International Journal of Education, Vocational and Social Science
https://e-journal.citakonsulindo.or.id/index.php/IJEVSS
E-ISSN : 2963 4954
IJEVSS

Exploring Sasaknese EFL Students' Perceptions Of Using ELSA Speak Application As An AI-Based Tool For English Pronunciation Practice

Ardiani Safitri[1]
[1]State Islamic University of Mataram
Ika Rama Suhandra[2]
[2]State Islamic University of Mataram
Sri Hastuti Novilla Anggraeni[3]
[3]State Islamic University of Mataram

Corresponding Author: *ardianisafitri18@gmail.com , ikaramasuhandra@uinmataram.ac.id, srihastuti@uinmataram.ac.id*

Abstract

This qualitative study aims to explore the perceptions of Sasaknese EFL students on ELSA Speak, an AI-powered tool for English pronunciation improvement. Data of the research were collected through semi-structured interviews, classroom observations and reflective journaling from twelve students of UIN Mataram and then analyzed thematically. The findings revealed three main themes: positive perceptions, challenges, and recommendations for education. The application's gamified motivational features, real-time phoneme-level feedback, and low-anxiety setting for independent pronunciation practice were highly appreciated by participants, especially for sounds that are not included in the Sasak phonological system. Participants did, however, also note several difficulties, such as restricted internet connectivity in rural locations, insufficient explanatory feedback for articulatory correction, and accent bias in the voice recognition engine. Instead of using ELSA Speak as a stand-alone tool, participants suggested incorporating it into a teacher-mediated blended learning framework. The study provides theoretical insight into AI-assisted pronunciation learning among an underrepresented ethnolinguistic minority community and is based on Vygotsky's sociocultural theory and the Technology Acceptance Model. The results add to the small amount of qualitative research on AI pronunciation aids outside of East Asian and mainstream student populations and have pedagogical implications for EFL education in multilingual environments.

Keywords: : Elsa Speak; AI-based pronunciation tool; EFL learners; Sasaknese

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

International Journal of Education, Vocational and Social Science
https://e-journal.citakonsulindo.or.id/index.php/IJEVSS
E-ISSN : 2963 4954
IJEVSS

INTRODUCTION

English has become the dominant language of global communication and the primary language of instruction and interaction in education, business, technology, and diplomacy. Among the various dimensions of the English language competence, pronunciation is of particular importance in the overall effectiveness of oral communication. (Celce-Murcia, 1996) claims that a bad pronunciation can alter the speaker's meaning and eventually make communication meaningless. The substantial phonetic differences between English and their mother languages are particularly challenging for students learning English as a foreign language (EFL). A phrase that is mispronounced has no meaning, even if it is grammatically correct. Phonetic mistakes can generate serious misconceptions and change the meaning of a message.To overcome profound cultural misunderstandings, develop professional confidence, and attain true fluency in a multicultural environment, EFL students must become proficient in certain phonetic nuances. By mastering these articulatory elements, non-native speakers are guaranteed to be both grammatically correct and intelligibly functioning, preventing breakdown in communication. Ultimately, targeted pronunciation instruction transforms theoretical language knowledge into a powerful, reliable tool for global collaboration and effective daily communication.

The Sasak are the primary indigenous ethnic group in the Indonesian province of West Nusa Tenggara on the island of Lombok. The Sasak people, an indigenous community with a strong feeling of cultural identity, speak Sasak, which is the greatest way to express Sasak. Phonological features, morphological features and syntactical features have their own in the Sasak language, It belongs to the Austronesian language family, related to Javanese, Balinese. The Sasak language is very important for the people and has a rich tradition of oral literature, handed down from generation to generation. The primary form of daily communication among the Sasak people.It is used in a number of important social areas, from family life and the carrying out of sacred, traditional ceremonies, to the local commercial activities on the market place. This multifaceted role has kept the Sasak language as a pillar that maintains the integrity of the social order while preserving noble values and cultural heritage across generations in Lombok.

Therefore, English serves as a third language for the majority of Sasak EFL learners. They live in a highly complex multilingual environment, where the process of mastering English pronunciation often faces significant obstacles due to double phonological interferenceoriginating from Sasak as their native language and Indonesian as the national language. From a sociocultural perspective, the majority of the Sasak population is Muslim, with many of them pursuing their education at various Islamic schools in the province. Although the participation rate of Sasak

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

learners in higher education is steadily increasing supported by expanded access to English through the academic ecosystem and digital media this unique population group remains significantly underrepresented. Their linguistic phenomena remain largely marginalized within the broader landscape of applied linguistics research networks and global studies on international language education curricula at present.

Sasaknese learners are likely to find it difficult to produce these sounds accurately in English. In a study of EFL From the linguistic point of view, the phonological system of Sasak is quite different with that of English. Sasak lacks some phonemes that are central to English, such as dental fricatives /θ/ and /ð/, the labiodental fricative /v/ and many suprasegmentals such as lexical stress and patterns of intonation at the sentence level. These sounds are not systematically encountered by Sasaknese learners during their early language development. in Lombok, found that the negative phonological transfer from the students' native language caused more than 67% of the participants to have problems with pronunciation. These findings highlight the urgent need for targeted pronunciation instruction and innovative pedagogical interventions that are sensitive to the specific linguistic background of Sasaknese learners.

Given the growing need for effective and accessible language learning tools, the Indonesian Ministry of Education, Culture, Research and Technology (Kemendikbudristek) has been keen to promote the use of technology in educational practices. This is a global trend, with more than 60% of schools in developing countries using computer-assisted language learning (CALL) in their programs, according to a 2023 UNESCO report. This suggests that everyone views technology as a transformative tool for language learning. Speak or English Language Speech Assistant has received quite a lot of attention among all AI products appearing on the market. ELSA Speak aims to help users learn English pronunciation (ELSA, n.d.) with Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) technologies that analyze the learner's speech in real-time, provide instant feedback at the phoneme level, and offer customized pronunciation practice based on each user's needs. The app has been downloaded over 22 million times in 195 countries and is used as a communication training tool by over 100 Fortune 500 companies, according to a 2023 report from (ELSA, n.d.). ELSA Speak is a promising instructional tool for EFL contexts in formal education and is already used in over 1,000 institutions across Asia including Indonesia.

There aren't many empirical research on how Sasaknese EFL learners view and interact with AI-based pronunciation aids, despite their increasing availability and use. East Asian nations

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

including China, Korea, and Japan were the sites of the majority of earlier research on ELSA Speak and associated technology (Kholis, 2021) . Southeast Asian students and members of ethnic linguistic minority, such the Sasaknese, were largely ignored. (Akhmad & Munawir, 2022; Sholekhah & Fakhrurriana, 2023)conducted studies in Indonesia that mostly employed a quantitative method with an emphasis on quantifiable improvements in pronunciation ratings. Relatively little research has been done on the qualitative aspect, which includes students' subjective experiences, perceived difficulties, and subjective assessments of the application. Technology Acceptance Model (TAM), which identifies perceived utility and perceived ease of use as the two primary determinants of technology adoption, serves as the theoretical foundation for understanding students' perceptions. (Dörnyei, 2014) highlighted the following about the motivating aspect of language learning.

Motivation is one of the most crucial elements in determining success when learning a second language (L2). These internal and external motivations serve as the primary driving force behind beginning the difficult process of learning a new language. Furthermore, motivation serves as a consistent source of energy to support the learning process, which is sometimes difficult, tiresome, and demanding.Even those with the most exceptional cognitive skills or academic aptitude will undoubtedly fail and be unable to accomplish their long-term objectives in the absence of strong enough drive. This demonstrates that a well-designed curriculum and effective teaching methods from an instructor, on their own, will never be sufficient to guarantee a student’s overall success in mastering a second language unless they are accompanied by a high level of commitment and enthusiasm for learning from within the learner themselves.

The research is based on a sociocultural perspective to language acquisition(Vygotsky, 1980) and the Technology Acceptance Model (Davis, 1989). The results are anticipated to fill the lacunae of the current research, add to the knowledge on the impact of learners’ ethnicity and linguistic background on their engagement with AI-based pronunciation tools and offer useful information for the development of culturally responsive English language instruction in Indonesia. In order to capture the subjective experiences, attitudes, and evaluative judgments of Sasak learners that cannot be sufficiently addressed by quantitative assessment alone, this study employs a qualitative approach. The study's theoretical foundations are a sociocultural view of language acquisition (Vygotsky, 1980) and the Technology Acceptance Model (Davis, 1989). The results are anticipated to close research gaps, improve knowledge of how learners' linguistic and ethnic backgrounds affect their use of AI-based pronunciation tools, and offer helpful information for the creation of culturally sensitive English language instruction in Indonesia.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

METHODOLOGY

This study was a qualitative study using a collective case study design. The approach is selected since the research is aimed to investigate the perceptions of Sasaknese EFL students on the use of ELSA Speak as an AI-based pronunciation practice tool which is a phenomenon in a high context and not measurable by numbers. According to (Creswell & Creswell, 2017) qualitative research is most appropriate when the aim is to obtain a nuanced understanding of human experience in its natural environment. (Yin, 2017) further argues that case study design is particularly fitting when the focus of inquiry involves a contextually bounded case. In this study, multiple individual cases were examined collectively to illuminate a shared phenomenon from varying perspectives, consistent with what (Stake, 1995) refers to as a collective case study.

The venue was chosen because it has an active English education program, the majority of the pupils are Sasaknese, and it is easily accessible for gathering detailed data. The participants were selected using purposeful selection, a sampling technique that selects individuals based on their relevance to the research issues rather than for statistical representativeness (Creswell & Creswell, 2017) .A Sasaknese ethnic background with Sasak being the primary language of communication in the family since childhood, using the ELSA Speak application for at least four weeks, and being willing to participate voluntarily by signing an informed consent form were all prerequisites for inclusion in the English Education Department at UIN Mataram. The inclusion criteria required participants to be active students in the English Education Department at UIN Mataram, of Sasaknese ethnic background with Sasak used as the primary language of communication within the family since childhood, to have used the ELSA Speak application for a minimum of four weeks, and to be willing to participate voluntarily by signing an informed consent form. The target number of participants was 8 to 12 students, in line with the principle of theoretical saturation the point at which further participants no longer yield substantially new insight. To supplement student perspectives, 2 to 3 lecturers teaching pronunciation-related courses were also involved as supporting informants.

Data were collected through multiple techniques to ensure triangulation, which is fundamental to establishing the trustworthiness of qualitative findings (Lincoln & Guba, 1988). The primary method was in-depth semi-structured interviewing, conducted individually with each student participant. This method was preferred because it allows the researcher to follow a prepared guide while remaining sufficiently flexible to pursue emerging and relevant threads through probing questions (Creswell & Creswell, 2017). With the participants' permission, each interview was audio recorded for verbatim transcription, lasted between 45 and 75 minutes, and

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

was conducted in Indonesian to facilitate comfortable and natural self-expression. To record how participants truly engaged with ELSA Speak during practice sessions, non-participant observation was conducted in addition to interviews. The researcher recorded in-depth field notes on observable behaviors, including patterns of interaction with the application interface, responses to AI-generated feedback, repetition techniques, and outward manifestations of engagement or difficulty, during each participant's two to three sessions of observation. In order to record shared experiences and peer dynamics that might not come up in individual interviews, a Focus Group Discussion (FGD) was also conducted once or twice with groups of four to six participants (Morgan, 1997). analysis served as a supplementary technique, involving the review of participants' ELSA Speak progress screenshots obtained with consent, as well as relevant departmental documents including pronunciation course syllabi.

Research Instrumen and Data Analysis The researcher served as the principal instrument of data collection and analysis in accordance with the epistemological attitude of qualitative inquiry; this function is referred to as the human-as-instrumen (Lincoln & Guba, 1988). A semi-structured interview guide organized around five thematic dimensions reported usefulness, perceived ease of use, motivation, self-confidence, and problems encountered was one of the auxiliary tools created to support this job. An FGD facilitator guide, an observation checklist, and a daily reflection sheet that participants filled out during the data collecting period were additional aids. The thematic analysis framework created by (Braun & Clarke, 2006), which provides a methodical yet adaptable process for finding, examining, and summarizing patterns within qualitative data, was used to examine the data. The analysis proceeded through six recursive phases: familiarization with the data, generation of initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the analytical report. The coding process was conducted manually and supported, where appropriate, by qualitative data analysis software such as NVivo.

Trustworthiness and Ethical Considerations Four criteria defined by (Lincoln & Guba, 1988) were used throughout the research procedure to guarantee the study's rigor and reliability. Data source triangulation, member checking returning preliminary results to participants for confirmation and extended fieldwork to foster rapport and enhance contextual understanding were all used to establish credibility. Thick description was used to address transferability, giving readers enough contextual information to evaluate the findings' applicability in different contexts. A thorough audit trail recording every methodological choice made throughout the study ensured dependability. Ongoing researcher reflexivity was used to assure confirmability, and personal

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

presumptions and potential biases were regularly recognized and recorded to reduce their impact on data interpretation. Additionally, this study complied with accepted ethical guidelines, such as voluntary participation, the use of pseudonyms to preserve participant identity, and the safe preservation of all recorded data solely for scholarly reasons.

RESULT AND DISCUSSION

This study investigated how Sasaknese EFL students felt about ELSA Speak, an AI-powered tool for practicing English pronunciation. Semi-structured interviews, classroom observations, and reflective journals were used to collect data from twelve purposefully chosen participants who were enrolled in the State Islamic University of Mataram English Education Department. Three main topics emerged from the data's thematic analysis: (1) favorable opinions of ELSA Speak as an independent pronunciation learning tool; (2) difficulties faced when using the application; and (3) pedagogical suggestions for incorporating ELSA Speak into formal English training. Table I presents these themes together with their supporting participant voices and sub-themes, which are further explained in the following subsections.

Table I. Thematic Summary of Sasaknese EFL Students' Perceptions of ELSA Speak

No.	Theme	Sub-Theme	Key Finding from Participants
1	**Positive Perceptions of ELSA Speak**	Real-time phoneme feedback	In order to recognize and self-correct mispronounced phonemes like /θ/, /ð/, and /v/, which are lacking from the Sasak phonological system, participants desired immediate corrective input.
		Gamification and motivation	Daily streak features, accomplishment badges and cumulative scores motivated participants to practice English pronunciation more frequently and autonomously outside of the classroom.
		Low-anxiety learning environment	Practicing in a private digital setting reduced fear of speaking compared to in-person classroom activities, allowing for more

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

No.	Theme	Sub-Theme	Key Finding from Participants
			unrestrained pronunciation experimentation.
2	**Challenges Encountered During Use**	Accent bias in AI recognition	When the speech recognition engine frequently marked understandable Sasaknese-accented English as incorrect, several participants reported feeling agitated and less productive.
		Insufficient remedial feedback	Participants discovered that the app's feedback was insufficient to direct articulatory modification; they were unable to determine how to physically modify tongue position or airstream to create the desired sounds.
		Connectivity and access constraints	The internet access in semi-urban and rural areas of Lombok was not always stable, which caused disrupted practice sessions and some participants were not able to use the app regularly.
3	**Pedagogical Recommendations for Integration**	Blended learning framework	Participants suggested integrating ELSA Speak homework with in-class debriefing sessions, where instructors can provide articulatory coaching and metalinguistic explanation that the app cannot offer.
		Formative assessment supplement	Lectures could use in-app progress reports to identify common phonological problems and provide targeted pronunciation practice in class, suggested participants.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

No.	Theme	Sub-Theme	Key Finding from Participants
		Peer collaborative practice	Participants proposed structured peer review sessions where students compare ELSA Speak feedback results, discuss shared pronunciation errors, and support one another's improvement within the sociocultural learning context.

Note. Themes and sub-themes were derived from thematic analysis of interview transcripts, classroom observation notes, and reflective journals (N = 12 participants).

Theme 1: Positive Perceptions of ELSA Speak as a Pronunciation Learning Tool

All twelve participants reported positive overall perceptions of ELSA Speak, describing the application as an effective and accessible tool for autonomous pronunciation practice. Participants generally valued the real-time, phoneme-level feedback of the application as more immediate and targeted than the corrective feedback available in a typical classroom setting (Liakin et al., 2015). Participants reported improvement in producing sounds not found in the Sasak phonological inventory. This includes the labiodental fricative /v/, and the dental fricatives /θ/ and /ð/. In English with Sasaknese accent, the sounds are often replaced because of the influence of the first language (Davis, 1989). One participant said: “I was not aware that I was pronouncing ‘think’ as ‘tink’ until I used ELSA. “The app told me what sound was wrong and I had to repeat it to get it right.

Another key element that helped to affect positive perceptions was the gamification features of ELSA Speak. Daily streak rewards and achievement badges motivated participants to practice pronunciation more often and on their own. This corroborates findings on how gamified features support language learners’. Moreover, participants experienced less anxiety about speaking when they used the app to practice pronunciation over classroom activities. The digital space at low pressure allowed the learners a safe space to experiment with new sounds without fear of peer feedback or judgment. This is in line with previous studies on the emotional benefits of technology use for language practice (Ducate & Lomicka, 2009)

Theme 2: Challenges Encountered During ELSA Speak Use

Despite their broadly positive attitudes, participants identified several challenges that

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

moderated their experience with ELSA Speak. The most pervasive difficulty was the application's limited sensitivity to Sasaknese-accented English. Participants recounted repeated instances in which phonetically intelligible utterances were scored as incorrect by the AI speech recognition engine, a phenomenon attributable to accent bias embedded in the system's underlying training data (Khakurel, 2025). One participant expressed frustration: "I said the word correctly my friend and my lecturer confirmed it but ELSA kept giving me a low score. It made me feel like giving up." This recurrent misevaluation was reported to negatively affect learner self-efficacy and, in some cases, diminished participants' motivation to continue regular practice.

A second cluster of challenges concerned the quality of the application's remedial feedback. The participants enjoyed hearing about phonetic errors but, uniformly, reported that the explanatory component was insufficient to guide physical articulatory adjustment. Learners struggled with how to adjust tongue position, lip configuration, or airstream dynamics to approximate the target phoneme. This gap points to a more general limitation of current AI-based pronunciation tools which tend to be good at identifying errors but not so good at providing pedagogical scaffolding for remediation (Wood, 2022). Furthermore, the intermittent network connectivity disrupted the continuity of practice, especially for those residing in semi-urban and rural areas of Lombok where the broadband infrastructure is still not robust

Theme 3: Pedagogical Recommendations for Integrating ELSA Speak into Formal Instruction

The third theme consolidates participants' recommendations for maximising ELSA Speak's pedagogical value within the formal EFL curriculum. The majority of participants advocated for a blended learning approach in which structured ELSA Speak homework tasks are paired with in-class debriefing sessions facilitated by the lecturer. According to the suggested approach, the diagnosis would begin with the AI-generated feedback from the application, and the teacher would supply the contextualized articulatory instruction and metalinguistic explanation that the application is unable to provide. This suggestion aligns with (Vygotsky, 1980) sociocultural framework, which suggests that technology-mediated learning is most effective when it is scaffolded within the Zone of Proximal Development through expert mediation, and the Technology Acceptance Model, which contends that perceived usefulness is increased when technology tools are embedded in familiar instructional routines (Davis, 1989; Venkatesh et al., 2003)

Additionally, some suggested that instructors use ELSA Speak's in-app progress reports as formative assessment data, with focused classroom assistance for common phonological

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

challenges. In order to reflect the collectivist learning orientation typical of Sasaknese educational culture and in line with sociocultural approaches to language learning, others proposed structured peer collaborative sessions where students compare their ELSA Speak feedback, discuss common errors, and support one another's pronunciation development (Godwin-Jones, 2017; Vygotsky, 1980). Together, these suggestions, as shown in Figure 1, position ELSA Speak as a supplemental tool that, when pedagogically guided and culturally contextualized, can significantly support Sasaknese EFL learners' pronunciation development within their complex multilingual environment rather than as a substitute for teacher-led pronunciation instruction (Shadiev et al., 2026)

This study explored the opinion of the Sasaknese EFL students of State Islamic University of Mataram towards ELSA Speak, an AI-based English pronunciation practice tool. (Vygotsky, 1980) sociocultural theory of language learning and the Technology Acceptance Model (Davis, 1989) are used to interpret the results, which are then discussed in relation to the three themes that arose from the thematic analysis: positive perceptions, challenges encountered, and pedagogical recommendations. The findings are analyzed in the context of broader research on mobile-assisted language learning (MALL), AI-assisted pronunciation acquisition, and the unique sociolinguistic context of Sasaknese EFL learners in Eastern Indonesia .

P*ositive Attitudes and AI Feedback's Contribution to Pronunciation Development*

The conclusion that all twelve participants had favorable opinions of ELSA Speak overall is in line with an increasing amount of studies showing that EFL students had favorable opinions of AI-based pronunciation aids in a variety of Asian situations (Kholis, 2021). The application's real-time, phoneme-level corrective feedback is valued by users. This is consistent with known criteria for effective pronunciation instruction, which emphasize the importance of timely, targeted, and tailored feedback in fostering phonological awareness and articulatory adjustment (Liakin et al., 2015). According to the TAM framework, learners' desire to adopt and continue using the technology is significantly influenced by how beneficial they believe ELSA Speak's feedback mechanism to be (Davis, 1989; Venkatesh et al., 2003). It is confirmed that L1 phonological transfer is a persistent source of pronunciation difficulty for this learner population because the phonemes that participants reported improving, especially /θ/, /ð/, and /v/, are exactly those that are missing from the Sasak and Indonesian phonological inventories (Insani et al., 2026). Thus, ELSA Speak's ability to identify and consistently target these particular phonemes fills a real and established pedagogical need that traditional classroom education, limited by time and class size,

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

is rarely able to achieve with equal precision.

The motivational benefits associated with the application's gamification features daily streaks, achievement badges, and cumulative fluency scores further corroborate the relevance of self-determination theory in understanding technology-enhanced language learning engagement (Dörnyei, 2005; Reinhardt, 2019). ELSA Speak seems to meet the psychological needs for competence and autonomy that underlie intrinsic motivation by giving students observable indicators of incremental success. For Sasaknese EFL learners, who work in a complicated multilingual environment where English is a distant third language and there are few opportunities for meaningful pronunciation practice outside of formal instruction, this is especially important (Insani et al., 2026). Similar to previous research showing that technology-mediated speaking tasks reduce foreign language anxiety by removing the social evaluative pressures inherent in face-to-face interaction, the application's private, low-stakes digital environment provides affective comfort (Ducate & Lomicka, 2009; Shadiev et al., 2026). When taken as a whole, these affordances present ELSA Speak as a tool that addresses both the cognitive and affective aspects of pronunciation learning a dual purpose that is rarely accomplished by traditional pedagogical approaches alone.

Challenges: Accent Bias, Feedback Limitations, and Infrastructure Constraints

Participants' concerns highlight crucial issues regarding the suitability of existing AI-based pronunciation aids for linguistically marginalized learner communities. Sasaknese learners have accent bias in native-speaker variations of English, where phonetically understandable sentences are frequently misclassified as incorrect. This is a well-known disadvantage of automatic speech recognition (ASR) systems, whose training data are mainly derived from mainstream Sasaknese learners.AI systems that consistently penalize non-mainstream accents run the risk of reinforcing deficit orientations toward non-native pronunciation and undermining the communicative confidence of learners whose English is shaped by rich and legitimate multilingual repertoires (Godwin-Jones, 2025; Kramsch, 2018). This institutional prejudice has serious consequences for schooling. This restriction is not just a technical annoyance for Sasaknese EFL learners whose English pronunciation is naturally influenced by the phonological systems of both Sasak and Indonesian; rather, it is a significant equity issue that calls for consideration from both application developers and language education policymakers.

Findings from more general research on computer-assisted pronunciation training (CAPT), which has long struggled with the conflict between diagnostic accuracy and instructional

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

scaffolding, are consistent with the inadequacy of ELSA Speak's remedial feedback (Al-Yafaei et al., 2026). A necessary but insufficient prerequisite for improving pronunciation is realizing that a phoneme is mispronounced; learners also need explicit metalinguistic guidance on the articulatory parameters place and manner of articulation, voicing, and airstream mechanism that distinguish the error from the target form. For Sasaknese speakers attempting to learn English phonemes without articulatory counterparts in their L1, the absence of this kind of help in ELSA Speak is particularly crucial. However, according to the Vygotskian perspective, the application's feedback no matter how accurate cannot replace the more experienced person in the Zone of Proximal Development; it can identify the gap but cannot help the learner close it (Guerrero Nieto, 2007). In particular, access to the tool is restricted by infrastructural obstacles such as uneven internet connectivity in Lombok's rural and semi-urban areas, which could worsen already-existing educational disparities between urban and rural students .

Pedagogical Implications: Toward a Culturally Responsive Blended Learning Model

Participants' suggestion of a blended learning integration model for ELSA Speak is consistent with a growing body of research in the MALL literature that suggests AI pronunciation tools have the most pedagogical impact when used as parts of a teacher-mediated instructional ecology rather than as independent, stand-alone resources. (Godwin-Jones, 2025; Guerrero Nieto, 2007) sociocultural framework, which views learning as essentially social and mediated, and the TAM's emphasis on perceived usefulness as a function of how well a technology is integrated into established instructional routines are both strongly aligned with the model that participants envisioned, in which ELSA Speak homework tasks are paired with in-class debriefing sessions. In this model, the teacher takes on the mediating role that the application itself cannot fulfill: providing metalinguistic explanation, contextualizing pronunciation norms within the students' multilingual reality. The TAM's focus on perceived usefulness as a function of how well a technology is integrated into established instructional routines and (Guerrero Nieto, 2007)sociocultural framework, which views learning as essentially social and mediated, are both highly compatible with this model (Davis, 1989).

Peer collaborative practice may be an especially fruitful addition to ELSA Speak's individualized feedback model, given the collectivist orientation of Sasaknese learners shaped by Islamic educational traditions, communal learning norms, and the sociolinguistic dynamics of a minority ethnolinguistic community (Giollagáin et al., 2025). Participants' suggestions for organized peer review sessions, where students compare application feedback and work together

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

to solve pronunciation problems, show an intuitive alignment with sociocultural learning principles and imply that the social aspects of Sasaknese ed ucational culture can be effectively mobilized to expand and enhance the affordances of AI-based pronunciation tools . These results add to the growing body of research on the necessity of culturally sensitive methods for integrating technology into language instruction, especially in Southeast Asian contexts where ethnolinguistic minorities are underrepresented.

Overall, the results of this study indicate that ELSA Speak has significant pedagogical potential for Sasaknese EFL learners, given that its implementation is coupled with thoughtful instructional scaffolding, fair infrastructure support, and critical awareness of the accent biases present in existing AI speech recognition systems. Affective, social, and metalinguistic aspects of the learner experience that are not captured by the quantitative evaluation of pronunciation scores alone have been revealed by the study's qualitative methodology (Creswell & Creswell, 2017; Yin, 2017). The experiences of EFL teachers tasked with incorporating AI pronunciation tools into their pedagogy, the potential of co-designed, culturally contextualized AI feedback systems that incorporate non-mainstream accent data into their training corpora, and longitudinal trajectories of pronunciation development among Sasaknese EFL learners using ELSA Speak could all be fruitfully investigated in future research (Godwin-Jones, 2025).

CONCLUSION

This study investigated how Sasaknese EFL students felt about ELSA Speak, an artificial intelligence tool for practicing English pronunciation. Based on a thematic analysis of twelve participants' reflective journals, classroom observations, and interview data, the results showed that students had largely favorable opinions of the application, especially appreciating its gamified motivational features, real-time phoneme-level feedback, and the low-stress environment it provided for independent pronunciation practice (Bashori et al., 2024). Simultaneously, participants noted significant difficulties, such as accent bias in the speech recognition engine of the application, inadequate explanatory feedback for articulatory remediation, and infrastructure obstacles pertaining to internet connectivity in Lombok's rural and semi-urban areas). These challenges are closely tied to the unique sociolinguistic position of Sasaknese learners, whose English pronunciation is shaped by phonological interference from both Sasak and Indonesian (Utami et al., 2026)

When combined, the three themes found in this study positive perceptions, difficulties encountered, and pedagogical recommendations indicate that ELSA Speak is best viewed as a

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

supplementary tool whose pedagogical value is maximized when integrated within a teacher-mediated blended learning framework rather than as a stand-alone substitute for classroom pronunciation instruction (Davis, 1989; Godwin-Jones, 2025; Vygotsky, 1980) sociocultural theory, which emphasizes that technology-mediated learning reaches its maximum potential when scaffolded through professional guidance within the learner's Zone of Proximal Development, is in line with this result. Thus, the study confirms the usefulness of qualitative research in capturing the affective, sociocultural, and metalinguistic aspects of learners' experiences with AI-based pronunciation tools that are not sufficiently represented by quantitative pronunciation scores alone (Creswell & Creswell, 2017; Yin, 2017)

THEORETICAL AND PEDAGOGICAL IMPLICATIONS

The findings of this study carry several theoretical and pedagogical implications. The study shows that perceived utility and perceived ease of use continue to be significant predictors of technology adoption even in complex multilingual contexts shaped by L1 phonological interference, theoretically extending the application of the Technology Acceptance Model to an underrepresented ethnolinguistic minority population. By showing that learner adoption of AI-based tools is influenced by both perceived utility and the availability of social and instructional mediation, the study further highlights the explanatory significance of combining sociocultural theory with TAM (Davis, 1989; Vygotsky, 1980). Future studies looking at technology uptake among various ethnolinguistic minority EFL communities in Indonesia and the larger Southeast Asian area may benefit from the application of this integrated theoretical lens.

According to the findings, English instructors who work with Sasaknese and similarly situated EFL students should think about implementing a blended learning approach wherein AI-based pronunciation programs, like ELSA Speak, are given as structured homework, followed by in-class debriefing sessions that offer the metalinguistic explanation and articulatory scaffolding that the program itself cannot provide. In order to prepare teachers to assist students in interpreting and contextualizing AI-generated feedback rather than blindly accepting it, curriculum developers and teacher educators should also think about including training on the critical evaluation of AI language tools (Shadiev et al., 2026). At the institutional level, the study highlights the importance of investing in digital infrastructure in rural and semi-urban areas of Lombok to ensure equitable access to AI-based learning tools across the student population (Muliyah et al., 2021).

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

SUGGESTION FOR FUTURE RESEARCH

Based on the findings and limitations of this study, A number of avenues for further investigation are proposed. In order to get insight into the developmental trajectory of AI-assisted pronunciation learning, future research might first adopt a longitudinal approach to investigate how Sasaknese EFL learners' pronunciation and views of ELSA Speak change over a prolonged time of continuous use. Second, a more comprehensive knowledge of blended learning implementation in this context would be provided by research involving EFL instructors' perspectives on incorporating AI pronunciation tools into official curricula, which would complement the learner-centered focus of the current study. Third, in order to directly address the concerns stated by survey participants regarding accent bias, future research might examine the viability of creating or modifying AI speech recognition systems trained on Indonesian-accented and ethnolinguistic minority voice data. Finally, comparative studies examining the perceptions of AI-based pronunciation tools across other Indonesian ethnolinguistic minority groups would help to further generalise and contextualise the findings of this study within the broader landscape of applied linguistics research in underrepresented populations (Kholis, 2021).

REFERENCES

Akhmad, N. W., & Munawir, A. (2022). Improving the Students' Pronunciation Ability by Using Elsa Speak App. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, *10*(1), 846–857. https://doi.org/10.24256/ideas.v10i1.2868

Al-Yafaei, Y. M., Jomaa, N. J., & Attamimi, R. A. (2026). Human-Centered Approaches to AI-Enhanced English Language Learning and Teaching. *IGI Global*. https://www.researchgate.net/profile/Nayef-Jomaa/publication/399083879_Human-Centered_Approaches_to_AI-_Enhanced_English_Language_Learning_and_Teaching/links/694e579e06a9ab54f84af2de/Human-Centered-Approaches-to-AI-Enhanced-English-Language-Learning-and-Teaching.pdf

Bashori, M., Van Hout, R., Strik, H., & Cucchiarini, C. (2024). *I Can Speak*: Improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching*, *18*(5), 443–461. https://doi.org/10.1080/17501229.2024.2315101

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

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Celce-Murcia, M. (1996). Teaching Pronunciation: A series of booknotes. *Issues in Applied Linguistics*, *7*(1). https://doi.org/10.5070/L471005226

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications. https://www.google.com/books?hl=id&lr=&id=335ZDwAAQBAJ&oi=fnd&pg=PT16&dq=Research+design:+Qualitative,+quantitative,+and+mixed+methods+approaches+(5th+ed.).+SAGE+Publications.&ots=YEAQKRtqoO&sig=QG3fxt8Le_atwjri7MqJsIFATfs

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Dörnyei, Z. (2014). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge. https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9781410613349&type=googlepdf

Ducate, L., & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation? *Language Learning & Technology*, *13*(3), 66.

ELSA, C. (n.d.). *to help users learn English pronunciation (ELSA Corp, 2023)—Penelusuran Google*. Retrieved June 28, 2026, from https://www.google.com/search?q=to+help+users+learn+English+pronunciation+%28ELSA+Corp%2C+2023%29&rlz=1C1JJTC_idID1141ID1141&oq=to+help+users+learn+English+pronunciation+%28ELSA+Corp%2C+2023%29+&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIHCAEQABjvBTIHCAIQABjvBTIHCAMQABjvBTIHCAQQABjvBdIBCTM1NjFqMGoxNagCCLACAfEFGA7EUZx_oXQ&sourceid=chrome&ie=UTF-8

Giollagáin, C. Ó., Bourgeois, D., Curnáin, B. Ó., Caimbeul, I., & Cameron, G. (2025). Language Dynamics in Society (LanDS): The LanDS Analytical Framework for Majority and Minority-Language Ethnolinguistic Vitality. *Treatises and Documents, Journal of Ethnic Studies / Razprave in Gradivo, Revija Za Narodnostna Vprašanja*, *94*(1), 25–75. https://doi.org/10.2478/tdjes-2025-0002

Godwin-Jones, R. (2017). *Scaling up and zooming in: Big data and personalization in language learning*. https://scholarspace.manoa.hawaii.edu/bitstreams/0b0d10ed-fb75-455e-94c0-2bc54ace1f41/download

Godwin-Jones, R. (2025). *Ecological semiotics: Multimodality, multilingualism, and situated language learning in the AI era*. https://scholarspace.manoa.hawaii.edu/items/4430d479-3e40-4349-90af-5cf1493c0947

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

Guerrero Nieto, C. H. (2007). Applications of Vygotskyan concept of mediation in SLA. *Colombian Applied Linguistics Journal*, (9), 213–228.

Insani, I. A., Azizah, A. N., & Lubis, Y. (2026). An ANALYSIS OF VOWEL AND CONSONANT ERRORS IN ENGLISH SPEECH FOR EFL STUFENTS IN INDONESIA. *JALC: JOURNAL OF APPLIED LINGUISTIC AND STUDIES OF CULTURAL*, *4*(1), 12–19.

Khakurel, U. (2025). *Towards Responsible AI: Methods for Fairness and Bias Mitigation Across Structured and Unstructured Datasets* [PhD Thesis, Howard University]. https://search.proquest.com/openview/7cfd9a0c35012d1549aa3c6cbd8f6902/1?pq-origsite=gscholar&cbl=18750&diss=y

Kholis, A. (2021). 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

Kramsch, C. (2018). Is there still a place for culture in a multilingual FL education? *Language Education and Multilingualism-The Langscape Journal*, *1*, 16–33.

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

Lincoln, Y. S., & Guba, E. G. (1988). *Criteria for Assessing Naturalistic Inquiries as Reports.* https://eric.ed.gov/?id=ed297007

Morgan, D. L. (1997). *Focus groups as qualitative research* (Vol. 16). Sage. https://www.google.com/books?hl=id&lr=&id=iBJZusd1GocC&oi=fnd&pg=PP7&dq=Morgan,+D.+L.+(1997).+Focus+Groups+as+Qualitative+Research+(2nd+ed.).+Thousand+Oaks,+CA:+Sage+Publications.&ots=mBRgfMhAX2&sig=Wcn2mCya1A0aRNdfE6NKt7vgm9M

Shadiev, R., Kaliyeva, A., Kairatova, A., Yerbulatova, Z., Ryskulbek, A., & Beisembayeva, Z. (2026). Exploring the role of conversational AI in developing communicative skills: A systematic review. *Digital Applied Linguistics*, *4*, 103058–103058.

Sholekhah, M. F., & Fakhrurriana, R. (2023). The Use of ELSA Speak as a Mobile-Assisted Language Learning (MALL) towards EFL Students Pronunciation. *JELITA: Journal of Education, Language Innovation, and Applied Linguistics*, *2*(2), 93–100. https://doi.org/10.37058/jelita.v2i2.7596

Stake, R. E. (1995). *The art of case study research*. sage.

Utami, E., Wahyuni, S. N., Sulistiyono, M., Raharjo, S., Hartanto, A. D., Rohman, A. N., Triwijoyo, B. K., Miswaty, T. C., & Patty, E. N. S. (2026). A Multimodal Transformer-LSTM Framework

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License

for Cross-Lingual Lexical Alignment of Indonesian Regional Languages. *Engineering, Technology & Applied Science Research*, *16*(2), 34283–34292.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view1. *MIS Quarterly*, *27*(3), 425–478.

Vygotsky, L. S. (1980). *Mind in Society: Development of Higher Psychological Processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wood, K. S. (2022). *Increasing learner engagement in, and performance of, technical threshold concepts in accounting and finance education through a gamified learning experience* [PhD Thesis, University of Tasmania]. https://figshare.utas.edu.au/articles/thesis/Increasing_learner_engagement_in_and_performance_of_technical_threshold_concepts_in_accounting_and_finance_education_through_a_gamified_learning_experience/23254121/1

Yin, R. K. (2017). *Case study research and applications: Design and methods*. Sage publications.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License