ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Effectiveness of AI-Powered Speech Recognition App in Enhancing Speaking Skills of Intermediate-Level ELLs in Mianwali

1. Asad Imran Shah (asadimran328@gmail.com) – MEN-S24-26 – M. Phil Scholar at University of Mianwali
2. Tariq Usman, Assistant Professor of English, University of Okara, (corresponding author, tariq.usman@uo.edu.pk
3. Rida Fatima (bestrida17@gmail.com) - MAI-F25-08 – MS Scholar at University of Mianwali

DO1- https://doi.org/10.5281/zenodo.20337858

Abstract

Traditional means of learning English, such as the Grammar Translation Method (GTM), have focused on the development of writing skills while neglecting speaking skills. This negligence has resulted in poor pronunciation, low proficiency, and speaking anxiety among English Language Learners (ELLs). In order to address this gap, the effectiveness of an AI-powered Automatic Speech Recognition ASR application, ELSA Speak, was assessed in this study among the intermediate-level students in the Mianwali district. 60 male students, who were recruited via purposive sampling at Government. Graduate College, Mianwali, were randomly assigned to either the Experimental Group, who used the designated application for 6 weeks, or the Control Group, who were taught traditionally. Their speaking performance was assessed through pre-tests and post-tests via a rubric. Semi-structured interviews were also conducted with 8 participants to explore their experience of and challenges regarding using this application and their perspectives were weighed through the Technology Acceptance Model. The quantitative analysis of the post-tests showed that EG outperformed the CG with mean Composite Scores of 19.33 and 14.90 respectively. The analysis of the interviews also yielded high perceived usefulness due to the application providing a non-judgemental environment that propelled learner autonomy, reduced their speaking anxiety, and boosted their confidence. However, some learners were also quick to point out challenges such as unreliable internet connections that made this application hard to use and the increased cost of mobile data. This study concludes that AI-powered ASR applications are an effective intervention to address the lack of attention given towards the development of speaking skills in overcrowded ESL classrooms.

Keywords: *AI-Powered Language Learning, Automatic Speech Recognition, ELSA Speak, Speaking Skills, ELLs, MALL*

Introduction

The influence of globalization, coupled with the advancements in technology, has reshaped language acquisition. English, after having solidified its status as the global lingua franca, is no longer a mere academic subject but is an integral instrument for achieving success in academic and professional spheres of life. As a result, the language learners worldwide have set out to develop proficiency in the English language, particularly in oral communication. But traditional methods of teaching English in many developing countries, such as Pakistan, focus on the development of other language skills, especially writing, and students struggle to acquire necessary speaking skills in such environments. This study has been undertaken to investigate a technology-driven solution to the aforementioned problem: the use of AI-powered Automatic Speech Recognition (ASR) applications, such as ELSA Speak, to enhance the speaking skills of English Language Learners (ELLs).

Background of the Study

Being able to communicate effectively in the English language is considered an important and valuable skill in the 21st century. Sound proficiency in spoken English unlocks access to higher education opportunities in another country, increases the chances of employability in both local and international job markets, and most importantly, facilitates cross-cultural communication (Thomas et al., 2016). Above all the subskills needed for effective oral communication, it is the pronunciation that directly affects the intelligibility of a learner's spoken words and then his/her confidence. As has been argued by the scholars, poorly developed pronunciation leads to the creation of breakdowns during

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

communication, which results in frustration, not only among the listener(s) but also by the speaker, even when the sentences uttered are grammatically correct and are complemented with appropriate vocabulary (Derwing & Munro, 2005).

The situation of the English language in Pakistan is rather unique and, to put it bluntly, paradoxical. On one hand, it is the official language of the country and a gateway to socioeconomic advancement for the Pakistanis, yet the methodologies of teaching English, especially in the public sector education system, are ill-suited for the development of communicative competence in the English language. One factor that has contributed to the negligence in speaking skills is the dominance of traditional teaching methodologies, chiefly the Grammar Translation Method. Another factor that makes the teaching of spoken skills difficult for the learners is the overcrowded classrooms, which are a common sight in many public sector educational institutes. The progress of the students in the Pakistani education system is determined by the written examinations conducted annually. Since the speaking and listening skills are not formally assessed in these examinations, these core language skills are entirely ignored in favour of memorizing the content that can be reproduced in the written form (Yaqub, 2009).

Due to these limitations of the traditional teaching methodologies, teachers, learners, and researchers have started taking advantage of the technology to enhance the language learning. This aspect of language learning, which has been known as Computer-Assisted Language Learning (CALL), has gone through many phases of evolution over these decades. This evolution has recently transitioned into Mobile-Assisted Language Learning (MALL). The tools of language learning have been placed directly in the hands of a multitude of the learners through the widespread adoption of smartphones and tablets, alongside the affordability of the internet services. The advancements in the field of Artificial Intelligence (AI) and its integration in MALL have also revolutionized language learning and the development of speaking skills, mainly through the AI-powered Speech Recognition System (ASR) technology. These applications contain features that offer instant feedback, something that is absent in a traditional, overcrowded English as a Second Language (ESL) classroom. Most crucially, such applications provide a language learner with a private space where his errors in speech would not be judged as his weakness. They offer a solution for language learning that is easily scalable, accessible, and personalized in accordance with the needs and goals of the individual students, with a safe space where individual corrective feedback is provided that cannot be offered by the traditional ESL classroom settings.

Statement of the Problem

Despite the very apparent potential of the AI-powered ASR language learning applications, made specifically for the improvement of speaking skills, there exists a significant gap in the academic literature regarding their effectiveness, particularly within the Pakistani ESL context. The research problem that this study has been enacted to address is the lack of empirical evidence on determining the impact of such MALL tools on the speaking skills of intermediate-level ELLs in a semi-urban district of Pakistan. Thus, in light of the problems highlighted, the main goal of this study is to investigate and determine the effectiveness of using AI-powered ASR applications for the enhancement of the speaking skills (pronunciation and fluency) of the intermediate-level ELLs in Mianwali district, as compared to traditional teaching methods.

Research Objectives

1. To determine the effect of using AI-powered ASR app on the pronunciation and fluency of intermediate-level ELLs in Mianwali.
2. To explore the opinions of Intermediate Level ELLs in Mianwali on the use of AI-powered ASR app for speaking skills development.

Significance of the Study

From academic point of view, this study has contributed to the scholarly literature related to CALL and MALL by directly addressing the technological, demographic and geographical gaps in the contemporary literature through the empirical data on the effect of AI-powered ASR technology, conducted among an under-researched learner group in a developing district. The primary beneficiaries of the findings and insights gathered from this study are the students themselves through

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 2 2026

the validation of an accessible and autonomous tool that empowers them to overcome the limitations of the traditional ESL classroom by practicing their speaking skills in a low-anxiety environment. For English Language Teachers (ELTs), this study has offered the concrete evidence of the effectiveness of a language learning tool that they can directly integrate into their pedagogy.

Literature Review

The integration and recent developments in the field of Artificial Intelligence (AI), especially in the form of Automatic Speech Recognition technology, have opened new opportunities to address the challenges with regards to the development of speaking skills among ESL learners. Usman et al, (2025) argue that there are many interesting methods of teaching the English language; however, teachers are not fully trained to execute them, therefore, they mostly rely upon the lecture method which is one of the leading barriers in the learner's autonomy. Applications and software built on top of ASR technologies take spoken text as input, transcribe it into text and also offer immediate feedback on core aspects of speech, such as pronunciation and fluency. Some modern language learning apps are designed specifically for the enhancement of speaking skills. ELSA Speak is one of the prominent apps that is designed solely for the development of speaking skills with feedback on pronunciation errors and offering correct pronunciation playback (Gusrianto, 2025).

AI-powered ASR apps, such as ELSA Speak, are built solely for the enhancement of weaker speaking skills among ESL learners. A study conducted by Anggraini (2022) showed that the ELSA Speak application helped in the improvement of students' pronunciation skills. Bashori et al. (2024) also explored the effectiveness of two ASR-based systems, namely ILI and NovoLearning on improving word-level and sentence-level pronunciation among Indonesian high school students. Another mixed-method study conducted by Sun (2023) on Chinese EFL showed that an experimental group making use of apps built on top of ASR technology with correction from their peers performed well above a control group whose participants were taught with traditional teacher-led feedback in L2 pronunciation, boosted comprehensibility and speaking skills in the language that is spoken globally.

These MALL apps let ESL learners practice their speaking repeatedly with instant feedback in a low-anxiety, personal environment and thus, helping them acquire fluency in the L2. For instance, Jiang et al. (2022) examined the effectiveness of ASR technology among Chinese college students, with results revealing that the experimental group performed exceptionally well on all metrics of lexical complexity and increased the speed of fluency. The personalized lessons and non-judgemental feedback provided by these language learning applications have the potential to increase the confidence among ESL learners. Li (2024) also discovered that an experimental group who used MALL apps showed significant improvement in their speaking skills as compared to the learners who were being taught in traditional classrooms.

The opinions and experiences of the language learners serve as the catalyst behind the adoption of ASR-enhanced language learning applications. A study conducted by Akman and Karahan (2023) on university ELT learners in Türkiye found that students considered MALL not only effective but also beneficial for the improvement of pronunciation, thus helping them to be motivated and engaged. ESL language learners from the field of engineering, as investigated by Moulieswaran and Kumar (2023), favoured AI-powered language learning tools to help them learn English fast. Metruk (2021) also investigated the perceptions of Slovak EFL learners regarding these applications, with positive views and improvement in various language skills reported.

Although the adaptation of CALL and MALL is not as widespread in Pakistan as it is in developed countries, still this aspect of English language teaching is being explored by scholars across Pakistan. An experimental study was conducted by Bashir et al. (2022) in Lahore, revealing that MALL had a positive effect on the acquisition of vocabulary among ESL learners compared to the traditional methods of teaching English. Shaheen et al. (2024) further set out to explore how emerging technologies, such as smartphones, are changing the process of foreign/second language learning while also revealing the supportive views of university students regarding the effectiveness of mobile devices and MALL tools for learning English. Although the studies regarding CALL, MALL and the usage of AI tools in ELT are not slowing down, but much of them focus on general language learning, vocabulary acquisition or listening comprehension. Studies regarding the exploration of AI-powered ASR

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 2 2026

applications such as ELSA Speak, which are designed specifically to enhance speaking skills (such as fluency and pronunciation), are very small in number in Pakistan.

Recent scholarship has further solidified the efficacy of MALL in enhancing speaking skills of Pakistani ELLs. For instance, Riaz and Kausar (2024) demonstrated through a 32-week experimental study that AI applications such as "Readlee" and "@Voice Aloud Reader" yielded significant gains in college-level ESL students' speaking proficiency compared to traditional methods. Similarly, Zohaib et al. (2025) utilized a quantitative approach to show that even social tools like WhatsApp can significantly enhance university students' fluency, vocabulary and grammatical competence by providing an informal yet interactive setting for practicing speaking. While Farhat and Dzakiria (2017) noted that CALL tools are particularly effective in reducing pronunciation barriers, specifically regarding stress and intonation patterns at the school level. Newer research by Fatima et al. (2025) has shifted focus toward rural learners. Their mixed-methods study highlights the potential of AI-powered chatbots to bridge pedagogical gaps for learners in less developed regions by integrating socio-linguistic and pedagogical evaluations.

Methodology

This study utilises the Mixed-Methods Embedded Experimental Design. The main component of this study design was a true experimental pre-test-post-test, control group design. The target population for this particular study consisted of all the male intermediate-level (Higher Secondary School Certificate, HSSC; Classes 11 and 12th) English Language Learners (ELLs) within the administrative jurisdiction of Mianwali District, Punjab, Pakistan. 60 students were then recruited from the initially selected pool of 80 ELLs and randomly assigned to either the control or experimental group.

Data was collected over a predefined period through a combination of qualitative and quantitative instruments which were designed to be valid and reliable to answer the research questions. Read-aloud tasks were used both for the pre-tests and the post-tests. The audio recordings of the students were scored by an independent but trained rater through a rubric that was inspired by the speaking assessment of the International English Language Testing System (IELTS). The ELSA Speak application, a prominent AI-powered ASR app. was used as the intervention for the EG. The last instrument used in collecting the data for this study was the semi-structured interviews for qualitative interpretation. The numerical data which was obtained from the pre-test and post-test scores were analyzed through the Statistical Package for the Social Sciences (SPSS) Version 26.0. Qualitative content analysis was used to analyse the semi-structured interviews.

Data Analysis and Results

The main objective of this study was to empirically determine the effectiveness of AI-powered Automatic Speech Recognition (ASR) application, namely the ELSA Speak mobile application in enhancing the speaking skills of intermediate-level ELLs in the district of Mianwali. The sample was further divided into the 30 students as the Experimental Group (EG) who made use of the designated application for a period of six weeks, while the remaining 30 students were added into the Control Group (CG), who continued to receive instruction from the traditional classroom.

Quantitative Analysis

The analysis of pre-test scores is important to establish the baseline speaking proficiency of the ESL learners. The descriptive statistics of the pre-test scores have been displayed in the table 1.

Table 1

Group Statistics: Pre-Test Comparison

Score Type	Group	N	Mean	Std. Deviation	Std. Error Mean
Composite Score	EG	30	12.60	2.527	.461
	CG	30	12.87	2.501	.457
Total Fluency Score	EG	30	12.37	2.327	.425
	CG	30	12.77	2.445	.446
Total Pronunciation Score	EG	30	12.47	2.688	.491
	CG	30	12.57	2.648	.483

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

As has been statistically demonstrated in the above Table 1, the speaking assessment conducted before the intervention yielded similar scores for EG and CG. The composite score for the EG was 12.60 (SD = 2.527), while the composite score for the CG was 12.87 (SD = 2.501). The difference between the mean scores of the two groups was just 0.27, which is statistically negligible. The similarity of the scores between the aforementioned groups also continued in the subcategories of the rubric to assess fluency and pronunciation. The EG had a Total Fluency mean of 12.37 as well as a Total Pronunciation mean of 12.47. The Total Fluency mean for the CG was 12.77 and the average score for the Total Pronunciation for the CG was 12.57.

After the six-week led intervention period, the descriptive statistics calculated after the conduction of the post-test showed that a significant divergence in the speaking performance of the two groups was apparent. Table 2 shows this difference in the performance of the ELLs in speaking skills.

Table 2

Group Statistics: Post-Test Comparison

Score Type	Group	N	Mean	Std. Deviation	Std. Error Mean
Composite Score	EG	30	19.33	1.583	.289
	CG	30	14.90	2.023	.369
Total Fluency Score	EG	30	18.57	1.716	.313
	CG	30	14.50	1.852	.338
Total Pronunciation Score	EG	30	19.43	1.591	.290
	CG	30	14.87	2.013	.367

A clear gap in the performance of the two groups has been illustrated in Table 2 with the increase of the Mean Composite Score as 19.33 (SD = 1.583) for the EG and 14.90 (SD = 2.023) for the CG. As far as the Mean Composite Scores are concerned, the Experimental Group outperformed the Control Group by an average of 4.43 points. The EG showed a remarkable improvement in fluency with an increase in their Mean Total Fluency Score of 18.57 (SD = 1.716) but the control group only scored 14.50 (SD = 1.852). The difference in the mean scores of the Total Pronunciation Score was even more apparent between the two cohorts, with EG scoring 19.43 (SD = 1.591), compared to the CG's mean of 14.87 (SD = 2.013).

An independent samples t-test was conducted on the pre-test scores to statistically confirm the equivalence of the EG and the CG.

Table 3

Independent Samples t-test - Pre-Test Scores

Score Type	Levene's F	Levene's Sig.	t	df	Sig. (2-tailed)
Composite Score	.127	.723	-.411	58	.683
Total Fluency	.600	.442	-.649	58	.519
Total Pronunciation	.166	.686	-.145	58	.885

As has been shown in Table 3, the t-test results support the null hypothesis of no difference, as the t value for the Composite Score is -0.411 and the p value is .683. Because the significance value was much higher than the alpha level of 0.05, the high p-value statistically validates that there is no difference in the speaking proficiency between the EG and the CG.

Paired samples t-tests were run to assess the magnitude of improvement within each group. The paired samples t-test for the EG demonstrated remarkable growth in their speaking proficiency owing to the use of the AI-powered ASR application. The improvement in the Composite Score jumped from a mean of 12.60 to 19.33. The t-test for this change was $t(29) = 25.668$ and $p < .001$, signaling that this change in score was statistically significant. The CG who received teaching through traditional means also demonstrated improvement in their spoken skills, but that improvement was not as pronounced as the EG. Their mean composite score improved from 12.87 to 14.90 and the t-tests for the improvement of their scores were $t(29) = 11.143$, $p < .001$, making their improvement also statistically significant. Table 4 shows the details of the paired samples t-tests for the two groups.

Table 4

Paired Samples t-test Results

Group	Pair	Mean Difference	Std. Deviation	t	df	Sig. (2-tailed)
EG	Composite Score	6.733	1.437	25.668	29	< .001
	Total Fluency	6.200	1.186	28.630	29	< .001

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

	Total Pronunciation	6.967	1.847	20.656	29	< .001
CG	Composite Score	2.033	.999	11.143	29	< .001
	Total Fluency	1.733	1.388	6.840	29	< .001
	Total Pronunciation	2.300	1.149	10.962	29	< .001

An independent samples t-test was also conducted to compare the post-test scores of the EG and the CG with the main goal to answer the first research question. The results obtained from these tests favour the experimental group, as a very noticeable difference emerged between the EG (M = 19.33) and the CG (M = 14.90), t(54.827) = 9.453 in the Composite Score. The t-value reported here has been based on "Equal variance not assumed" in Table 5.

Table 5

Independent Samples t-test for Post-Test Scores

Score Type	Levene's F	Levene's Sig.	t	df	Sig. (2-tailed)	Mean Difference
Composite Score*	9.396	.003	9.453	54.827	< .001	4.433
Total Fluency	.710	.403	8.822	58	< .001	4.067
Total Pronunciation*	7.463	.008	9.751	55.060	< .001	4.567

The EG also showed a remarkable performance over fluency (Mean Difference 4.07), t(58) = 8.822 with p < .001 and pronunciation with a mean difference of 4.57, t(55.060) = 9.751 with p < .001. A one-way analysis of covariance was conducted in which "Post-test Composite Score" was a dependent variable, experimental and control groups were the fixed factor, and "Pre-test Composite Score" was the covariate.

The result for ANCOVA, presented in Table 6, revealed a significant main effect for the Group variable: F(1,57) = 437.906,p < .001. The analysis showed a partial eta squared value of .885. The value of .885 obtained in this study is very large, suggesting that 88.5% of the variance in the post-test speaking assessment can be attributed to the intervention used (AI-powered ASR app vs. traditional).

Table 6

ANCOVA: Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	444.918	2	222.459	307.284	< .001	.915
Intercept	175.336	1	175.336	242.192	< .001	.809
Composite_Scores	150.101	1	150.101	207.336	< .001	.784
Group	317.023	1	317.023	437.906	< .001	.885
Error	41.265	57	.724			
Total	18065.000	60				

Item Level Analysis

The rubric used for the assessment of pronunciation and fluency of the intermediate-level ELLs consisted of 5 separate items for each subcategory of the speaking skills. The statistical analyses shown above have mainly displayed an aggregate score for the pronunciation and fluency, without getting into a deep. Item-level analysis of what certain aspects of pronunciation and fluency were improved. The scores for rubric items were collected on a Likert scale from 1 to 5. An item-level breakdown of the post-test mean scores among the two groups has been presented in Table 7.

Table 7

Item-Wise Comparison of Post-Test Scores

Assessment Item	Group	Mean	t-value	df	p-value
Does the student read at a natural, steady pace?	EG CG	3.87 2.93	9.011	42.647*	< .001

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Are pauses logical (at punctuation)?	EG CG	3.60 2.50	7.940	58	< .001
Does the student group words into meaningful phrases?	EG CG	3.90 3.53	2.408	45.135*	.020
Does the student frequently stop to fix mistakes?	EG CG	3.50 2.53	5.491	58	< .001
Does the student repeat words or stutter?	EG CG	3.70 3.00	7.167	45.727*	< .001
Are vowel and consonant sounds produced accurately?	EG CG	4.10 3.47	5.356	44.298*	< .001
Is the stress placed on the correct syllable?	EG CG	3.97 2.97	7.902	33.300*	< .001
Does the student emphasize key content words?	EG CG	3.73 2.53	8.050	58	< .001
Are word endings and clusters clear?	EG CG	4.00 3.47	4.000	47.038*	< .001
Is the rise and fall of the voice natural?	EG CG	3.63 2.43	8.266	58	< .001

Pronunciation. *It was identified from the literature reviewed in Chapter 2 of this thesis that Pakistani English language learners grapple with poor pronunciation, both in the production of individual sounds (segmental features) and speaking words with appropriate stress and intonation (suprasegmental features). The data presented in Table 7 reveals that the AI-powered ASR application was very effective in addressing issues related to incorrect pronunciation.*

Production of Accurate Speech Sounds. *For the item that assessed whether the learners produced correct consonant and vowel sounds, the students in the Experimental Group scored 4.10 as compared to the mean score of 3.47 among the CG on the same rubric item in the post-test. What this difference in the mean scores of the two groups suggests is that the "phoneme breakdown" feature of the ELSA Speak application was very helpful for the students in reducing incorrect pronunciation, as the teachers in the overcrowded classrooms could not correct the mispronunciations of individual students in their overcrowded classrooms. The color-coded feedback, especially the phonemes highlighted in red, helped learners identify the exact speech sounds that were being pronounced incorrectly and to correct them on their second try.*

Word Stress. *The most prominent difference in the mean scores between the EG and the CG was observed in the rubric item for assessing the placement of correct word stress on a specific syllable. The mean score for the EG was 3.97 and 2.97 for the CG. The Grammar Translation Method (GTM) favours the teaching of vocabulary to be carried out through memorization of lists. Students learn the meaning of the words but not their correct stress pattern. This gap is very apparent in the mean performance of the CG in that particular rubric item. The students within the EG were facilitated through the app that displayed the correct stress pattern to pronounce a word, and incorrect word stresses were marked by the app as well. Moreover, the app also helped students in the CG helped overcome their instinct to drop the word endings in certain words, such as "-ed" in past tense and "-s" in plural or in present tense, with them scoring 4.00 on the rubric assessing word endings as compared with the 3.47 scored by the CG.*

Fluency. *It is not easy for English language learners to develop fluency in standard ESL classrooms. The numerical data in the table above shows that the low-anxiety environment provided by MALL tools, such as ELSA Speak, helped students speak English, with natural pacing and a reduction in hesitation was observed.*

Natural Pacing. The mean score for the EG on the term about the students reading at a steady and natural pace was 3.87, with the CG only scoring 2.93. The score of 3 reveals that CG's speech was robotic and stilted (forced and awkward), a trait much common among the learners who are transitioning from L1 to L2. Since the students within the EG were facilitated by the application to listen and repeat sentences in English, they were more successful in mimicking the natural rhythm and pacing.

Pauses. *Naturally, it is expected by the speakers to pause between sentences. But language learners often pause in the middle of the sentences to remember how the upcoming words will be*

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

pronounced. Thus, the rubric also accounted for the pauses at different punctuation markers, such as full stops, commas, etc. The difference in the mean scores regarding this point of the rubric was staggering between the two groups. The mean score for EG was 3.60 as compared to the 2.50 for the CG. The students in the CG paused more frequently in the sentences than the students from the EG.

Reduction in Stuttering. *This was the last item in the fluency rubric, which was used to assess whether the student repeats words or stutters, with higher scores indicating fewer stutters and reduced repetitions, to keep the overall rubric score consistent. The score for EG on this rubric item was 3.70, much higher than the 3.00 scored by the CG. The ELLs in the EG gathered enough confidence by practicing their speaking skills through the application, which resulted in the reduction of repetition, such as "the- the- the bank of the river."*

Qualitative Analysis

After the demonstration of the effectiveness of the AI-powered ASR application in improving the pronunciation and fluency among intermediate-level ELLs through quantitative analysis, semi-structured interviews were conducted among 8 intermediate-level ELLs to explore their experience of using the prescribed application alongside the challenges they came across.

Students from the EG considered the ELSA Speak application not only as a novel tool but also as a very effective one to help them achieve tangible improvements in their speaking skills in the English language. Participant 1 observed, "Before this, I couldn't even say 'education system' properly… Last week our English sir asked me to read a passage, and he paused and said, 'Your pronunciation is much clearer now.'" Participant 2 noted, "My biggest issue was word endings - I would say 'lookin' instead of 'looking'… The app caught this every single time. Now I consciously remember to pronounce the -ing properly." Participant 4 reported, "My pronunciation of medical terms improved dramatically. Words like 'diagnosis,' 'pharmaceutical,' and 'anatomy'… I can now pronounce correctly."

This observable improvement in their speaking skills directly resulted in boosting their confidence. Participant 2 highlighted, "Last Eid, my cousin from Karachi visited. She… always laughed at my 'desi' accent. This time, we had a 10-minute conversation in English, and she actually complimented my improvement." Participant 3 shared, "My father was so proud, he bought me chicken karahi that night." The boost in confidence was even more apparent for participant 6, who was once an avoider but the application encouraged him to speak for 2 minutes in his college debating society and received an award. He stated, "In our college's debating society, I always avoided participating. Last Friday, I gave a 2-minute speech… and won third prize. That wouldn't have happened before ELSA." There is no better evidence of the sheer utility of ELSA Speak than the reduction in stuttering as reported by Participant 8. He confessed, "I have a stuttering problem… The app's patience - letting me repeat without rushing - helped reduce my stuttering by 50-60%… My parents cried when they saw me on stage."

The direct interaction of the participants with the technology has been explored in this theme that aligns with the "Perceived Ease of Use" (PEoU) component of TAM. Participant 1 noted, "The red and green colors were the most helpful part, honestly… It's simple - even my younger brother could understand it." Participant 2 added, "My little sister watched me once and said, 'Bhai, it's like a game!' That's exactly what it feels like - an educational game." However, the initial learning curve was steep for some. Participant 5 admitted, "It was difficult at first because I had never used a smartphone before. The scroll, the tap, the microphone permission - everything was new." Participant 7 shared, "I'm not very tech-friendly, so the first week was frustrating. But my younger sister helped me."

In this theme, the most significant psychological advantage of using applications like ELSA Speak, the creation of a private and non-judgmental learning environment, is being explored. Participant 1 stated, "No fear of 60 classmates listening to my mistakes. No one laughing. Just me and my phone." Participant 5 reflected, "Some boys laugh. They called me 'paindu' (villager) when I said 'development' wrong. With ELSA, no one judges. I can make mistakes privately. That's the biggest difference - dignity." The inexhaustible patience of the ELSA Speak application was a recurrent theme in the interviews. Participant 1 described the app as a "private tutor who never gets tired." Participant 7 noted, "I could repeat one word 50 times without sir getting annoyed. That's impossible in class." Participant 3 emphasized learner autonomy, stating, "With ELSA, there was no front or back bench. Just me learning at my own speed. No comparison with brilliant students, no embarrassment."

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Although the participants acknowledged the benefits of the designated application, the path to accessing these benefits was riddled with numerous challenges. The most frequent challenge, according to the learners, was the poor and unreliable internet connectivity. Participant 1 reported, "In our village, we have Zong 4G, but during rain or wind, the connection drops… there were three days straight where the app wouldn't load at all." Participant 3 shared, "Our village has Telenor network, which is weak inside our house. I had to go to the rooftop to get good signal." Participant 5 had to struggle the most, adding, "I had to travel 2km to my cousin's house to use his Wi-Fi because we don't have internet at home."

Another challenge that is linked to the unreliable infrastructure of the internet is the high cost of internet packages. Participant 2 stated, "The app uses a lot of data - about 2-3 GB per week… My monthly data package is only 8GB." Participant 3 captured the financial strain, saying, "I want to, but my father says we can't afford extra data charges." Furthermore, hardware limitations affected users. Participant 2 noted, "Battery drainage was serious. My old Huawei phone would heat up after 25 minutes and the battery would drop 30-40%." Participant 4 added, "Storage space… My phone is 32GB… it kept showing 'Storage Full.' I had to delete my photos to make space."

A quiet space is necessary for an application that requires its users to speak and listen to correct pronunciation. Participant 7 shared, "The app requires quiet environment, but in our joint family system, there's always noise - TV, siblings, guests. I had to practice under my blanket sometimes!" Participant 1 also noted, "Background noise from my mother's cooking or tractors passing by was a real problem."

Conclusion

In the light of the findings stated above, this study concludes that the mobile phone applications powered by ASR and AI serve as an effective intervention to enhance the speaking skills of intermediate-level ELLs in the Mianwali district. The positive impact of this technology on students' weaker speaking skills in the English language is seen through the statistically significant improvements in pronunciation and fluency of the EG, and these improvements are also complemented by the reduction of anxiety and the growth in learners' autonomy. The ELSA Speak app, which was designated as the intervention, was effective at the reduction of fossilized errors in segmental and suprasegmental aspects of pronunciation, which were neglected by the GTM method commonly used in Pakistan. The positive influence of ASR technology is not limited to the enhancement of the speaking skills of individual students but also to empowering them even in overcrowded classrooms.

Although the positive influence of the ASR technology on learners' speaking skills has been demonstrated, this study also reveals that technological solutions further amplify the existing inequalities if they are deployed without solving the problems faced by the students. The socio-technical challenges, such as the unavailability or recurrent internet disconnection, expensive mobile data packages, inadequate smartphone hardware, and the joint family system, where finding quiet space for practicing speaking English is difficult, reduce the adoption of these ASR-based interventions. The findings also caution against the techno-optimism. It is established in this study that the tools like ELSA Speak are very effective in correcting the mispronunciations of ELLs, it cannot be a substitution for the contextually rich feedback provided by the teachers on communicative appropriateness. To conclude, the AI-powered ASR applications like ELSA Speak serve as a scalable and evidence-based solution to address the speaking skills crisis in the Pakistani ESL context for the language learners who are preparing for higher education and professional life.

References

Akman, E., & Karahan, P. (2023). ELT students' perceptions toward mobile-assisted language learning (MALL): Exploring its effects on motivation and learner autonomy. *International Journal of Educational Researchers (IJERs)*, 14(2), 1-20.

Anggraini, A. (2022). Improving students' pronunciation skill using ELSA Speak application. *Journey*, 5(1), 135-141.

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.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL quarterly*, 39(3), 379-397.

Farhat, P. A., & Dzakiria, H. (2017). Pronunciation barriers and computer assisted language learning (CALL): Coping the demands of 21st century in second language learning classroom in Pakistan.

Fatima, N., Firdaus, A., & Saleem, Z. (2025). Leveraging AI-powered chatbots to enhance English speaking skills among rural Pakistani learners: A socio-linguistics and pedagogical evaluation. *Policy Journal of Social Science Review, 3*(5), 289–301.

Gusrianto, E. (2025, March). Exploring learning experience with ELSA Speak for independent learning: A case study. In *International Conference on Advances in Humanities, Education and Language (ICEL 2024)* (pp. 219-229). Atlantis Press.

Jiang, M. Y. C., Jong, M. S. Y., Wu, N., Shen, B., Chai, C. S., Lau, W. W. F., & Huang, B. (2022). Integrating automatic speech recognition technology into vocabulary learning in a flipped English class for Chinese college students. *Frontiers in Psychology*, 13, 902429.

Metruk, R. (2021). The use of smartphone English language learning apps in the process of learning English: Slovak EFL students' perspectives. *Sustainability*, 13(15), 8205.

Moulieswaran, N., & NS, P. K. (2023). Investigating ESL learners perception and problem towards artificial intelligence (AI)-assisted English language learning and teaching. *World Journal of English Language*, 13(5), 290-290.

Riaz, Y., & Kausar, G. (2024). Improving Pakistani college students' speaking skills using AI-driven linguistic input: An experimental study. *Journal of Asian Development Studies, 13*(4), 853–870.

Shaheen, R., Soomro, A. R., & Ali, H. (2024). Effect of mobile assisted language learning (MALL) attitude and practices in university students. *Journal of Asian Development Studies*, 13(2), 101-113.

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in psychology*, 14, 1210187.

Thomas, A., Piquette, C., & McMaster, D. (2016). English communication skills for employability: The perspectives of employers in Bahrain. *Learning and Teaching in Higher Education: Gulf Perspectives*, 13(1), 36-52.

Usman, T., Kharal, Q. A., Arsalan, M., & Riasat, A. (2025). Predominance of Lecture Method as a Barrier to Learners' Autonomy: A Survey of Pakistani Universities in the Punjab Province. Journal of Arts and Linguistics Studies, 3(3), 3541–3563.

Yaqub, S. (2009, March 16). *ELT in Pakistan* [Blog post]. English Language Learning Forum. https://englishlanguagelearningforum.blogspot.com/2009/03/elt-in-pakistan.html

Zohaib, Naveed, S., & Shah, S. H. R. (2025). Effectiveness of WhatsApp as a tool to enhance English-speaking skills of university students. *International Journal of Academic Research for Humanities*, *5*(1), 43–53. https://jar.bwo-researches.com/index.php/jarh/article/view/544