Upper Primary Students' Perceptions Towards the Use of Elsa Speak in Improving English Pronunciation

***Grace Kho Hui Xuan, Melor Md Yunus, Hanita Hanim Ismail**

Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

***Corresponding Author**

DOI: https://dx.doi.org/10.47772/IJRISS.2025.91100176

Received: 19 November 2025; Accepted: 26 November 2025; Published: 04 December 2025

ABSTRACT

English pronunciation is a significant barrier for many Malaysian primary students, hindering their oral participation and causing persistent low confidence and speaking anxiety. AI-powered tools, such as ELSA Speak, offer personalised, real-time feedback that could address these limitations; however, relevant research among young Malaysian learners is scarce. This mixed-methods study investigated the effectiveness of ELSA Speak among 62 upper primary students (Year 5 and Year 6) in an SJKC school over four weeks. A questionnaire measured students' perceptions, confidence, motivation and speaking anxiety, while interviews with eight pupils provided deeper insights. Quantitative analysis indicated generally positive perceptions and significant improvements in students' reported pronunciation clarity, higher motivation, increased confidence and reduced anxiety. Qualitative findings supported these outcomes, revealing that gamified features, colour-coded feedback and private practice encouraged greater engagement and readiness to speak. Challenges identified included speech-recognition inaccuracies, occasional technical issues and difficulty understanding linguistic terminology. Overall, ELSA Speak demonstrates strong potential as a supplementary pronunciation tool, though improvements in accuracy and child-friendly feedback are needed.

Keywords: Artificial Intelligence (AI), ELSA Speak, pronunciation training, Primary ESL Learners, perceptions

INTRODUCTION

Pronunciation is widely recognised as a critical component of English language proficiency, as it shapes learners' intelligibility, communicative effectiveness and overall self-confidence (Derwing & Munro, 2015; Isaacs & Trofimovich, 2017). In Malaysia, the importance of spoken English has intensified under the Common European Framework of Reference (CEFR)-aligned curriculum, which foregrounds learners' ability to communicate confidently in real-world situations (Ministry of Education Malaysia, 2021). Despite these reforms, many primary school learners continue to struggle with accurate pronunciation, resulting in hesitation during oral tasks and restricted participation. These difficulties are commonly attributed to limited opportunities for individualised practice, lack of corrective feedback, and the fear of making mistakes in front of peers (Ch'ng, Chan, & Saufi, 2025; Lau et al., 2025).

The emergence of Artificial Intelligence (AI)-enhanced language learning applications has created new possibilities for addressing these longstanding issues. Among these tools, ELSA Speak has gained traction as a popular mobile application that provides automated, instant feedback on pronunciation, using colour-coded indicators and articulatory explanations to identify mispronounced sounds. By allowing learners to practise independently at their own pace, AI pronunciation tools serve as valuable supplements to classroom instruction, providing targeted and self-paced practice essential for improving pronunciation skills. Studies conducted between 2024 and 2025 demonstrate that ELSA Speak significantly improves pronunciation accuracy, learner confidence, and satisfaction across diverse contexts (Adawiah et al., 2024; Pham & Pham, 2025; Aulia & Santosa, 2025). These affordances are particularly relevant in the Malaysian primary context, where large class sizes and limited instructional time pose challenges for teachers to provide targeted pronunciation support (Ag-Ahmad et al., 2025; Mohammed et al., 2021).

Although studies have documented the benefits of AI-driven pronunciation tools among adult and tertiary learners internationally, relatively little is known about their impact on younger ESL learners in Malaysia. Primary pupils differ from older learners in terms of cognitive development, motivation and affective needs, and their experiences with AI-based tools warrant dedicated investigation. Gaining insight into how AI tools influence their pronunciation development, confidence, motivation and speaking anxiety is therefore crucial for determining their suitability for integration into primary ESL classrooms.

To address this gap, the present study investigated the use of ELSA Speak among upper primary learners in an SJKC school across four weeks. By employing a mixed-method design involving a student questionnaire and semi-structured interviews, this research sought to provide a comprehensive understanding of pupils' perceptions, emotional responses and learning challenges when engaging with AI-driven pronunciation practice. The findings contribute to the growing body of knowledge on AI-assisted language learning and offer insights into how pronunciation training tools can be meaningfully integrated within the Malaysian primary English language curriculum.

Research Problem and Its Importance

Despite sustained efforts to strengthen English language education in Malaysia, many primary school pupils continue to struggle with pronunciation accuracy and oral fluency. These challenges are often reflected in classroom observations, where pupils demonstrate hesitation when reading aloud, limited participation during speaking tasks, and persistent reliance on teacher modelling. Such difficulties not only impede learners' oral intelligibility but also contribute to low self-confidence and heightened speaking anxiety—affective barriers that significantly influence learners' willingness to communicate. Current research indicates that Malaysian ESL learners often struggle with speaking anxiety, inhibition and fear of negative evaluation, all of which adversely affect their oral performance (Hashim, 2024; Aziz & Kashinathan, 2021).

A key issue is the limited access to personalised pronunciation practice in the primary school setting. Teachers frequently struggle to provide personalised feedback due to time constraints, large class sizes and competing curricular requirements. Prior studies highlight that large class sizes and curriculum pressures restrict teachers' ability to provide targeted oral feedback, resulting in pupils receiving insufficient guidance on features such as phoneme production, stress, and rhythm (Hazita, 2016; Mohammed, Raof, & Yusof, 2021). As a result, learners have few opportunities to refine their pronunciation. Given these constraints, there is an increasing need to explore alternative instructional tools that can complement classroom practices.

AI-driven applications, such as ELSA Speak, offer promising solutions by providing instant, precise, and repeated feedback—key components for effective pronunciation improvement. However, despite the growing use of such applications globally, there is insufficient evidence on how Malaysian primary learners perceive these tools, whether they positively influence affective factors such as motivation and confidence, and what challenges may hinder their effective use. This gap is especially critical given that many speech recognition technologies are primarily designed for adult voices, which may pose additional challenges for younger users.

Hence, it is essential to investigate how upper primary students engage with AI pronunciation tools, how they perceive their usefulness, and how these tools influence their emotional readiness to speak English. Insights into these areas can inform pedagogical decisions, technology adoption and future curriculum planning within the Malaysian ESL context. Addressing this research problem also aligns with broader national efforts to enhance English language proficiency through technology-enhanced learning.

Research Objectives

The objectives of this research are:

1. To examine students' perceptions of ELSA Speak as a pronunciation learning tool
2. To determine whether ELSA Speak improves students' confidence, motivation and reduces anxiety in speaking English

3. To find out the challenges and limitations experienced by the students when using ELSA Speak

LITERATURE REVIEW

Advancements and Impacts of AI on Pronunciation Training

Recent developments in AI have progressively reshaped pronunciation pedagogy by offering learners personalised, real-time feedback that was once difficult to implement in traditional classroom settings. AI systems equipped with automatic speech recognition (ASR) technology can analyse learners' spoken output in detail, identify errors and provide corrective guidance almost instantly, which is particularly beneficial for younger learners who thrive on timely correction during oral practice (Prakash & Kausalya, 2025; Sungkur & Shibdeen, 2025).

AI-based pronunciation training has been consistently shown to strengthen learners' segmental accuracy, fluency and intelligibility. Studies indicate that ASR-supported tools enable learners to repeatedly practise problematic phonemes in an adaptive, learner-centred environment (Shak, Lee, & Loo, 2025). Moreover, research in ESL contexts reveals that AI pronunciation tools enhance learners' opportunities for speaking practice, particularly in educational settings where teachers may lack sufficient time to provide one-on-one feedback (Thomas, 2025; Vincent, Yunus, & Said, 2025). As such, AI functions as a complementary support system, enabling learners to rehearse pronunciation independently and consistently.

Another major contribution of AI pronunciation tools lies in learner engagement. Many AI applications incorporate gamified features, such as point systems, progress trackers and badges, which have been found to significantly boost learner motivation and sustain interest in pronunciation tasks (Caiza, Villafuerte, & Guanuche, 2025). Moreover, the low-stress environment created by AI tools for oral practice reduces learners' anxiety and encourages a greater willingness to speak (Li, Mohamad, & You, 2025). Despite these advantages, challenges remain, including accent misrecognition, inconsistent scoring and variations in feedback accuracy, particularly when tools are used by younger learners with developing speech patterns (Lai & Chen, 2022). These limitations underscore the need for further investigation into how primary pupils respond to AI pronunciation applications in real classroom settings.

ELSA Speak as a Pronunciation Training Tool

ELSA Speak (English Language Speech Assistant) has emerged as a prominent AI tool designed specifically for pronunciation training. The application employs advanced ASR and deep learning algorithms to analyse learners' pronunciation at phoneme, word and sentence levels, while also assessing suprasegmental features such as stress and intonation. Dhivya et al. (2023) note that the app's colour-coded scoring system and step-by-step corrective feedback allow learners to pinpoint articulatory issues more effectively than traditional drills.

ELSA Speak also supports learner autonomy. With intuitive features such as visual diagrams of mouth positions, automatically generated lesson pathways and personalised progress tracking, the app encourages learners to take charge of their own pronunciation practice (Dhivya, Hariharasudan, Ragmoun, & Alfalih, 2023). Learners can monitor their progress, set goals and repeat exercises freely, which strengthens ownership of learning and contributes to improved fluency. Research further underscores ELSA Speak's alignment with Education 4.0 principles, highlighting its emphasis on learner-centred, technology-enhanced, and self-paced practice environments (Dhivya et al., 2023).

However, most existing studies on ELSA Speak focus on adults or tertiary-level learners. For example, studies with university students report notable gains in oral proficiency after sustained use of ELSA Speak but also highlight mixed perceptions regarding feedback accuracy and phoneme scoring (Pham & Pham, 2025). Far fewer studies have examined how younger learners interpret and benefit from the tool's feedback systems. Primary-aged pupils often require simpler instructions, clearer scaffolding and teacher mediation to comprehend automated feedback (Kamaruddin, 2023; Zulkarnain, & Yunus, 2023). This gap in the literature suggests a need to evaluate the suitability and effectiveness of ELSA Speak for younger ESL learners, particularly in the Malaysian primary school context, where exposure to English varies widely.

Students' Perceptions of AI Pronunciation Tools in English language learning

Learners' perceptions play a central role in determining the success of AI-assisted pronunciation learning. When learners feel supported, motivated or empowered by a tool, they tend to use it more actively; however, negative perceptions may reduce engagement and limit learning gains (Amin, 2024). Research consistently indicates that AI pronunciation tools can significantly enhance learners' confidence, especially because they allow learners to practise privately without the fear of being judged or criticised by peers (Sari & Kurniawan, 2025). Consistent practice supported by clear, non-judgmental feedback contributes to improved self-belief in oral production tasks.

Motivation is another key factor influenced by AI-based tools. Studies show that AI applications with gamified features, such as levels, achievements and interactive visual feedback, contribute to increased enjoyment and commitment to practice (Yaccob et al., 2022; Liu, 2024). Learners are often encouraged by the incremental progress and rewards provided by tools like ELSA Speak, which reinforce persistence and sustained effort. This motivational impact is especially vital for primary school pupils, who often thrive in visually rich, reward-based learning environments (Hung et al., 2018).

AI pronunciation tools have also been shown to reduce speaking anxiety. Speaking anxiety is a common obstacle in ESL learning that often leads to avoidance of oral tasks. The private, self-paced nature of AI tools helps learners practise without pressure, reducing the fear of being criticised for mispronunciations (Nadeem, Iftikhar, & Rasheed, 2025). By offering a rehearsal platform before speaking publicly, these tools help learners feel more prepared and less anxious during classroom speaking tasks. This reduction in anxiety is linked to increased willingness to communicate, which is a key factor in successful oral language development.

However, despite these benefits, researchers also caution that technological challenges, such as inconsistent speech recognition, misunderstanding of children's voices and unclear scoring, can negatively affect learners' confidence and motivation (Ngo, Chen, & Lai, 2024). For younger learners, misinterpreting feedback may lead to confusion or frustration, particularly when scores appear inaccurate or seem difficult to improve. These challenges suggest that teacher support and guided practice remain essential when integrating tools like ELSA Speak into primary ESL classrooms.

METHODOLOGY

Research Design

This study adopted an explanatory sequential mixed-methods design, in which quantitative data were gathered first, followed by qualitative data. The initial quantitative phase involved administering a questionnaire to capture overall trends in pupils' perceptions of ELSA Speak as a pronunciation training tool. The subsequent qualitative phase consisted of semi-structured interviews, which were conducted to elaborate, clarify, and provide deeper insight into the patterns identified in the survey responses. This design aligns with Creswell and Plano Clark's (2017) recommendation for studies seeking to explore the underlying explanations behind quantitative patterns. Combining both statistical and narrative data enhanced the depth, validity and interpretive strength of the findings.

Participants and Sampling

The participants in this study were 62 upper primary pupils from a Chinese national-type primary school (SJKC) in Selangor, Malaysia. The sample comprised 31 Year 5 pupils and 31 Year 6 pupils, aged 11 and 12 years old. All pupils were enrolled in English as a Second Language (ESL) classes and had learned English throughout their primary schooling.

A convenience sampling method was used due to the school's accessibility and willingness to participate. This approach was appropriate given the structured school setting, the age of the pupils and the practical constraints involved in conducting classroom-based research. All participants possessed basic digital literacy and had access to a mobile device either personally or through shared family usage. Participation was voluntary, and written informed consent was obtained from parents or guardians. To maintain confidentiality, pseudonyms were

assigned, and identifiable information was removed from all datasets.

Instruments

Two instruments were used in this study: a questionnaire and a semi-structured interview protocol. These instruments were selected to capture both the quantitative patterns and qualitative depth necessary to understand upper primary pupils' experiences with ELSA Speak. The questionnaire was the main quantitative instrument and consisted of 25 Likert-scale items designed to measure five constructs related to pupils' use of the application, namely perceived usefulness, pronunciation confidence, learning motivation, speaking anxiety and challenges encountered during practice. All items were rated on a five-point scale ranging from Strongly Disagree to Strongly Agree. The items were adapted from established instruments used in previous studies on AI-assisted pronunciation learning, technology adoption and young learners' speaking anxiety, but were modified to suit the linguistic and developmental level of primary school pupils. To ensure clarity and content validity, the questionnaire was reviewed by two TESL lecturers and an experienced ESL teacher who evaluated the suitability, readability and alignment of each item with the target constructs. Their suggestions led to several refinements, particularly in simplifying technical terms and improving the wording for young learners. In addition to expert review, the questionnaire underwent a brief pilot check with a small group of pupils outside the main sample.

Feedback from the pilot confirmed that the instructions, item phrasing and response options were understandable for upper primary pupils. The reliability of the questionnaire was subsequently tested using Cronbach's alpha, which produced a coefficient of 0.87, indicating strong internal consistency and confirming that the items were appropriate for measuring the intended constructs.

The second instrument used in this study was a semi-structured interview protocol. The interview component was designed to complement the questionnaire by exploring pupils' perceptions, motivation, confidence and speaking anxiety in greater depth. Eight pupils were purposively selected based on their questionnaire response patterns to ensure variation in perspectives. The interview guide consisted of open-ended questions framed in simple, age-appropriate language to help pupils express their thoughts comfortably. The semi-structured format allowed the researcher to ask follow-up questions when necessary, enabling a more nuanced understanding of pupils' experiences with ELSA Speak. Interviews were conducted in a quiet classroom environment, and pupils were given the option to respond in English, Mandarin or Bahasa Melayu to ensure full comprehension and comfort. All interviews were audio-recorded with parental consent, transcribed verbatim and supplemented with researcher field notes describing pupils' non-verbal expressions and reactions. These qualitative data later played an important role in enriching the interpretation of quantitative findings by highlighting pupils' emotional responses, preferred features of the application and the specific challenges they encountered during the intervention period.

Data Collection and Data Analysis

Data collection took place over two sequential phases in accordance with the explanatory sequential mixed-methods design. During the initial quantitative phase, pupils completed the questionnaire in their respective classrooms under the supervision of the researcher and the English teacher. Clear verbal instructions were provided, and where necessary, additional explanations were offered in Mandarin or Bahasa Melayu to ensure comprehension without influencing responses. Pupils completed the questionnaire within approximately twenty minutes, and all forms were collected immediately to maintain data accuracy and avoid missing responses. Following the completion of the quantitative phase, the qualitative phase commenced with semi-structured interviews involving eight purposively selected pupils. These pupils represented a diverse range of questionnaire response patterns. Interviews were held in a quiet classroom space, audio-recorded with parental consent and accompanied by researcher field notes that documented pupils' non-verbal cues, hesitations, enthusiasm and moments of uncertainty. Each interview lasted between ten and fifteen minutes and allowed pupils to express their experiences and feelings regarding their use of ELSA Speak in greater depth.

Following data collection, the questionnaire responses were analysed using the Statistical Package for the Social Sciences (SPSS) Version 27. Responses were coded numerically on a five-point Likert scale, enabling the computation of descriptive statistics such as frequencies, percentages, means and standard deviations. These

descriptive measures provided a clear overview of pupils' perceptions across the constructs of usefulness, confidence, motivation, anxiety reduction and challenges. Reliability analysis using Cronbach's alpha was conducted within SPSS as part of the data screening process, confirming the strong internal consistency of the questionnaire. Before analysis, the dataset was examined for missing values and inconsistencies; however, all questionnaires were completed fully and no invalid entries were detected.

The qualitative interview data were analysed using Braun and Clarke's (2006) six-phase framework for thematic analysis. This process involved familiarising oneself with the transcripts, generating initial codes, identifying and reviewing potential themes, refining theme definitions and integrating the findings into a coherent narrative. Themes were generated inductively to allow pupils' voices and personal experiences to emerge naturally from the data. This analytic process provided richer contextual insights that explained, supported or expanded upon the quantitative results. The integration of both datasets ultimately strengthened the interpretation of findings by offering a more comprehensive understanding of pupils' perceptions and experiences with ELSA Speak.

FINDINGS

This section presents the findings of the study, which investigated the perceptions of 62 upper primary pupils (31 Year 5 and 31 Year 6) from a Malaysian SJKC towards the use of ELSA Speak for English pronunciation learning. The results are organised into three main sections in accordance with the structure of the questionnaire: (i) pupils' perceptions of ELSA Speak as a pronunciation training tool, (ii) the influence of ELSA Speak on pupils' confidence, motivation, and speaking anxiety, and (iii) challenges and limitations encountered during the use of the application. Qualitative findings from semi-structured interviews are also presented to complement the quantitative data. The chapter concludes with a discussion integrating both sets of findings.

Survey

Pupils' Perceptions of ELSA Speak

Table 1 presents pupils' responses to Items 1 to 5, which examined their perceptions of ELSA Speak in terms of usefulness, clarity of feedback, phoneme guidance, and overall contribution to pronunciation improvement.

Table 1 Pupils' Perceptions of Using ELSA Speak

No.	Item	SD	D	N	A	SA	Mean	SD
1	ELSA Speak helps me pronounce words correctly.	0	2	6	20	34	4.38	0.72
2	The feedback helps me improve my pronunciation.	0	2	7	22	31	4.32	0.74
3	I can learn correct mouth and tongue positions.	0	3	5	23	31	4.32	0.78
4	ELSA makes my speech clearer.	0	2	6	21	33	4.37	0.73
5	ELSA Speak is useful for learning pronunciation.	0	2	6	21	33	4.37	0.73

The results indicate overwhelmingly positive perceptions of ELSA Speak. More than 87% of pupils responded with Agree or Strongly Agree across all perception items. The highest endorsement was for Item 1 (M = 4.38), suggesting that pupils strongly believed the app helped them pronounce words correctly. Similarly, high means for Items 2 to 5 (M = 4.32 to 4.37) reinforced the view that the app's colour-coded feedback and articulatory guidance were both accessible and useful. This aligns with ASR-supported pronunciation literature emphasising the effectiveness of real-time phoneme-level feedback.

Influence on Confidence, Motivation, and Speaking Anxiety

Table 2 summarises Items 6 to 20, grouped under the constructs of pronunciation confidence, motivation, and anxiety reduction.

Table 2 Influence of ELSA Speak on Confidence, Motivation, and Anxiety

Construct	Mean	SD
Pronunciation Confidence (Items 6–10)	4.20	0.79
Learning Motivation (Items 11–15)	4.44	0.69
Speaking Anxiety Reduction (Items 16–20)	4.10	0.82

No.	Item	SD	D	N	A	SA	Mean	SD
6	I feel more confident speaking English after using ELSA.	0	2	7	22	31	4.32	0.75
7	I am less afraid of making mistakes.	0	3	6	23	30	4.29	0.79
8	I am more willing to speak English in class.	0	3	8	21	30	4.26	0.81
9	I feel proud when I get a high score.	0	2	6	19	35	4.40	0.74
10	I can pronounce difficult words more confidently.	0	2	8	21	31	4.29	0.78
11	I enjoy using ELSA Speak.	0	2	5	20	35	4.43	0.72
12	I want to use ELSA even without teacher instruction.	0	3	6	20	33	4.35	0.78
13	I like the stars, coins, and scores.	0	2	6	18	36	4.45	0.71
14	ELSA makes me want to improve my pronunciation.	0	2	5	19	36	4.45	0.70
15	ELSA Speak is fun and interesting.	0	3	5	18	36	4.42	0.74
16	I feel less nervous speaking English.	0	4	7	22	29	4.23	0.85
17	I feel comfortable practising alone with ELSA.	0	3	8	23	28	4.23	0.82
18	I am less scared of being laughed at.	0	5	7	22	28	4.18	0.89
19	I prefer practising before speaking in class.	0	4	9	23	26	4.13	0.87
20	ELSA helps me overcome fear of speaking.	0	4	8	25	25	4.13	0.86

The construct with the highest mean score was motivation (M = 4.44, SD = 0.69). Items related to gamification elements, such as enjoyment (Item 11), willingness to practise independently (Item 12), and appreciation of stars, coins and scores (Item 13), recorded consistently high means (M = 4.35 to 4.45). These findings indicate that the motivational design of ELSA Speak successfully sustained pupils' interest and shaped positive learning behaviours.

Pronunciation confidence was also recorded high endorsement (M = 4.20, SD = 0.79). Pupils reported feeling more confident speaking English (Item 6: M = 4.32), less afraid of making mistakes (Item 7: M = 4.29) and more willing to participate in class (Item 8: M = 4.26). The data suggest that regular practice with ELSA Speak helped pupils rehearse pronunciation privately, increasing their readiness for classroom speaking tasks.

Speaking anxiety showed moderately high improvement (M = 4.10, SD = 0.82). Items such as "I feel less nervous speaking English" (Item 16: M = 4.23) and "I am less scared of being laughed at" (Item 18: M = 4.18) reveal that pupils experienced notable affective benefits. These findings imply that ELSA Speak functioned as a low-pressure environment where pupils could practise without fear of judgment, improving their emotional comfort during oral tasks.

Challenges and Limitations of Using ELSA Speak

Table 3 presents Items 21 to 25, which address potential difficulties encountered when using ELSA Speak.

Table 3 Challenges and Limitations of Using ELSA Speak

No.	Item	SD	D	N	A	SA	Mean	SD
21	Sometimes ELSA Speak does not recognise my voice correctly.	2	9	21	20	10	3.50	1.03
22	I find some ELSA instructions or feedback difficult to understand.	1	8	18	23	12	3.68	0.96
23	I feel frustrated when my score is low even when I try my best.	2	7	19	22	12	3.69	0.99
24	I need help to understand how to use ELSA Speak.	3	10	18	20	11	3.53	1.05
25	It is difficult to use ELSA when the Internet connection is slow.	2	7	16	24	13	3.74	0.99

The highest-rated challenge was related to internet connectivity (Item 25), with M = 3.74 (SD = 0.99), suggesting that technical infrastructure played an important role in shaping pupils' experiences. Difficulties understanding feedback terminology (Item 22: M = 3.68) and frustration with low scores (Item 23: M = 3.69) also appeared prominently. Challenges in voice recognition were noticeable but less severe (Item 21: M = 3.50), reflecting occasional mismatches between pupils' pronunciation and the app's scoring algorithms.

These findings indicate that while ELSA Speak was generally well-received, certain technological and linguistic barriers influenced pupils' overall satisfaction and ease of use. These challenges align with known limitations in ASR-based tools, particularly when used with younger learners whose voices may not match adult-focused training models.

Interview

The semi-structured interviews with eight upper primary pupils provided more in-depth insights into how ELSA Speak supported their English pronunciation, influenced their confidence, motivation and speaking anxiety, and what challenges they encountered during the four-week intervention. Thematic analysis yielded three overarching themes: (1) pupils' perceptions and experiences of using ELSA Speak, (2) the influence of ELSA Speak on confidence, motivation and speaking anxiety, and (3) challenges and limitations of using the app. Selected verbatim excerpts are presented to illustrate each theme.

Perceptions and Experiences with ELSA Speak

Across all eight participants, there was a consistently positive perception of ELSA Speak as a fun, game-like and useful tool for pronunciation practice. Many contrasted the app with traditional classroom reading tasks, suggesting that the interactive features made learning feel less like schoolwork and more like play.

For instance, one of the pupils described this contrast clearly, stating that:

"Sometimes English lessons can be a bit boring, but ELSA was like a game." (P3)

Similarly, another pupil emphasised how the game-based design made pronunciation practice less intimidating and more fun:

"It was fun because it's not like normal English work. It's more like playing a game but still learning." (P5)

Beyond enjoyment, pupils valued ELSA Speak's clarity in highlighting mispronounced sounds. Many appreciated the colour-coded feedback and visual diagrams showing mouth and tongue positions. One of the pupils described how the system guided self-correction:

"If I say wrongly, it underlines the part in red… red means very wrong and yellow means a bit wrong. So I try again until it becomes green." (P1)

Another pupil similarly valued the articulatory feedback:

"Before this, I didn't know that for some sounds you need to put your tongue in a certain way. The picture helps me understand better." (P6)

Influence on Confidence, Motivation and Speaking Anxiety

A second major theme concerned how ELSA Speak influenced pupils' confidence, motivation and speaking anxiety in relation to English oral tasks. All eight participants indicated that their speaking confidence increased over the four weeks. They attributed this to having opportunities to practise privately, receive feedback and rehearse difficult words before reading in front of the class. One of the pupils described a noticeable shift in how they read aloud in class:

"Last week teacher asked me to read one paragraph. Usually I read very softly, but this time I spoke louder. Because I practised the difficult words the night before in ELSA." (P3)

Another pupil reported a similar change in volume and self-assurance:

"Before this, I always read in a small voice. I scared people will hear my mistake. But now I speak louder. Because I know I practised already." (P4)

Motivation was another strong theme. Pupils discussed the satisfaction of earning streaks, stars and coins, which encouraged them to practise regularly. One pupil described how gamification shaped their behaviour:

"The daily streak. I wanted to see how many days I can continue. And also the leaderboard. I see my friend's score and I want to beat it." (P1)

Similarly, others described how friendly competition motivated more frequent use:

"The leaderboard! My friends and I compare scores. If they get a higher score, I want to beat them. It makes me want to practise more." (P7)

Beyond confidence and motivation, pupils consistently reported that ELSA helped reduce speaking anxiety. Pupils expressed relief about being able to practise without fear of judgment. One pupil noted:

"When I practise with the app, I don't feel shy because nobody is listening except the app. I can try many times." (P2)

Similarly, another pupil noted that the privacy of AI-mediated practice made them feel calmer:

"When I practise alone with ELSA, I don't feel shy. I can try many times. Nobody is listening except the app." (P5)

Challenges and Limitations

Despite the positive experiences, pupils also faced several challenges while using ELSA Speak. One of the most prominent concerns was inaccurate voice detection and inconsistent scoring, which at times reduced the reliability of the feedback. Pupils expressed frustration when the app failed to recognise their pronunciations correctly or assigned low scores despite repeated attempts. One remarked:

"Sometimes the app thinks I said it wrongly, even when I try very hard to say correctly… Sometimes I feel angry because I already try so many times but still get red." (P6)

Another pupil shared a similar sentiment regarding the inconsistency of the scoring system:

"Sometimes the mic doesn't hear me properly. Even when I say correctly, it gives me red. Then I feel frustrated." (P4)

Another common limitation is related to technical issues, particularly slow internet and app freezing. One shared:

"Sometimes the internet is slow at home. When the Wi-Fi slow, the app loads very long. One time it suddenly froze. I had to close and open again." (P5)

In addition, many pupils found some of the feedback terms and instructions too difficult for their age. Words such as intonation, stress, syllable and linking sounds were not easily understood without teacher or parent support. As one explained:

"When I see words like 'intonation' or 'stress', I don't understand. I needed to ask my father to explain." (P2)

Another pupil also commented:

"Some instructions are difficult. Like 'intonation' or 'stress patterns'. I don't know what they mean. I had to ask teacher to explain." (P7)

A further challenge concerned sentence-level pronunciation tasks, where a single error could significantly lower the overall score. One explained:

"The sentence ones. If I forget one word, it marks everything wrong. So I need to practise more times." (P5)

Overall, the interview data indicate that while ELSA Speak was perceived as an enjoyable and beneficial tool for pronunciation practice, refinements are needed in recognition accuracy, linguistic scaffolding and task difficulty, especially when used with younger learners in primary school contexts.

DISCUSSION

Upper primary students' perceptions of using ELSA Speak for English pronunciation practice

The quantitative findings from the perception items (Items 1–10) revealed that students generally held positive perceptions towards ELSA Speak as a tool for pronunciation practice, with mean scores ranging from 4.20 to 4.40. These results indicate strong agreement that the application was enjoyable, user-friendly and beneficial for improving pronunciation accuracy. Items related to perceived usefulness, such as "ELSA helps me speak more clearly," consistently received high ratings, suggesting that learners recognised the app's capacity to support accurate articulation and clearer speech.

The qualitative findings reinforce these findings. Interview participants consistently described ELSA Speak as "fun," "interesting," and "like a game," highlighting its gamified interface, such as stars, coins, streaks and leaderboards, as core motivators. For example, P3 shared that "ELSA was like a game," while P5 emphasised that it "felt like playing but still learning." These remarks resonate with the high mean values in the quantitative data for items on engagement and enjoyment.

The clarity and usefulness of feedback also emerged as a strong theme in both data sets. Pupils valued colour-coded scoring and articulatory guides, which helped them identify specific pronunciation errors. For instance, P1 noted: "If I say wrongly, it underlines the part in red… so I try again until it becomes green." This echoes the high quantitative ratings for items related to correction and error identification.

Collectively, the quantitative and qualitative findings converge to indicate that pupils perceive ELSA Speak as an effective, accessible and motivating tool for pronunciation practice. The combination of immediate feedback, self-paced practice and interactive design appears to enhance their willingness to engage in pronunciation activities, echoing findings from past research on AI-assisted oral practice.

ELSA Speak's influence on students' confidence, motivation and speaking anxiety

The quantitative findings for confidence, motivation and anxiety (Items 11–20) showed that students largely agreed that ELSA Speak enhanced their confidence ($M > 4.20$), increased their motivation to practise English pronunciation ($M > 4.10$), and reduced their speaking anxiety ($M > 4.00$). These positive outcomes were further supported by qualitative interview findings, which provided rich explanations for the shifts.

Pupils frequently reported feeling "more ready," "less afraid," and "braver" after using ELSA Speak. Many explained that the opportunity to practise pronunciation privately helped reduce the fear of making mistakes in front of classmates. For example, P4 described a noticeable change in classroom reading: "Before this, I always read in a small voice… but now I speak louder because I know I practised already." Similarly, P3 recounted speaking more confidently in class after rehearsing difficult words using ELSA. This aligns closely with the high quantitative ratings for confidence-related items, indicating that repeated, individualised practice strengthened pupils' sense of preparedness and self-assurance.

Motivation also emerged as a strong theme across both data sets. Pupils described ELSA Speak's gamified features, such as stars, coins, streaks and leaderboards, as powerful motivators that encouraged them to practise more frequently and persistently. Quantitative items related to enjoyment and willingness to practise achieved some of the highest mean scores in the questionnaire. The interview excerpts illuminate this pattern. For instance, P7 explained, "My friends and I compare scores. If they get higher, I want to beat them," while P6 noted that watching their score increase motivated continued practice. These comments reinforce how reward elements played a significant role in sustaining interest, making pronunciation practice an enjoyable and self-driven activity rather than a compulsory task.

Reductions in speaking anxiety were also evident. The quantitative results for anxiety-related items showed that many pupils felt less nervous, less shy and more comfortable speaking English after using the app. The qualitative data reveal that the private, non-judgmental nature of ELSA Speak was a key factor contributing to this emotional improvement. P5 articulated this clearly: "When I practise alone with ELSA, I don't feel shy. I can try many times. Nobody is listening except the app." Pupils also mentioned feeling calmer and more emotionally steady after using the app, particularly before oral tasks. P3 described this shift: "Last time my heart beat so fast when talking. But now I feel more steady because I practised already."

Overall, the integration of quantitative and qualitative findings demonstrates that ELSA Speak plays a meaningful role in enhancing upper primary students' pronunciation confidence, sustaining their motivation and reducing their speaking anxiety. The combination of instant feedback, gamified rewards and a private practice environment creates favourable affective conditions that support oral language development. These findings underscore the potential of AI-based tools to address both linguistic and emotional needs in primary ESL learning contexts.

Challenges and limitations experienced by the students when using ELSA Speak

Although the overall findings revealed highly positive perceptions of ELSA Speak, both quantitative and qualitative data indicated several challenges. These included inaccurate voice detection, inconsistent scoring, technical disruptions, difficulty understanding linguistic feedback, and task demands that were sometimes unsuitable for younger learners. While these limitations did not substantially diminish pupils' overall enjoyment

or willingness to use the app, they highlight areas in which AI-based pronunciation tools require refinement when used with younger learners.

One of the most frequently reported challenges involved the inaccuracy of ELSA Speak's speech-recognition system. This concern corresponded with comparatively lower SPSS mean scores for the related questionnaire items and was further illuminated by the interview data. Several pupils reported frustration when the app marked what they believed to be correct pronunciations as "wrong." For example, P6 explained, "Sometimes the app thinks I said it wrongly even when I try very hard… I feel angry because I already try so many times but still get red." P4 expressed a similar experience, noting that "Sometimes the mic doesn't hear me properly. Even when I say correctly, it gives me red." These examples reflect known challenges in automatic speech-recognition technology, which is often trained primarily on adult speech and may struggle with children's developing phonological patterns and pitch ranges.

Technical issues also emerged as a recurring limitation. Pupils described encountering slow internet connections, long loading times and occasional app freezing, particularly when practising at home. These difficulties are consistent with the higher SPSS mean score for the item related to connectivity problems. P3 shared, "When my Wi-Fi slow, the app loads very long. One time the whole screen froze and I had to restart everything." Such interruptions disrupted the flow of practice and occasionally reduced learners' motivation, suggesting that stable digital infrastructure is essential for AI-mediated pronunciation activities to function effectively.

Another challenge related to pupils' difficulty understanding some of the linguistic terminology used in ELSA Speak's feedback. Terms such as "intonation," "stress patterns," and "linking sounds" were unfamiliar to many pupils and required teacher or parental explanation. P2 noted, "When I see words like 'intonation' or 'stress', I don't understand," while P7 added that some feedback instructions were "too difficult." This suggests that although ELSA Speak offers detailed linguistic feedback, the complexity of its terminology may not be developmentally appropriate for younger ESL learners, who may need simplified explanations or additional scaffolding.

Sentence-level tasks posed another significant limitation. Pupils frequently explained that these tasks were challenging because a single mispronunciation could result in a low overall score. P5 expressed this frustration, stating, "If I forget one word, it marks everything wrong. So I need to practise more times." This feedback highlights how overly stringent scoring mechanisms may discourage young learners, particularly those who are still developing foundational reading and pronunciation skills.

Overall, while these challenges did not deter pupils from using ELSA Speak, they did contribute to moments of frustration and indicate important areas for improvement. Enhancing speech-recognition accuracy for children's voices, simplifying feedback terminology, adjusting scoring mechanisms for longer responses and ensuring stable technological access may help improve the usability of ELSA Speak for younger learners. Addressing these limitations is crucial for ensuring that AI-based pronunciation tools remain accessible, developmentally appropriate and supportive of meaningful learning experiences in primary school contexts.

CONCLUSION

This study investigated upper primary students' perceptions of ELSA Speak as an AI-driven pronunciation training tool, focusing on its impact on their confidence, motivation, speaking anxiety and the challenges they encountered during a four-week implementation period. Overall, the findings revealed that learners responded positively to the app, perceiving it as enjoyable, accessible and effective for improving pronunciation. Features such as instant colour-coded feedback, articulatory guides and gamified elements, including stars, streaks and leaderboards, created an engaging learning environment that encouraged active participation and made pronunciation practice less daunting for young ESL learners.

In addition to supporting linguistic development, ELSA Speak made a meaningful contribution to learners' emotional readiness for speaking tasks. Many pupils reported feeling more courageous, better prepared and less anxious about making mistakes after practising privately in a non-judgmental digital space. This improvement in confidence and reduction in anxiety was consistently evident across both the quantitative scores and

qualitative remarks, suggesting that the app has the potential to strengthen not only pronunciation but also learners' self-belief and willingness to speak in class. These affective gains highlight the valuable role AI pronunciation tools can play in creating a supportive and motivating environment for young ESL learners.

Despite these strengths, several limitations surfaced, particularly relating to technological constraints and the developmental needs of younger users. Issues such as inaccurate speech recognition, inconsistent scoring, slow internet connectivity and occasional app freezing disrupted learning and caused frustration. Some pupils struggled to understand feedback terms like intonation or stress, highlighting the need for more age-appropriate explanations and simplified scoring mechanisms.

In light of these challenges, several improvements are recommended. To facilitate better understanding of key pronunciation concepts, learners would benefit from supplementary materials that simplify complex linguistic terms, ideally incorporating visual representations, diagrams or child-friendly explanations. Furthermore, ensuring stable internet access, updated devices or providing offline practice options would also help minimise disruptions caused by freezing or delayed voice recognition. Finally, additional assessment opportunities beyond the four weeks are encouraged to determine the long-term impact of ELSA Speak on sustained pronunciation development and oral performance. While these refinements, tools like ELSA Speak hold strong potential for effective integration into Malaysian primary ESL classrooms, supporting both linguistic growth and learners' confidence in spoken English.

REFERENCES

1. Adawiah, R., Muliati, A., & Samtidar, S. (2024). The effect of ELSA Speak application on students' English pronunciation development. Journal of Excellence in English Language Education, 3(1).
2. Ag-Ahmad, N., Syed Mohamed, A. T. F., & Majilang, D. B. (2025). Voices from ESL classrooms: Overcoming challenges and enhancing English language teacher education in Malaysia. Studies in English Language and Education, 12(1): 328-345 DOI:10.24815/siele.v12i1.39831
3. Amin, E. A.-R. (2024). EFL students' perception of using AI text-to-speech apps in learning pronunciation. Migration Letters, 21(3), 887–903.
4. Aulia, L. G. Y. M., & Santosa, M. H. (2025). AI in language education: Systematic review of ELSA Speak's impact on English. Jurnal Pendidikan Bahasa Inggris Indonesia, 13(1).
5. Aziz, A. A., & Kashinathan, S. (2021). ESL learners' challenges in speaking English in Malaysian classroom. International Journal of Academic Research in Progressive Education and Development, 10(2), 983–991. https://doi.org/10.6007/IJARPED/v10-i2/10355
6. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
7. Caiza, G., Villafuerte, C., & Guanuche, A. (2025). Interactive application with virtual reality and artificial intelligence for improving pronunciation in English learning. Applied Sciences, 15(17), 9270. https://doi.org/10.3390/app15179270
8. Ch'ng, L. C., Chan, A. N., & Saufi, N. A. (2025). Exploring ESL public speaking anxiety among undergraduates. Journal of Creative Practices in Language Learning and Teaching, 13(1), 146–169. https://doi.org/10.24191/cplt.v13i1.4234
9. Creswell, J. W., & Plano Clark, V. L. (2017). Designing and conducting mixed methods research (3rd ed.). Thousand Oaks, CA: SAGE Publications.
10. Derwing, T. M., & Munro, M. J. (2015). Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research. John Benjamins.
11. Dhivya, D. S., Hariharasudan, A., Ragmoun, W., & Alfalih, A. A. (2023). ELSA as an Education 4.0 tool for learning business English communication. Sustainability, 15(4), 3809. https://doi.org/10.3390/su15043809
12. Hashim, H. (2024). Speaking anxiety among ESL learners in the Malaysian classroom. UiTM eBulletin APB, 13, 22–28. Universiti Teknologi MARA.
13. Hazita, A. (2021). Implementation and challenges of English language education reform in Malaysian primary schools. 3L The Southeast Asian Journal of English Language Studies, 22(3), 65-78. DOI:10.17576/3L-2016-2203-05

14. Hung, H.-T., Yang, J.-C., Hwang, G.-J., Chu, H.-C., & Wang, C.-C. (2018). A scoping review of research on digital game-based language learning. Computers & Education, 126, 1–13. https://doi.org/10.1016/j.compedu.2018.06.001
15. Isaacs, T., & Trofimovich, P. (Eds.). (2017). Second Language Pronunciation Assessment: Interdisciplinary Perspectives (Vol. 107). Multilingual Matters & Channel View Publications. http://www.jstor.org/stable/10.21832/j.ctt1xp3wcc
16. Kholis, A. (2021). ELSA Speak app: Automatic speech recognition (ASR) for supplementing English pronunciation skills. Pedagogy: Journal of English Language Teaching, 9(1), 1–14. https://doi.org/10.32332/joelt.v9i1.2723
17. Lai, K. W. K., & Chen, H. J. H. (2022). An exploratory study on the accuracy of three speech recognition software programs for young Taiwanese EFL learners. Interactive Learning Environments, 32(5), 1582–1596. https://doi.org/10.1080/10494820.2022.2122511
18. Lau, K. Y. L., Adi Badiozaman, I. F., & Leong, H. J. Y. K. (2025). Predictors and moderators of ESL classroom speaking anxiety: A structural equation modelling approach. Journal of Applied Research in Higher Education. https://doi.org/10.1108/JARHE-05-2025-0380
19. Liu, L. (2024). Impact of AI gamification on EFL learning outcomes and nonlinear dynamic motivation: Comparing adaptive learning paths, conversational agents, and storytelling. Education and Information Technologies, 30, 11299–11338. https://doi.org/10.1007/s10639-024-13296-5
20. Li, W., Mohamad, M., & You, H. W. (2025). Exploring the effects of using automatic speech recognition on EFL university students with high speaking anxiety. International Journal of Information and Education Technology, 15(1), 187–194. https://doi.org/10.18178/ijiet.2025.15.1.2231
21. Ministry of Education Malaysia. (2021). English language education reform 2015–2025: Planning in retrospect, implementation and prospective planning. English Language Teaching Centre Ministry of Education Malaysia.
22. Mohammed, N. A., Raof, A. H. A., & Yusof, M. A. M. (2021). Challenges in implementing the CEFR in English language classrooms. LSP International Journal, 8(2), 115–128. https://doi.org/10.11113/lspi.v8.17977
23. Nadeem, S., Iftikhar, S., & Rasheed, A. (2025). Addressing the fear of speaking: The role of AI in reducing speaking anxiety in ESL learners. Journal of Applied Linguistics and TESOL, 8(2), 825–833.
24. Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. ReCALL, 36(1), 4–21. https://doi.org/10.1017/S0958344023000113
25. Pham, V. T. T., & Pham, A. T. (2025). English major students' satisfaction with ELSA Speak in English pronunciation courses. PLoS ONE 20(1): e0317378. https://doi.org/10.1371/journal.pone.0317378
26. Prakash, T., & Kausalya, S. (2025). Speech-to-text AI for improving English pronunciation in ESL learners. International Journal of Innovative Science and Research Technology, 10(8), 1341–1349. https://doi.org/10.38124/ijisrt/25aug1065
27. Vincent, J., Yunus, M. M., & Said, N. E. M. (2025). Using AI platforms to improve listening and speaking skills in ESL primary students. International Journal of Academic Research in Progressive Education and Development, 14(1), 1829–1843. http://dx.doi.org/10.6007/IJARPED/v14-i1/24848
28. Sari, F. K., & Kurniawan, P. Y. (2025). EFL students' perceptions of AI speech recognition in pronunciation practice: The case of Google Read Along. English Review: Journal of English Education, 13(1), 205–214. https://doi.org/10.25134/erjee.v13i1.10751
29. Sungkur, R. K., & Shibdeen, N. (2025). Pronunciation trainer for second language learning using generative AI. International Journal of Educational Technology in Higher Education, 22(64). https://doi.org/10.1186/s41239-025-00561-x
30. Yaccob, N. S., Rahman, S. F. A., Mohamad, S. N. A., Abdul Rahim, A. A., Rashid, K. K. A., Aldaba, A. M. A., Yunus, M. M., & Hashim, H. (2022). Gamifying ESL classrooms through gamified teaching and learning. Arab World English Journal (AWEJ) Special Issue on CALL, 8, 177–191. https://doi.org/10.24093/awej/call8.12
31. Zulkarnain, N. S., & Yunus, M. M. (2023). Primary teachers' perspectives on using artificial intelligence technology in English as a second language teaching and learning: A systematic review. International Journal of Academic Research in Progressive Education and Development, 12(2), 861–875. https://doi.org/10.6007/IJARPED/v12-i2/17119