



Vol. 4 No. 5 (May) (2026)

Impact of ELSA Speak Application on English Consonant Pronunciation Accuracy among Grade 10 ESL Learners in the Educators School, Nowshera

Umar Javed

MPhil English Linguistics Scholar, Riphah International University Islamabad

Shahrukh Khan

MPhil English Linguistics Scholar, Riphah International University Islamabad

ABSTRACT

English pronunciation is a key skill for communicating with others, but it is among the least developed in a typical Pakistani schools ESL (English as a second language) class. One of the common challenges Pakistani learners face with English is its phonological system, which lacks consonant sounds from their L1. This study investigates English consonant accuracy gain between the use of ELSA Speak, an AI-based pronunciation application, and Grade 10 English as an additional language (ESL) students in a Private school located in Nowshera, Pakistan. Thirty-five Grade 10 students were involved during a 5-week period. The instruments used were a consonant word list, a pronunciation test and a perception questionnaire. Means, percentages, and a paired-samples t-test were used to analyze the data. Results have shown that consonant accuracy significantly improved after the intervention. There were some positive changes for the /θ/, /v/, /z/ and /ð/ sounds. Students' views about ELSA Speak were very positive. The study validates that ELSA Speak is an effective, easily accessible tool for pronouncing consonants at the word level in under-resourced Pakistani ESL contexts.

Keywords: ELSA Speak, consonant accuracy, ESL learners, Mobile Assisted Language Learning, AI pronunciation feedback, Pakistani schools

INTRODUCTION

English is the international language of education, commerce and communication. For ESL students, correct pronunciation is very important in interactions as it paves the way towards clear and effective communication. Even with the decent grammar and vocabulary, clarity in communication can still be affected by poor pronunciation. Intelligibility in spoken English relies heavily on the accuracy of consonant production: misspoken consonants change the meaning and will affect intelligibility (Yulmania, 2024).

ESL learners from Pakistan face a special challenge with English consonants that do not exist in their native languages. Urdu, Punjabi, Pashto, and Sindhi differ significantly from English in their phonological make-up. The learners have difficulty with the sounds /θ/, /ð/, /v/, and /z/, as they sometimes replace /θ/ with /t/ or /s/, and /v/ with /w/ (Hoeriyah, 2024). Little systematic correction during schooling allows these patterns to turn into habits. Only little attention is given to pronunciation in Pakistani schools, as English instruction is mainly focused towards grammar, reading and exam preparation. The problem is aggravated in semi-urban localities such as Nowshera, where there is little or no exposure to English, and teachers lack adequate Phonetics training.

With recent advances in Mobile Assisted Language Learning (MALL), Mobile pronunciation learning has become more flexible. ELSA Speak is one such tool in which an artificial intelligence checks one's pronunciation, and the LMS provides corrections in



Vol. 4 No. 5 (May) (2026)

real time (Adillah, 2022; ANISSA, 2025).

The present study examines how ELSA Speak affects consonant accuracy and identifies the primary area for improving it among Grade 10 students in ESL programs in Nowshera, as well as their opinions on AI pronunciation exercises. The study has three research questions:

RQ1: What impact does ELSA Speak have on the accuracy of English consonants among Grade 10 ESL students?

RQ2: Which English consonant sounds progress the best when using ELSA Speak?

RQ3: What are the perceptions of the use of ELSA Speak for the purpose of pronunciation learning from Grade 10 ESL learners?

Literature Review

The effective intelligibility and communication competence of the second language learner are closely related to pronunciation, a key factor in L2 acquisition (Mishra & Mallik, 2011). Consonants, for example, have high functional load among the segmental features; that is, saying something incorrectly disrupts meaning and understanding (Madrid Valencia, 2024). This means that today, pronunciation instruction is more about intelligibility than native accents. The difficulty ESL learners in Pakistan face with certain consonants are pronounced due to phonological gaps between L1 and L2. This sustainedness is attributed to instructional neglect by Mirzayev (2026), who claims that the private school English syllabus is imbalanced, emphasising grammar and exam performance rather than language use in speaking. There is evidence that this type of learning, which involves mobile applications, can positively impact language development, with specially designed apps offering interactive feedback (Indrayadi, 2026; Ismahani et al., 2026).

The capacity to analyze speech data as it is captured and provides corrective feedback unique to the learner is enabled by AI-based tools for pronunciation analysis. Reskiawan & Riyani (2026) state that interactive feedback is considered to be one of the most effective features of language learning with the aid of technology, particularly in situations where there are no trained pronunciation teachers. In addition, Rikwan (2026) found that CAT significantly improves pronunciation across a variety of educational contexts. In Pakistan, Ulugbek found that MALL with mobile devices drew positive responses from students and increased engagement, but the majority of studies focused on university students, a population still underutilized in MALL research in Pakistan. ELSA Speak has been applied in various international research projects. After 8 weeks of pronunciation training, Malle (2026) reported considerable improvement in the pronunciation of Vietnamese university students. The results of Awololon 2025 found that gains in fricatives were the highest among Indonesian high school students compared to other groups.

The researchers in Aldaghri (2025) found that Thai learners' speaking confidence and pronunciation accuracy are higher. To evaluate the effectiveness of ELSA application, one of the study (Bakori 2025) reported that ELSA Speak users performed better than students who received only traditional classroom instruction, indicating that ELSA Speak yielded better results than the conventional classroom approach. Although significant research on this topic is available, as previously discussed studies focused on Southeast Asian university students learning English as a Foreign Language. Secondly, secondary school learners in the private sector of Pakistan have not been studied in depth, and few studies use consonant-level accuracy as the primary outcome variable.

The present study directly tackles these gaps. The theoretical approach is based upon the sociocultural theory of learning (Vygotsky 1978), which holds that learning is a socially mediated process within the Zone of Proximal Development, through which



Vol. 4 No. 5 (May) (2026)

interactions guide the learning process. The ZPD is the gap between what a learners can do at their own and how it is different when the teacher helps them. ELSA Speak is a learning tool used as a technological scaffold in this study. The application detects when students enunciate a consonant incorrectly and provides an example of correct pronunciation. Learners internalize accurate sound production through repeated structured practice and will gradually reach an independent competence level. Sociocultural Theory is a suitable theory for addressing AI-assisted pronunciation learning because it supports the scaffolding model. The scaffolding model makes Sociocultural Theory a suitable theory for addressing pronunciation learning assisted by AI.

Methodology

Research Design

This study adopted a quantitative, supervised, experimental research design with a pretest and a posttest, in which one group was studied to explore the effect of ELSA Speak on English consonant accuracy (Creswell, 2023). A baseline pronunciation exam was administered at the outset of the study, and a repeat exam was administered following 5 weeks of practice. The pronunciation test consisted of three items: a 30-item test focusing on some problematic consonants in the input /v/, /z/, /θ/, /ð/, /f/, /tʃ/, /dʒ/, /f/; a consonant analysis sheet for calculating sound-wise accuracy and a 12-item Likert scale questionnaire for scrutinizing the learner perceptions. The accuracy of students' recorded answers was checked by two assessors, with inter-assessor agreement at a high level (Cohen's Kappa = .87).

Settings and Participants

The study was conducted at The Educators School in Nowshera, KPK. The sample consisted of 35 students (15 – 16 years old), Grade 10, from one intact classroom. All participants were native Urdu / Pashto speakers and had studied English as a compulsory subject. None were experienced with ELSA Speak or other AI-based pronunciation software. The researchers used convenience sampling because of the access to the school. They were given the option to volunteer and provided written permission to the parents or guardians prior to participation. Ethical principles were followed throughout the study. The anonymity of students was ensured through data coding; participants could withdraw from the study at any time, and the data was securely stored. The present study had several limitations, including a small sample size, a school-based design, a short time frame, and the absence of a control group.

Data Collection Method and Analysis

Five weeks of data were collected. In Week 1, students were given the pre-test and training on installing and using ELSA Speak. They received practice using the application for 20-30 minutes per day during Weeks 2-5. The post-test and questionnaire were given in Week 6. Data were analyzed using SPSS version 26 and included mean scores, percentages, and paired-samples t-tests. The frequencies and percentages were used to analyze the responses from the questionnaires.

The pretest and posttest were conducted from the same 35 students of the class and results attained are illustrated in the following table:-

Table 1: Consonant Overall Accuracy: Pre-test and Post-test Comparison

Test	N	Mean Score (%)	Standard Deviation
Pre-test	35	51.4	8.72
Post-test	35	73.8	7.16

The mean score rose from 51.4% to 73.8%, with a difference of 22.4 percentage points. This is both a numerical and an educational improvement, reflecting a clear and



Vol. 4 No. 5 (May) (2026)

measurable increase in consonant accuracy after five weeks of structured practice with AI support. The pre-test mean is 51.4%, which substantiates what existing literature has already reported: that private school learners at secondary level in Pakistan suffer from poor accuracy of consonants despite the number of years these students have spent learning English and that, broadly, pronunciation is not addressed in students' classrooms systematically (Akhtar et al., 2020; Manzoor et al., 2021). The decrease in SD (8.72 to 7.16) is also statistically significant. An expanded range of performance was observed in this group after the intervention, suggesting greater homogeneity.

The data suggest that improving speaking skills, specifically pronunciation, was most advantageous for weaker learners, who are likely the population most in need of specific pronunciation assistance. This levelling effect is probably due to the application providing personalized, specific feedback regardless of a learner's proficiency level. Unlike whole-class instruction, which targets instruction to the average student, ELSA Speak was tailored to focus instruction on each student's error profile, with practice targeting each student's specific needs.

Statistical Significance

Paired Sample T-test Table 2 shows the Paired Sample T-test results to determine if the perceived improvement is considered statistically significant.

Comparison	Mean Difference	SD	t-value	Df	p-value
Pre-test – Post-test	22.4	5.63	23.57	34	.000

The difference in consonant accuracy was statistically significant at $p < .001$. The large t-value of 23.57 indicates a large, consistent treatment effect across the treatment group as a whole, rather than in a smaller subgroup of high performers. The first research question is directly answered, and ELSA Speak had a statistically significant positive effect on the accuracy of English consonant production in Grade 10. The results of traditional phonics instruction in Pakistani private schools haven't yielded the same improvements year after year. There is an important distinction between interactive, feedback-based practice and passive exposure, highlighted here. ELSA Speak puts learners in an active position, asking them to make the sound again and again, providing real-time error identification, and adjusting their pronunciation accordingly (Mirza, 2015). To answer the second research question, individual consonant accuracy rates are shown for each administration in the Sound-Specific Improvement.

Table 3: Improvement Rates for Individual Consonants

Consonant	Pre-test (%)	Post-test (%)	Gain (%)
/θ/	31.4	68.5	+37.1
/v/	38.2	74.6	+36.4
/z/	42.3	77.8	+35.5
/ð/	34.7	69.2	+34.5
/ʃ/	60.1	82.4	+22.3
/tʃ/	65.4	84.7	+19.3
/f/	70.5	87.3	+16.8
/dʒ/	67.2	83.6	+16.4

Improvements in the individual sounds are coherent as a pattern and pedagogically informative. These are the four consonants /θ/ /v/ /z/ and /ð/, which show the greatest improvement; all of them are not included in the Urdu phonology. The pre-test scores are very low for each of them (31.4%–42.3%), indicating that learners had not received meaningful instruction or natural exposure to these sounds before the intervention



Vol. 4 No. 5 (May) (2026)

commenced. Despite this initial deficiency, all four sounds earned a score above 68% on post-test, and /z/ earned 77.8% and /v/ earned 74.6% on post-test. The achievement of these learners over 5 weeks is impressive, and the phoneme-level identification of their errors ensured an effective level of targeted teaching.

The contrastive approach also has a predictability level, and Lado's (1957) approach, the Contrastive Analysis Hypothesis (CAH), says that learners will have the most difficulty with phonemes that appear in L1 but do not in L2, and that these phonemes will be the most successful in responding to targeted instruction once mastered in terms of articulatory requirements.

The greatest consonant changes in fricative sounds not present in the L1 of the Indonesian learners occurred when ELSA Speak was employed. When an application is consistent across two contexts, confidence is strengthened that it works for learners with no key English consonant contrasts in their L1. As per the outcomes of one of the studies by (Ahmed & Aziz, 2022), /θ/, /ð/, /v/ and /z/ were identified as the four most challenging consonants for Pakistani learners, and these consonants were found to be most sensitive to the effects of AI-assisted intervention in the present study. Response to key items from the perception questionnaire is summarized in Table 4, which provides an answer to the third research question.

Table 4 presents the students' Perception Questionnaire results

Statement	Agree / Strongly Agree (%)
ELSA Speak is easy to use	88.6
ELSA Speak improved my pronunciation	91.4
I enjoyed practicing with ELSA Speak	85.7
The feedback was clear and helpful	82.9
I feel more confident speaking English	80.0
I want to continue using ELSA Speak	94.3

The results revealed that 91.4% believed their pronunciation had improved with ELSA Speak, which again aligned with the objective tests and confirmed their perception of ELSA Speak. Overall, the overwhelmingly positive response to continuing to use the app after formal instruction (94.3%) is very encouraging, as ongoing, sustained language use by learners outside formal instruction is one of the best predictors of long-term language development. This high ease-of-use score is particularly important in the context of private schools in Pakistan, where students' use of digital tools varies and access to cutting-edge technology has been restricted. The rating here indicates that ELSA Speak's interface is somewhat user-friendly, especially for secondary learners, and not technologically sophisticated, making it a viable option for broader use in similar under-resourced contexts. The 80.0% who reported becoming more confident about their pronunciation when speaking in English is about an area of pronunciation learning not captured in test scores. Willingness to communicate is highly correlated with confidence in spoken production, which, in turn, predicts actual language use and long-term language acquisition. The positive ratings students reported for 85.7% of students further support this interpretation, as positive affect when told to practice the target language is correlated with effort, persistence, and ultimate achievement in learning it.

Discussion

The overall findings of this research study support the conclusion that the effect of ELSA Speak is statistically and linguistically significant, resulting in an overall improvement in English consonant (accuracy) attainment in a medium-sized Pakistani private school across the three dimensions studied: overall accuracy, sound-specific



Vol. 4 No. 5 (May) (2026)

development, and learner perception. The pattern of improvement, with the greatest improvement for sounds absent in L1 and a decreasing proportion for sounds with partial L1 equivalents, is in line with the theoretical framework of the CAH and with pedagogies that foreground intelligibility. Vygotsky's Sociocultural Theory states that ELSA Speak was a technological scaffold because it provided learners with corrective feedback and repetition of pronunciation within their Zone of Proximal Development (ZPD). There was a gradual internalisation of correct consonant production with this support.

Shadiev and Arosy (2025) highlight that technology is most effective in language learning when it provides responsive, interactive feedback rather than passive content delivery. ELSA Speak is an example of using this, with real-time error detection and personalised corrective feedback. The results of Aldaghri 2025, which revealed significant improvements in quantitative analysis and highly positive student perceptions, also supported the finding that AI pronunciation feedback produced better results than teacher-oriented conventional instruction, especially for students who had not received any phonological feedback in the conventional learning process.

Conclusion and Recommendations

The purpose of this study was to explore the role of ELSA Speak in promoting accuracy in English consonant production among grade 10 English as a second language (ESL) learners at a Private school located in Nowshera. The mean consonant accuracy scores increased from 51.4% in the pre-test to 73.8% in the post-test, with a gain of 22.4%. This improvement was confirmed as statistically significant using the paired-samples t-test ($p < .001$). Among them, the greatest improvement was achieved for /θ/, /v/, /z/, and /ð/, which are absent from Urdu phonology and are frequently mispronounced by the Pakistani students. The students also gave positive feedback: after using the application, 91% stated that it helped their pronunciation, and most expressed a willingness to continue using it.

The findings suggest that ELSA Speak is an effective and practical tool for implementation in under-resourced secondary schools serving ESL learners in Pakistan. It has not only value for enhancing test performance but also in tackling the need for pronunciation instruction (NPI). A large number of private school students have studied English for many years, yet have not acquired the fundamental prerequisite for the accurate use of consonants. Thus, ELSA Speak aims to bridge this gap by offering real-time feedback to each learner on a low-cost, widely accessible, and easy-to-use platform.

ELSA Speak is designed to serve as a complementary resource for pronunciation practice, enhancing classroom teaching and learning. There is some guidance that a teacher needs to provide a student on how to use the app and on encouraging the use of difficult consonant sounds, specifically /θ/, /v/, and /ð/. Regular monitoring will enable the identification of any pupils who require extra support.

In future studies, larger samples taken from schools across multiple school districts could facilitate generalizability. Further studies are necessary to investigate whether or not pronunciation improvements are maintained following intervention. Classroom observations and interviews are qualitative techniques that can give insight into learners' experiences. Comparative studies within the field of Artificial Intelligence-based pronunciation research and the study of suprasegmental features (stress, rhythm, and intonation) would also enhance understanding of technology-based pronunciation learning in Pakistan.

REFERENCES

Adillah, D. (2022). *IMPROVING STUDENTS' PRONUNCIATION THROUGH ELSA SPEAK APPLICATION OF MADRASAH ALIYAH OF DARUL ISTIQOMAH*



Vol. 4 No. 5 (May) (2026)

- CILALLANG BOARDING SCHOOL* (Doctoral dissertation, Institut Agama Islam Negeri (IAIN) Palopo).
- Aldaghri, A. A. (2025). Consonant Pronunciation Errors Made by Saudi EFL Students. *Journal of Language Teaching and Research*, 16(5), 1640-1646.
- ANISSA, W. S. (2025). *THE INFLUENCE OF ELSA SPEAK APPLICATION TOWARDS STUDENTSENGLISH PRONUNCIATION MASTERY* (Doctoral dissertation, UIN RADEN INTAN LAMPUNG).
- Arosy, H. N. (2025). *The Effectiveness Of The Elsa Speak Application in Facilitating Students' English Pronunciation In Story Telling* (Doctoral dissertation, IAIN Kediri).
- Awololon, Y. O. L., Erfiani, N. M. D., Dewi, N. L. D. S., & Wirasa, K. S. (2025). English Consonant Errors Pronounced by Seventh Grade Students in Dwijendra High School. *Jurribah: Jurnal Riset Rumpun Ilmu Bahasa*, 4(1).
- Bakori, H. A., & Alsofi, E. A. (2025). Pronunciation Challenges Faced by Libyan EFL Learners. *Journal of Human Sciences*, 24(2), 136-142.
- Hoeriyah, S. I. T. I. (2024). The Effectiveness of ELSA Speak Toward Students' Pronunciation in the Eleventh Grade of SMKN 2 Purwokerto. *State Islamic University Prof. KH Saifuddin Zuhri Purwokerto. University repository*.
- Indrayadi, T. (2026). PRONUNCIATION LEARNING STRATEGY RELATIONSHIP WITH PRONUNCIATION PERFORMANCE. *JURNAL BASIS*, 13(1), 29-40.
- Ismahani, S., Putri, M., Siregar, P. A. S., Zakiah, A. F. R., & Turnip, R. B. (2026). A CONTRASTIVE ANALYSIS OF ENGLISH AND INDONESIAN CONSONANT CLUSTERS AND THEIR IMPACT ON EFL LEARNERS' PRONUNCIATION ERRORS. *Journal Of Human And Education (JAHE)*, 6(1), 1-8.
- Madrid Valencia, A. J. (2024). *Error analysis of english consonant pronunciation in efl learners* (Doctoral dissertation, Ecuador: Pujilí: Universidad Técnica de Cotopaxi (UTC)).
- Malle, I. R., Fatsah, H., & Mahmud, M. (2026). Students' Beliefs And Strategies In Learning Pronunciation: A Narrative Research. *Journal of Language Education (JoLE)*, 4(1), 1-6.
- Mirza, H. S. (2015). ESL and EFL learners improve differently in pronunciation: The case of Lebanon. *Procedia-Social and Behavioral Sciences*, 199, 486-495.
- Mirzayev, E. (2026). Pronunciation Difficulties Among Azerbaijani Learners of English: A Phonetic Analysis. *Acta Globalis Humanitatis et Linguarum*, 3(1), 265-272.
- Mishra, S. K., & Mallik, A. (2011). Strength for Today and Bright Hope for Tomorrow Volume 11: 3 March 2011.
- Reskiawan, B., Abin, R., & Riyani, N. (2026). English Pronunciation Problems among Indonesian Students of English as a Foreign Language at University Level: A Descriptive Analysis. *Journal of English Language and Education*, 11(2), 828-836.
- Rikwan, M., Oktarini, W., & Nuraini, K. (2026). An Analysis of Madurese Language Interference in the English Pronunciation of EFL Students at SMP Negeri 1 Randuagung. *Journal of Educational Sciences*, 10(1), 1463-1472.
- Ulugbek, M. PHONETIC VARIATION AND TRANSFER PATTERNS IN ENGLISH LANGUAGE LEARNING.
- Yulmania, S. (2024). *Investigating EFL Learners' Pronunciation Learning Experience Using ELSA Speak Application* (Doctoral dissertation, UIN KH Abdurrahman Wahid Pekalongan).