sum Test, is used to compare two samples or groups. It is the non-parametric equivalent of the t-test for independent samples. Therefore, the Mann-Whitney U test is always used when the normal distribution requirement for the t-test is not met.

The finding shows that Asymp. Sig. (2-tailed) =0.45>0.05. It concluded that Ho was not rejected. It indicates that the difference in score between the control group and the experimental group is not significant. In other words, the difference score between the control and experimental groups is homogeneous or equal. No significant result exists between the students' pronunciation skills taught with AI-based and non-AI-based apps.

Human decision-making possesses its opacity and bias, making it a frequent alternative to AI decision-making. Baker et al. (2019) noted that the precision of artificial intelligence can be missed. This is in line with a study conducted by Huang et al. (2023) which stated that the feedback in AI speech evaluation algorithms was primarily limited by a lack of accuracy and challenges in monitoring improvement.

When individuals utilize artificial intelligence, they cannot comprehend the method by which it makes judgments. This, in turn, prevents users from truly comprehending the deep logic, which increases the difficulty of understanding artificial intelligence. AI's detecting accuracy is not perfect; occasionally, a mistake occurs. This relates to the issue of essential algorithms, and it is crucial to research how to enhance pertinent models and algorithms to increase detection accuracy. In this instance, a few sophisticated models or algorithms might be considered. (Qiu et al., 2022). There is currently a dearth of research on models and algorithms in this field.

All technologies occasionally have malfunctions, such as lost connections or malfunctioning applications or programs. One instance of AI technology failing was when the AI provided false information. Another intriguing difficulty that was exposed was limited capabilities. This code alluded to the desire of users for the AI systems to perform more tasks in a more sophisticated manner. According to Thompson et al. (2018), students

desired more incredible advancements with the AI chatbots they were utilizing. Some students thought chatbot conversations were strange. As a result, the students lost interest in utilizing the chatbot. Working with students in Sweden, Ericsson et al. (2023) noted numerous benefits of utilizing virtual persons as conversational agents. Nonetheless, reports indicated that occasionally, students were not heard or understood as intended, which is comparable to Thompson's study. In addition to appreciating the AI's ability to teach English, the study's participants demanded that its capabilities be expanded further to prevent such errors or constraints.

The effectiveness of AI-based apps varied across learners with different degrees of proficiency, emphasizing the necessity of tailoring these tools to meet learners' specific needs and expectations at different stages of language development. Many studies demonstrated the influence of learners' competence levels and familiarity with technology on their interactions with AI-based app. (Huang and Moore, 2023; Mavjudovna, 2023; Yesilyurt, 2023). The teacher and the students encounter certain constraints when utilizing ELSA Speak for teaching and learning pronunciation. Students who encountered difficulties navigating the system felt frustrated and lacked patience. The second problem arises from the fact that the method is time-consuming.

These findings validate the research conducted by Marcais (2002), as referenced in the study of Suppasetseree and Dennis (2010) which emphasizes the importance of incorporating practical and relevant activities in learning speaking and pronunciation. In this scenario, ELSA Speak, with its interactive activities, cannot fulfil this requirement or inspire students in pronunciation sessions.

In order to address these problems, teachers should allocate sufficient time for comprehensive ELSA Speak training for students before implementing the system in the pronunciation course. Thorough preparation from the beginning minimises time wasted. Thus, implementing this approach will result in various varied, creative, and advantageous activities for learners and teachers on ELSA Speak.

Conclusion

This study compared the students' pronunciation skills taught with AI-based apps and non-AI-based apps. The result reveals no significant difference between the students' pronunciation skills taught with AI-based apps and non-AI-based apps. However, the result shows that based on the results from the pre-post score test analysis, after one day of conversation class, the score test shows that both groups have improved.

However, it is essential to recognize the constraints of this study. Primarily, this study examines explicitly the senior high school level, which limits the applicability of the results to other educational levels. Hence, it is imperative to undertake additional and more comprehensive research to explore the potential uses of AI-based and non-AI-based apps

at another educational level. Due to the time constraints and the limited sample size, it is recommended that the scope of the research be expanded and the duration extended to achieve more dependable results.

References

Arikunto, S. (2006). *Prosedur penelitian suatu pendekatan praktek* (6th ed.). Rineka Cipta.

Arslan, R. S. (2013). Enhancing non-native prospective English language teachers' competency in sentential stress patterns in English. *Pamukkale University Journal of Education*, *2*(34), 183–195. https://doi.org/10.9779/PUJE625

Baker, T., Smith, Nandra, A., Sheehan, K., Ward, K., Waters, A., Berditchevskaia, A., Van Den Berg, C Campbell, N., Candsell, O., Casasbuenas, J., Cinnamon, J., Copeland, E., Duffy, E., Hannon, C., John, J., Grant, J., Klinger, J., Latham, M., Macken, C., & Ward-Dyer, G. (2019). *Educ-AI-tion rebooted? exploring the future of artificial intelligence in schools and colleges*. Nesta Foundation. www.nesta.org.uk

Bennett, K. B. (2013). Teaching pronunciation: an independent study course for teachers of adult english as a second language learners. *Journal of Language Sciences and Linguistics*, *Vol.1*, *(1)*, 1–6. http://jlsljournal.com

Benzie, Y. (2017). Contributions of new technologies to the teaching of English pronunciation. *Journal of Language Value*, *9*(1), 135.

Brown, G. and Yule, G. (1983). *Teaching the spoken language*. Cambridge University Press.

Creswell, J. W. (2008). *Educational research: planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education Ltd.

Dizone, G. (2017). Using intelligent personal assistants for second language learning: a case study of Alexa. *TESOL Journal*, *8*(4), 811–830.

Ericsson, E., Sofkova Hashemi, S., & Lundin, J. (2023). Fun and frustrating: Students' perspectives on practising speaking English with virtual humans. *Cogent Education*, *10*(1). https://doi.org/10.1080/2331186X.2023.2170088

Gelu, M. (2020). *Benefits of using ELSA speak in role play class of SMA Negeri 9 Yogyakarta*. Universitas Sanata Dharma.

Hrehova, D. (2010). The significance of communication and language competences for employment and career development. *Human Resources Management & Ergonomics Volume IV*, *2*.

Hsu, H., Chen, H., & G, T. (2021). Investigating the impact of the Amazon Alexa on the development of L2 listening and speaking skills. *Interactive Learning Environments*, 1–14.

https://www.oxfordlearnersdictionaries.com/. (n.d.). *Pronunciation guide*

(English/academic dictionaries). Retrieved July 10, 2021, from https://www.oxfordlearnersdictionaries.com/us/about/pronunciation_english.html

Huang, G., & Moore, R. K. (2023). Using social robots for language learning: are we there yet? *Journal of China Computer-Assisted Language Learning*, *3*(1), 208–230. https://doi.org/10.1515/jccall-2023-0013

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 112–131. https://doi.org/https://doi.org/10.30191/ETS.202301_26(1).0009

Liu, Y. (2008). *The effectiveness of integrating commercial pronunciation software into an ESL pronunciation class*. Iowa State University.

Mavjudovna, K. (2023). A differentiated approach to teaching the Russian language using artificial intelligence. *American Journal of Language, Literacy and Learning in STEM Education*, *6*(1), 210–216.

Mousalli, S., & Cardoso, W. (2020). Intelligent personal assistants: can they understand and be understood by accented L2 learners? *Computer Assisted Language*, 865–890.

Pangastuti, D. (2021). The effect of ELSA Speak aplication on students' pronunciation in English. *Prosiding Pekan Ilmiah Mahasiswa Unis*, 127–133.

Qiu, Y., Wang, J., Jin, Z., Chen, H., Zhang, M., & Guo, L. (2022). Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. *Biomedical Signal Processing and Control*, *72*, 103323. https://doi.org/10.1016/j.bspc.2021.103323

Sugiyono. (2004). *Statistik untuk penelitian*. Alfabeta.

Sugiyono. (2016). *Metode penelitian kuantitatif, kualitatif dan R&D*. Alfabet.

Suppasetseree, S., & Dennis, N. K. (2010). The use of Moodle for teaching and learning English at tertiary level in Thailand. *The International Journal of the Humanities: Annual Review*, *8*(6), 29–46. https://doi.org/10.18848/1447-9508/CGP/v08i06/42964

Suwartono. (2014a). *Dasar-dasar metodologi penelitian*. Penerbit Andy.

Suwartono. (2014b). Enhancing the pronunciation of english suprasegmental features through reflective learning method. *The Asian Conference on Language Learning: Conference Proceedings*, 1–11.

Thompson, A., Gallacher, A., & Howarth, M. (2018). Stimulating task interest: human partners or chatbots? In *Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018* (pp. 302–306). Research-publishing.net. https://doi.org/10.14705/rpnet.2018.26.854

Yesilyurt, Y. E. (2023). *AI-enabled assessment and feedback mechanisms for language learning* (pp. 25–43). https://doi.org/10.4018/978-1-6684-9893-4.ch002

Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*.