Article

Check for updates OPEN ACCESS

Journal of International Students
Volume 16, Issue 19 (2026), pp. 1-17
ISSN: 2162-3104 (Print), 2166-3750 (Online)
jistudents.org
DOI: https://doi.org/10.32674/es0wvt92

AI-mediated speaking practice in less-resourced EFL contexts: Effects on fluency and vocabulary

Saria Mazhar [1], **Dr. Bushra Sani** [2], **Dr. Muhammad Adnan Sial** [3],

[1] University of Education, Pakistan; sariamazhar890@gmail.com
[2] University of Education, Pakistan; bushra.sani@ue.edu.pk
[3] University of Education, Pakistan; adnan.sial@ue.edu.pk
* Correspondence: bushra.sani@ue.edu.pk

ABSTRACT

AI-based automation is reshaping the way people learn, work, and communicate. This study investigates the effectiveness of AI tools (ChatGPT, Call Annie, Grok, and Gemini) in developing English as a foreign language speaking skills, such as fluency and vocabulary, among learners in less-resourced contexts. The study also identifies participants' perceptions of the challenges and advantages faced during IELTS preparation. Guided by Vygotsky's sociocultural theory, the present research used a mixed-methods quasi-experimental approach. The study included two groups of 30 participants each: the experimental group engaged in AI-assisted speaking through one-to-one voice sessions three times a week for 12 weeks, while the control group practiced traditional speaking methods. We collected data through IELTS-formatted pre- and post-speaking assessments and semi-structured interviews. The findings revealed that the experimental group experienced moderate but statistically significant improvements in speech rhythm and lexical resource enrichment. Participants also perceived AI tools as accessible, low-anxiety speaking partners. The findings have implications for equitable language education in resource-limited contexts, as well as for the UNESCO 2030 agenda for inclusive digital learning.

Keywords: artificial intelligence, English as a foreign language, speaking skills, less-resourced contexts, fluency, vocabulary

© Author(s), 2026. Published by Star Scholars Press.

This article is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited. https://creativecommons.org/licenses/by/4.0/

INTRODUCTION

With the potential to offer personalized, accessible, and scalable learning experiences, artificial intelligence (AI) is quickly reshaping the landscape of language education (Holmes & Tuomi, 2022; Jeon et al., 2024). In this context, AI-powered conversational tools may serve as useful assistants for developing speaking skills, supporting structured communication, providing instant feedback, and offering nonthreatening speaking practice contexts. For nonnative English speakers, speaking is among the most challenging skills to master because, beyond language, they also need opportunities to communicate with others in a realistic and natural way, which is difficult to provide in a formal classroom environment (Morovat, 2014; Santoso et al., 2024).

These issues are particularly acute in less-resourced contexts where access to qualified teachers is limited, digital infrastructure is less developed, and financial constraints and opportunities to use authentic English are fewer (Curtis, 2021; Singh, 2024). In such settings (i.e., rural and semiurban Pakistan), students are exposed to a lack of native English speakers, limited contact with English as a language of instruction, and overcrowding in classrooms, which further hinders students' ability to develop oral proficiency (Bhutto et al., 2023; Majeed & Kareem, 2025). While interest in AI in L2 language learning has increased worldwide, empirical evidence on its effectiveness under these limited conditions remains limited.

The present study is situated in the broader context of international education, with a particular focus on the experiences of English as a foreign language (EFL) learners who struggle with language and institutional challenges that affect their educational path and drive them to move across borders. In this context, AI tools have special potential because they are widely accessible, not limited to classroom hours, and offer one-to-one, interaction-based practice that is often unavailable in low-resource schools (Huang & Zou, 2024; Yuan & Liu, 2025). This turns AI not only into a teaching tool but also into a potential equalizer for language-learning opportunities.

In this study, the terms “more knowledgeable other” (MKO) and “scaffold” are used to understand AI-mediated interactions as scaffolding tools for learners in their zone of proximal development (ZPD) (Vygotsky, 1978). The analysis focuses on two measurable subskills of speaking: fluency and vocabulary. To achieve this goal, four tools (ChatGPT, Gemini, Grok, and Call Annie) with one-to-one voice communication capabilities were chosen. This research aims to fill a significant gap in the literature, as no empirical research has examined the use of AI in oral practice in a less-resourced EFL context through an intervention approach. The research objectives are twofold: First, this study aims to investigate the impact of the voice characteristics of AI-powered chatbots on the acquisition of EFL speaking skills (particularly fluency and vocabulary) by learners from resource-poor settings. Second, it is intended to investigate participants' perceptions of the benefits and barriers of speaking practices with AI in the experimental group.

LITERATURE REVIEW

Development of AI-Powered Tools in EFL Speaking

Research on AI integration in language education has become increasingly prominent over the past decade. A chatbot refers to an artificial intelligence designed to replicate natural, human-like interactions in text and/or voice, enabling timely responses and effective support and tutoring (Kerlyl et al., 2007). Recent developments in large language models (LLMs) have significantly improved the ability of AI-powered chatbots to engage in context-sensitive and meaningful conversations, making them more suitable as speaking practice partners (Fryer et al., 2019; Jeon et al., 2024). Research on the use of speech-recognition chatbots in language learning indicates that such digital agents may help develop students' oral proficiency through repeated, nonthreatening interactive exposures that are difficult to achieve in the classroom setting (Jeon et al., 2023).

Most existing research has focused on writing and/or oral proficiency development in well-resourced institutional contexts (Singh, 2024), but empirical research on AI-supported speaking practices in less-resourced EFL contexts remains limited. The present study aims to fill this gap by highlighting the measurable subskills of speaking, namely, fluency and vocabulary, among EFL learners in under-resourced contexts.

Theoretical Framework

This study draws on three complementary language acquisition theories: Krashen's input, Long's interaction, and Swain's output hypothesis. Vygotsky's Sociocultural Theory lays the theoretical foundation for better comprehending AI-assisted language learning processes. Krashen's input hypothesis (1982) proposes that language acquisition occurs when learners are exposed to comprehensible input that propels their language competence slightly beyond their next developmental level (i + 1). This view is extended by Long's (1983, 1996) interaction hypothesis, which suggests that language development progresses through real communicative exchanges. The interactional patterns help the learners learn both receptive and productive skills together, reconfirming the communicative basis of learning. Krashen emphasized the importance of language exposure, but Long emphasized that meaningful communicative interaction is also essential for language acquisition, which has implications for the development of AI-assisted speaking tasks (Lichtman & VanPatten, 2021).

Building on prior input and interaction theories, Swain's Output Hypothesis (1995) highlights the significant role of language production as a mental engagement process that serves as a cognitive mechanism in second-language development (Poehner & Lantolf, 2024). It prompts learners to internalize different aspects of language, such as vocabulary and grammar, and to refine gaps in their knowledge through language production. Swain proposed "Pushed Output" as circumstances where learners are inspired to produce thoughts in communication, enabling them to operate beyond their existing communicative proficiency. This implies that students should be forced to speak; however, in a comfortable environment, they gradually become confident. His notion exemplifies AI-based speaking applications as an instant evaluation mechanism that offers awareness of linguistic forms and performance (Ismail et al., 2022). Such immediate feedback supports noticing crucial metacognitive acts that help learners become aware of linguistic discrepancies in their actual language use.

Overall, these three theories establish a coherent rationale for AI-mediated oral practice. AI-driven speaking platforms provide personalized speech activities and responsiveness evaluation in real time, are calibrated to students' capacity, facilitate meaningful interaction, and generate pushed output opportunities (Litchman & VanPatten, 2021). Vygotsky's theory of the Zone of Proximal Development (ZPD)—the concept of the zone where a learner can learn with the help of the more knowledgeable other (MKO)—is a perfect fit for the use of AI conversational tools in the context of the present study. AI acts as an ongoing MKO, providing scaffolded prompts, constructive and timely feedback, and levels of challenge within each learner's ZPD (Lantolf & Poehner, 2014; Moulavinafchi et al., 2026).

Less-Resourced Contexts and EFL Speaking Challenges

Less-resourced educational contexts are defined by a combination of economic, educational, and technological limitations that limit learners' opportunities for meaningful language development (Curtis, 2021). The settings are limited not only financially but also by the lack of qualified teachers, digital infrastructure, overcrowded classrooms, and authentic English-speaking opportunities (Singh, 2024; Majeed & Kareem, 2025). These obstacles are exacerbated in rural and semiurban Pakistan because of fewer instructional hours, limited professional development opportunities for teachers, and limited access to digital learning platforms (Rashid et al., 2018; Bhutto et al., 2023).

The implications for the development of spoken language skills are significant. In such contexts, learners often fail to develop high levels of achievement because their speaking level is limited to using rote memorization methods, which are not appropriate for spontaneous speech (Alessa & Hussein, 2023).

Additionally, the psychological aspects of speaking L2, such as L2 speaking anxiety, L2 low self-confidence, and L2 fear of negative evaluation, are exacerbated in the context of a lack of safe space for L2 oral experimentation (Xing & Bolden, 2019; Hao, 2026; Rattanayat & Boonsuk, 2026).

AI-powered conversational tools can offer contextually relevant answers to these problems. They are readily accessible in nonacademic settings, have low or no cost, and provide personalized, nonjudgmental interactions, which are particularly suitable in less-resourced environments (Yuan & Liu, 2025; Qin & Derakhshan, 2026). AI tools can tailor their content to the learner, thus engaging in a differentiated, scaffolded practice that is impossible in a one-size-fits-all traditional classroom because of structural constraints (Huang & Zou, 2024).

Research Gap

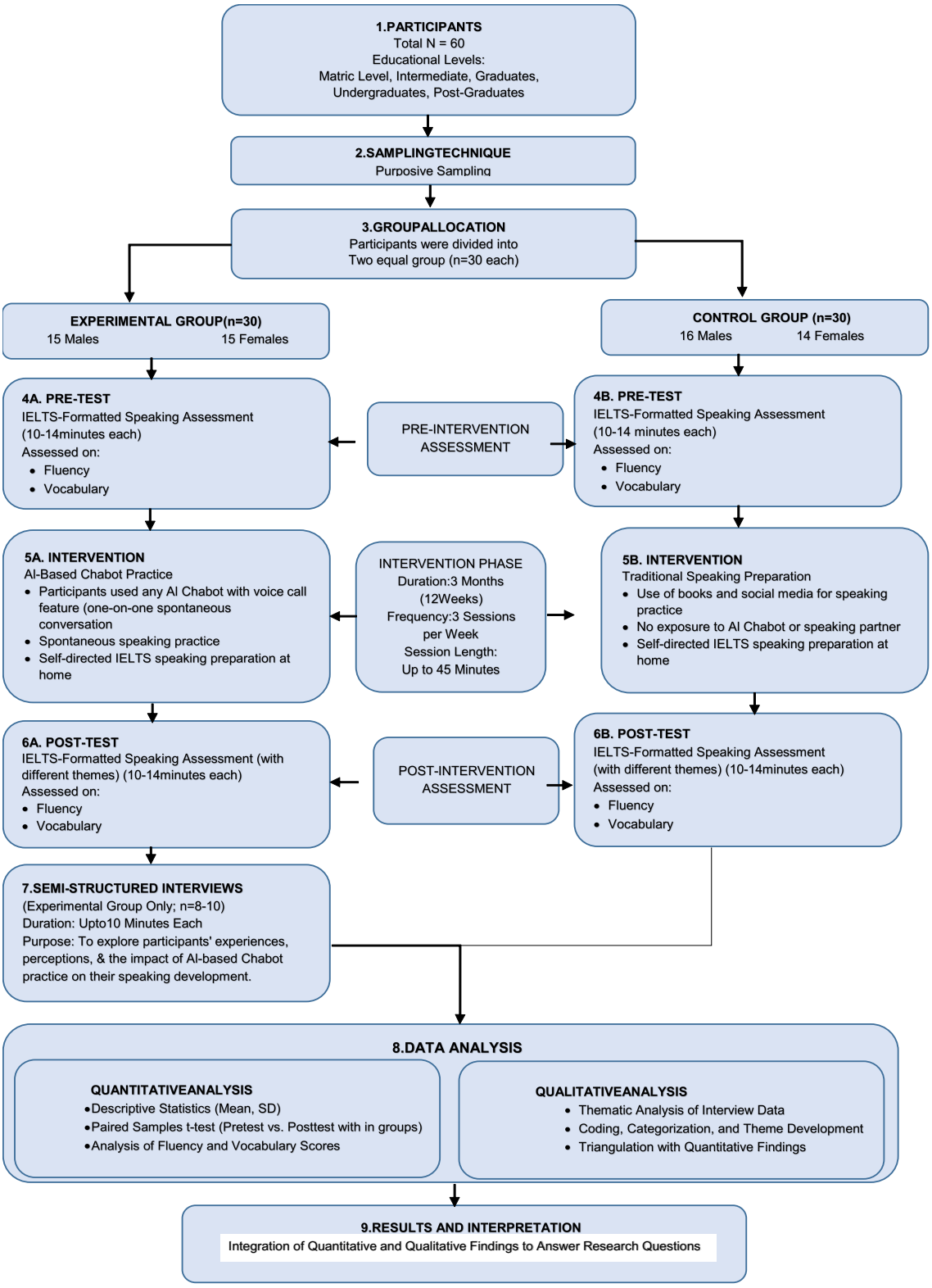
Although AI is being increasingly integrated into language instruction, few empirical studies focus on its effectiveness in promoting oral proficiency in language learning within less-resourced EFL contexts, particularly when quasiexperimental designs are used. Existing studies have focused on writing skills, technology-rich institutional settings, or attitudinal outcomes. Studies that measure concrete gains in speaking subskills among linguistically marginalized learner populations in the context of intervention are needed. The current study responds directly to this gap by identifying the effectiveness of AI-based interactive practices on EFL learners' oral proficiency, precision fluency and vocabulary development, in less-resourced settings.

RESEARCH METHOD

Research Design

This study employs a mixed-method, quasi-experimental design to explore the influence of AI-mediated speaking practices on EFL learners' fluency and vocabulary. A quasi-experimental design was suitable because random assignment was not feasible in the study's naturalistic educational setting; however, the pretest-posttest structure with parallel control and experimental groups provides a credible basis for examining intervention effects (Zargaran, 2025). The quantitative strand assesses increases in pre- and post-intervention speaking subskills, and the qualitative strand investigates participants' lived experiences and perceptions of using AI tools. A general overview of the research design and procedure is shown in Figure 1.

Figure 1



Research Methodology Overview: Quasi-Experimental Mixed-Methods Design

Research Context and Participants

Contexts with scarce academic materials commonly mark combinations of barriers that significantly hinder candidates' meaningful opportunities to develop oral language development (Anwar & Ihsan, 2016). This study was conducted in less-resourced settings, where EFL learners face well-documented barriers, such as a shortage of qualified teachers, poor digital infrastructure, overcrowded classrooms, and a lack of authentic English-speaking opportunities outside the classroom (Rashid et al., 2018; Abbasi et al., 2021). Purposive sampling (N = 60; Control = 30 (Male = 14 & Female = 16), and Experimental = 30 (Male = 15 & Female = 15)) was used to select participants in line with a specific learner profile: EFL candidates with backgrounds from underresourced areas actively preparing for the IELTS speaking examination. The participants included diverse levels of education in keeping with the diverse backgrounds of IELTS candidates from developing countries, with a particular focus on Pakistan.

Intervention Procedures

The experimental group participated in self-directed AI-supported speaking training three times per week for up to 45 minutes in each session over 12 weeks. The participants had the option to choose from four AI tools: ChatGPT, Gemini, Call Annie, and Grok. The only condition was that each tool's one-to-one voice call feature be used to participate in spontaneous spoken communication rather than in a text-based exchange. The condition was incorporated to ensure that practice would not be based on written exchanges but on oral production. These four tools were chosen because they are accessible, available at little or no cost, and can be used with voice calls, making them suitable for lower-resource settings.

To ensure reliability, a structured monitoring system was employed throughout the intervention. Each participant in the experimental group had a set time and day each week to practice. A special WhatsApp group was also created, and the group members sent screenshots of their interactions with the AI as proof of their speaking sessions. The research team was responsible for the participants checking in and out the participants of the session. Additionally, the duration of the session and the frequency of the participants' contributions were systematically recorded. The research team was available throughout the intervention for guidance and troubleshooting and to keep participants motivated. Researchers conducted visits at regular intervals to monitor progress and compliance with the research intervention protocol. To maintain the validity of the results, any dropouts that occurred during the pilot phase were not included in the main study.

To maintain methodological consistency between the two groups, the same monitoring procedures were used for the control group. Before the intervention, participants in the control group were explicitly asked to avoid using any AI-based conversational tools for speaking during the 12 weeks, and this was formally affirmed by all the participants. The control group participants were also given fixed training days and times and were asked to prepare their own speaking tasks using books, social media, and self-directed IELTS speaking practices. They were not exposed to AI speaking partners or conversational agents. The control group was checked by periodically using Zoom meetings to track adherence to the conventional practice protocol, to discuss any concerns about the protocol and to keep participants engaged during the intervention. No dropouts were observed during the main study for either group. As stated earlier, all the dropouts were identified and excluded during the pilot phase from both the experimental and control groups. The parallel monitoring structure allowed for the between-group comparison to be an actual difference in practice modality, not uncontrolled differences in participant practice or engagement.

Research Instruments

Two research instruments were employed to capture both quantitative linguistic gains and qualitative experiential insights.

IELTS-Formatted Speaking Assessment

The Cambridge band descriptor chart was employed in this research as both a pretest and a posttest instrument and was administered to all 60 participants before and after the intervention period. To ensure

authenticity and standardization, the subject matter of the IELTS assessment was chosen from the British Council's official preparation resources. Each assessment session lasted for 10–14 minutes. Tests were scored using the Cambridge IELTS band descriptor chart with particular attention to fluency and vocabulary. The final band score for each participant was the average score of the two independent rating participants. A pilot assessment was carried out with a subsample of 10 candidates to test the instrument before the main study and to fine-tune the procedures for administering the instrument.

Semi-Structured Interviews

Posttesting was followed by in-depth interviews with ten participants from the experimental group. This step was necessary to gain qualitative insights into their experience with AI-mediated speaking practices. The duration of each interview was approximately ten minutes. The questions were kept open-ended and were about the perceived advantages, obstacles, and general feedback that participants had regarding the use of AI tools in enhancing their speaking skills.

Ethical Considerations

Because this study was part of a larger research project, it followed standard ethical guidelines for research with human participants. The host university's ethics committee formally approved the study. Before participating, participants were briefed on the study objectives and procedures, data collection methods, and their right to withdraw consent at any time without adverse consequences. The researchers then obtained written consent from each participant. The researchers also respected participants' privacy, and all information was anonymized and kept secure.

RESULTS

As shown in Table 1, the experimental group's mean score for fluency improved from 5.01 to 5.90 (a gain of 0.89 band points), whereas the mean score for vocabulary improved from 4.93 to 5.92 (a gain of 0.99 band points). Notably, the SD for vocabulary widened considerably from 0.10 to 0.90, reflecting substantial individual variation in vocabulary increase, likely attributable to differences in autonomous engagement, proficiency level, and tool selection. The fluency SD narrowed slightly from 1.00 to 0.90, indicating more consistent fluency gains across participants.

Table 1

Descriptive Statistics for the Experimental and Control Groups

Subskills	No. of participants	Experimental group			Control group		
		Mean	SD	Skewness	Mean	SD	Skewness
Prefluency	30	5.01	.99	-0.13	5.55	0.57	0.35
Post-Fluency	30	5.7	0.94	-0.14	4.1	0.51	0.3
Pre-Vocabulary	30	4.93	0.1	-0.22	5.5	0.57	-0.11
Post-Vocabulary	30	5.92	0.9	0.04	6.36	0.57	-0.39
0 = does not attend				9 = Expert user			

For the control group, the mean score for fluency decreased from 5.55 to 4.10, indicating a decrease of 1.45 band points, whereas the mean value for vocabulary increased from 5.50 to 6.36 (an increase of 0.86 band points). The narrowing of the fluency SD from 0.60 to 0.50 indicates that this decline was uniform across participants rather than driven by individual outliers, indicating a systemic

effect of conventional preparation methods on spontaneous oral production. The control group's vocabulary gain is comparable in magnitude to that of the experimental group but is best understood as surface-level lexical retention through rote memorization rather than communicatively embedded acquisition.

As shown in Table 1, the skewness values before and after treatment for both groups (control and experimental) remained within the range of -1 and +1, indicating that the data fulfilled the assumption of normality (Hair et al., 2019).

Within-Group Analysis: Paired Samples T Tests

Referring to Table 2, the paired samples t test revealed statistically significant improvements in both fluency ($t = 12.3$ ($df = 29$), $p < .001$, $d = 0.96$) and vocabulary ($t = 12.5$ ($df = 29$), $p < .001$, $d = 0.96$) for the experimental group. Both effect sizes fall within the large range according to Cohen's (1988) conventions ($d \geq 0.80$), indicating that the intervention produced not only statistically significant but also practically meaningful gains in both speaking subskills.

Table 2

Paired Sample t test: Experimental Group

Subskill	Time	Mean	SD	t	df	P	Effect size (d)
Fluency	Pretest	5.01	1	12.3	29	0	0.96
	Posttest	5.9	0.9			0	
Vocabulary	Pretest	4.93	1.1	12.5	29	0	0.96
	Posttest	5.91	0.9			0	

Interrater Reliability

Interrater reliability analysis was conducted to check the reliability of the speaking assessments before the substantive findings were presented. Two independent raters evaluated all the pre- and posttest audio recordings using the Cambridge IELTS band descriptor chart. Intraclass correlation coefficients (ICCs) were calculated using a two-way mixed-effects model in SPSS.

Table 3

Interrater Reliability

Variable	ICC (Average)	95% CI
Pre-Fluency	.985	.967 – .993
Post-Fluency	.992	.987 – .995
Pre-Vocabulary	.990	.982 – .994
Post-Vocabulary	.988	.980 – .993

The data in Table 3 indicate that the interclass reliability of each variable with respect to time is high, indicating consistency throughout the assessment of audio recordings (Koo, 2016). Interrater

reliability was evaluated through the interclass correlation coefficient (ICC), which indicates whether the values given by the raters are consistent or not and was measured by a two-way mixed method.

Table 4
Multivariate tests

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Fluency[b]							
Time	Pillai's Trace	.444	46.225[a]	1	58	.000	.444
	Wilk's Lambda	.556	46.225[a]	1	58	.000	.444
	Hotelling's Trace	.797	46.225[a]	1	58	.000	.444
	Roy's Largest Root	.797	46.225[a]	1	58	.000	.444
Time * Group	Pillai's Trace	.881	4.29 [a]	1	58	.000	.881
	Wilk's Lambda	.119	4.29 [a]	1	58	.000	.881
	Hotelling's Trace	7.397	4.29 [a]	1	58	.000	.881
	Roy's Largest Root	7.397	4.29 [a]	1	58	.000	.881
Vocabulay[b]							
Time	Pillai's Trace	.862	3.625[a]	1	58	.000	.862
	Wilk's Lambda	.138	3.625[a]	1	58	.000	.862
	Hotelling's Trace	6.25	3.625[a]	1	58	.000	.862
	Roy's Largest Root	6.25	3.625[a]	1	58	.000	.862
Time * Group	Pillai's Trace	.032	1.917 [a]	1	58	.171	.032
	Wilk's Lambda	.968	1.917 [a]	1	58	.171	.032
	Hotelling's Trace	.033	1.917 [a]	1	58	.171	.032
	Roy's Largest Root	.033	1.917 [a]	1	58	.171	.032

In light of the guidelines given by Hair et al. (2019), as shown in Table 4, a repeated-measures multivariate analysis using a general linear model (GLM) was conducted to examine the effects of time and group on students' speaking subskills. The multivariate tests indicated a significant main effect of time, Wilks' $\Lambda = .556$, $F(1, 58) = 46.225$, $p < .001$, partial $\eta^2 = .444$, suggesting significant differences between the pretest and posttest scores for fluency.

Moreover, the interaction effect between time and group was statistically significant, Wilks' $\Lambda = .119$, $F(1, 58) = 4.29$, $p < .001$, partial $\eta^2 = .881$. These results indicate that the experimental and control groups for fluency differed significantly in their performance changes over time, demonstrating the effectiveness of the intervention. Multivariate tests revealed a statistically significant main effect of time, Wilks' $\Lambda = .138$, $F(1, 58) = 3.625$, $p < .001$, partial $\eta^2 = .862$, indicating significant differences between

the pretest and posttest scores for vocabulary. However, the interaction effect between time and group was not statistically significant, Wilks' $\Lambda = .968$, $F(1, 58) = 1.917$, $p = .171$, partial $\eta^2 = .032$. This suggests that the changes over time were similar for both the experimental and control groups, indicating that the intervention did not have a statistically significant differential effect on participants' performance (vocabulary).

Qualitative Findings: Thematic Analysis

To answer RQ2, semi-structured interviews were administered after the post-intervention assessment with ten participants from the experimental group. The interviews were audio-recorded, transcribed, and analyzed through an inductive thematic process consisting of open coding, categorization, and theme development. The findings were then triangulated with the quantitative results. Six themes emerged, each with multiple dimensions reflecting participants' perceptions and experiences of AI-mediated speaking practices.

Theme 1: Spoken Safe Setting

Participants spontaneously described feeling psychologically safe when interacting with the AI, with no judgment or negative evaluation. This perception was especially strong among participants with lower proficiency and beginners, who had experienced being embarrassed in public as one of the main challenges to practicing speaking English. A participant said that he felt free to make mistakes, correct himself, and play around with language without facing social repercussions, which is not easily found in formal classroom situations and informal social interactions in less-resourced contexts. The fluency gains for the experimental group in this theme are explained qualitatively through the reduction of the affective filter, meaning that AI tools created conditions for making spontaneous oral production accessible for learners, which had not been accessible before.

Theme 2: Affordable-Speaking Partnership

One key advantage in speaking preparation was the financial accessibility of AI tools. Several experiences were described about attending preparation classes for IELTS of 50 or more students, where no attention was given to individual speaking needs. Others said that no one truly spoke English in their social surroundings. Considering this, AI tools, particularly ChatGPT, were seen as truly fair conversational partners. This finding confirms the central argument of the study that AI tools can work as educational equalizers in less-resourced settings. AI tools can offer personalized and extended spoken practice, which is generally not available in financial or social terms.

Theme 3: Structured Spoken Format

Several respondents stated that they used carefully designed prompts to guide the AI tools to IELTS-specific oral formats, including DREC (Description, Reason, Example, Conclusion). The participants' use of the same model response allowed them to learn different words, develop fluency with transitions, and grasp the full learning value of one model response. This theme is a partial qualitative interpretation of the experimental group's vocabulary growth, where compared with memorization-based strategies, structured prompting strategies offered more frequent and more contextually rich experiences with vocabulary.

Theme 4: Autonomous Practice

Many respondents reported that they practiced more than the suggested amount, either by using an additional account to increase the length of time spent daily, by swapping between the applications to lengthen the practice time, or by practicing more in areas that they felt they needed to improve. The increased standard deviation in vocabulary (from 0.10 to 0.90) for the experimental group is the most direct qualitative explanation for this pattern of autonomous engagement. Participants with greater engagement were exposed to the highest volume and complexity of lexical input, resulting in more vocabulary

acquisition. However, respondents who barely met the requirements experienced the least gain in vocabulary.

Theme 5: Articulatory Development and Fluency Automaticity

A few respondents claimed that they gained familiarity and automaticity with the sounds of English after prolonged interaction with AI. One respondent shared that English sounds felt unfamiliar to her own ears before the intervention, implying that she had very little exposure to spoken English in everyday life. The same participant later reported an improvement in instant recall of words and reported that she could speak naturally and conversationally with the help of regular use of AI. This qualitative description validates the automaticity theory of Segalowitz (2010) and provides a plausible neurological and phonological explanation for the quantitative measures of fluency seen in the experimental group.

Theme 6: Technical limitations

A few respondents reported facing technical difficulties with AI tools. The Call Annie users experienced frequent interruptions to their tasks because of subscription advertisements, which made them feel frustrated and lose engagement. In addition, grok users reported that the informal register, which they believed had been biased by the training set from Twitter, was not consistent with the academic vocabulary needs of IELTS and that the customization function did not function as expected. A few participants also noted that AI tools need a stable internet connection, which is not always available in less-resourced settings where internet connectivity is not guaranteed. Some participants experienced technical challenges that tempered the benefits of AI-assisted practice, which helps account for the variation in results.

Tool-Specific Observations

There were significant differences in the experiences and outcomes across the four sets of AI tools used in the study. Each of the tools is summarized below in terms of characteristics observed by participants and confirmed by qualitative analysis. One hour of one-to-one free speech interaction a day (n = 16), a friendly and encouraging tone, and a wide vocabulary from a variety of data on the internet made ChatGPT well suited for participants with all skill levels. Despite its time limit (approximately 10–15 minutes per session), Gemini (n = 6) was appreciated for its formally diverse vocabulary and feedback, but its time limit made this process challenging. Furthermore, participants deemed Grok inappropriate for formal IELTS preparation in every scenario, pointing to the informal register, the speed and irregularity of the speech output, and the lack of a functional customization feature. These results contradict previous research by Preciado (2025) on the effectiveness of Grok's customization feature, suggesting that regional technical constraints or version differences may have affected participants' experience. Call Annie (n = 3) was evaluated the least, with a few technical disruptions and poor error feedback, and the sample size was not large enough for generalization.

DISCUSSION AND CONCLUSIONS

This study investigated the efficiency of conversational AI tools for improving the EFL fluency and vocabulary of less-resourced learners and the perceptions of participants regarding AI-supported speaking practices. The discussion of each research question is presented sequentially, followed by a synthesis of quantitative and qualitative results in an interpretive argument.

The quantitative results indicate that there was a statistically significant increase in fluency and vocabulary in the experimental group after the 12-week AI-mediated intervention. Fluency improved from a pretest mean of 5.01 to a posttest mean of 5.90 ($t(29) = 12.3$, $p < .001$, $d = 0.96$), whereas vocabulary improved from 4.93 to 5.92 ($t(29) = 12.5$, $p < .001$, $d = 0.96$). Both effect sizes are in the large range as defined by Cohen (1988), which means that the intervention was not only statistically significant but also practically significant in both subskills. The control group's fluency, on the other hand, significantly decreased, dropping from a pretest mean of 5.55 to a posttest mean of 4.10; however, there was a slight

increase in vocabulary from a pretest mean of 5.50 to a posttest mean of 6.36. The conclusions are generally congruent with those of previous studies showing that AI-supported interactions have a positive effect on EFL speaking development (Huang & Zou, 2024; Jeon et al., 2024) and are broadened to less-resourced contexts, where there are relatively few empirical studies.

A further detailed analysis of the descriptive data indicates additional nuances that go beyond the simple mean score. The experimental group showed a significant increase in the standard deviation of vocabulary from 0.10 at pretest to 0.90 at posttest, suggesting that there was significant variability in vocabulary gains among students after the intervention. This pattern indicates that the impact of using AI tools on vocabulary acquisition was not consistent among all the students but depended on individual factors such as the level of engagement, type of AI tool, independent use, and proficiency. In contrast, the standard deviation for fluency in the AI-mediated practice group was slightly smaller (1.00 to 0.90), indicating that there was less variation in fluency gains among participants in this group, presumably because the tools differed in their spontaneous speech components, but all of them offered a certain amount of spontaneous spoken interaction, which was completely absent in the other group.

Theoretically, this discrepancy in fluency and vocabulary achievement is noteworthy. Fluency development is seen as a function of automaticity—the gradual reduction in cognitive effort in the production of speech as a result of repeated practice (Lichtman & VanPatten, 2021). The group was homogeneous in its conditions for automaticity development, with all participants interacting regularly with AI tools, which resulted in relatively homogeneous fluency gains across the group. On the other hand, the depth, variety, and richness of language exposure are more important for vocabulary acquisition (Swain & Lapkin, 1995). Participants who used the most (more accounts, more switching between tools, and more use of a structured DREC format prompting) received increasingly diverse lexical inputs, resulting in greater vocabulary gains and a broad postintervention standard deviation.

The results from the control group should be interpreted with caution. The decrease in fluency from 5.55 to 4.10 (-1.45 band points) is significant, and there is a consistency of loss across the groups, as evidenced by the reduction in the standard deviation from 0.60 to 0.50. This pattern highlights the significance of active and regular oral practices to improve fluency as a productive language skill. Furthermore, the conventional preparation methods employed by the control group were not sufficient to maintain, and even more so, benefit their spontaneous speech production (Bhutto et al., 2023; Majeed & Kareem, 2025). This result is consistent with Swain's Output Hypothesis (1995) that productive language use is a prerequisite for attaining oral proficiency, not passive exposure.

The vocabulary gain of 0.86 band points for the control group is similar in size to the experimental group's vocabulary gain of 0.99 but of a different kind. Rote memorization strategies were employed mainly by the control group, which lacked the use of spontaneous communicative strategies to embed vocabulary (Alessa & Hussein, 2023). The standard deviation across time (0.60) is fairly stable, indicating the uniformity of this memorization-based strategy. The experimental group’s gains in vocabulary, although less consistent, were contained in the authentic spoken interaction, supporting more profound and lasting acquisition of vocabulary, which is consistent with Long’s (1983, 1996) interaction hypothesis.

Semistructured interviews with the experimental group resulted in six overlapping themes that together provide an understanding of the mechanisms of the quantitative findings: safe spoken setting, affordable speaking partnership, structured speaking format, autonomous practice, articulatory development, and technical limitations.

The most consistent reporting was the sense of psychological safety that came from the interaction with AI. Participants shared experiences of practicing without worrying about being ridiculed, humiliated, or judged negatively—a situation that is much less likely to occur in formal classroom settings or informal social environments in less-resourced contexts. These results confirm Krashen's affective filter hypothesis, which holds that anxiety is an impediment to language acquisition when there is input present (Lichtman & VanPatten, 2021). One participant reported feeling real physical symptoms of anxiety during speaking practice and later went on to report feeling natural and automatic with interaction with the AI. This

qualitative evidence provides a plausible mechanism for the increase in fluency in the experimental group, as it was found that AI tools allowed learners who were not given the opportunity to use spontaneous oral activity to produce naturally (Xing & Bolden, 2019; Hao, 2026). This theme also provides the most persuasive explanation for the decrease in fluency in the control group. The control group was not provided with any space for practicing in a psychologically safe environment or with an interactive speaking partner to enable participants to practice the spoken automaticity that fluency requires.

The financial accessibility of AI tools was continually emphasized as a game changer in the speaking preparation process. Some have described earlier experiences of attending overcrowded IELTS preparation institutes where they felt that their individual speaking needs were not met, and others have mentioned that interlocutors in their social environment spoke in home languages, thus denying them real exposure to English. In this context, AI tools were fair and friendly speaking partners. In particular, ChatGPT was available without any monetary cost and during nonformal hours. This insight highlights the main argument of this study that AI can act as an equalizer in less-resourced educational environments, which contributes to the UNESCO 2030 agenda on inclusive and equitable digital learning (Yuan & Liu, 2025; Duke, 2026).

As stated earlier, participants followed the DREC format to give specific prompts to AI tools to ensure that they generated structured, IELTS-focused spoken responses. One respondent reported that during IELTS preparation, a single-model response by AI simultaneously increased his vocabulary because of lexical variation, as well as fluency, by incorporating transition words. He further elaborated that the clarity of the structure was reinforced because of the seamless integration of these elements. A single prompted interaction stimulated three-dimensional learning. This finding is consistent with Swain's (1995) notion of "pushed output," which involved the need to use AI tools to generate structured, academically appropriate output; the participants thus received richer lexical input than they would have managed on their own and had to produce more sophisticated output than they could come up with on their own.

A significant number of participants indicated that they practiced more than the recommended amount and set up extra accounts to extend practice periods or switch between applications to increase interactions. The most direct explanation for the increasing standard deviation of the vocabulary (from 0.10 to 0.90) is this autonomous engagement pattern. Those who had the most contact with the AI tools received the most varied and detailed vocabulary input and acquired the most vocabulary. The results indicate that interventions promoting self-directed engagement may be more effective and lasting in achieving positive learning outcomes, strengthening the value of AI interaction in language learning (Huang & Zou, 2024).

Some participants reported experiencing a sense of familiarity and automaticity with English sounds as a result of prolonged engagement with AI. One participant described the feeling of unfamiliarity with the English sounds before the intervention and how there was little exposure to spoken English in their surroundings. After regular AI-assisted practice, the same player reported that he or she could speak freely and naturally, with a greater ability to recall vocabulary in real-time conversation. This qualitative description follows the automaticity theory, which states that fluency develops as the cognitive load required for speech production gradually decreases with practice.

The quantitative and qualitative results obtained from this study provide a solid argument for the effective use of AI-based tools through the lens of a Vygotskian-mediated approach within the ZPD of EFL learners in less-resourced settings when the tools used are linguistically suitable, technically stable, and accessible within the constraints of the learner environment. The experimental group's improvements in fluency and vocabulary are due not only to exposure to the input created by AI but also to a full language learning cycle, namely, comprehensible input (Krashen, 1982), meaningful interaction (Long, 1983), pushed output (Swain, 1995), and scaffolded mediation within the ZPD (Vygotsky, 1978), on one free, accessible platform. These improvements varied across individuals depending on the degree of autonomous engagement, choice of tools, and affective readiness, yielding findings of a standard deviation that is much broader than usual. The control group results confirm that normal, self-directed approaches to language learning are not enough for fluency development and can achieve only superficial vocabulary retention,

showing a clear need for an alternative to speaking practice in less-resourced educational settings that is accessible and interactive.

This study provides empirical evidence supporting the adoption of AI-powered conversational tools to achieve considerable improvements in fluency and vocabulary in EFL speaking among students in less-resourced contexts. After 12 weeks of AI-supported practice, the results for the experimental group revealed a large effect size for both fluency ($d = 0.96$) and vocabulary ($d = 0.96$), with six qualitative themes collectively indicating that by solving psychological, financial, and structural challenges, AI can help address drawbacks that traditional instruction cannot address. The results further support Vygotsky's Sociocultural Theory by showing that AI tools can serve as more knowledgeable others (MKOs) in the Zone of Proximal Development (ZPD) of learners and that it is possible to combine the constructs of Krashen's comprehensible input, Long's interactional hypothesis, and Swain's pushed output in a single AI-supported platform.

IMPLICATIONS

The results have implications for practitioners, policy makers, and developers, and they indicate that free AI conversational tools, such as ChatGPT and Gemini, are good, scalable, and cost-efficient supplements to formal EFL teaching under infrastructure conditions. These conclusions are supported by the limited intervention window (12 weeks), the limited transferability of the samples, and the fact that pronunciation and grammatical range are not assessed. Longitudinal designs are recommended for future studies, as they expand the scope of oral assessment, including tests in sub-Saharan Africa, South and Southeast Asia, and Latin America, to determine the durability and global generalizability of AI-mediated spoken language development.

Author Contributions: Conceptualization, S.M.; methodology, S.M. and B. S; software, A.S.; validation, S.M., B.S., and A.S.; formal analysis, B. S; A.S.; investigation, S.M.; resources, S.M.; data curation, S.M.; writing—original draft preparation, S.M.; writing—review and editing, B.S.; visualization, S.M. and B. S; supervision, B. S; project administration, S.M. and B.S. All the authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Before data collection, ethical approval was obtained from the Research Ethics Committee of the University of Education with reference number UE/DM&AS/UEBS/EC/2026/15. All the participants were informed of the purpose of the research and gave their consent to participate. Participants were assured of the confidentiality of their responses and of their right to withdraw from the study at any stage without consequence.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). The publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

REFERENCES

Alessa, I. A., & Hussein, S. (2023). Using traditional and modern teaching methods on the teaching process from teachers' own perspective. *Route Educational and Social Science Journal*, *10*(2), 65–92. https://doi.org/10.17121/ressjournal.3361

Abbasi, W. T., Ibrahim, A. H., & Ali, F. (2022). Perceptions about English as second language teachers' technology based English language teaching in Pakistan: Attitudes, uses of technology and challenges. In M. Al-Emran, M. A. Al-Sharafi, M. N. Al-Kabi, & K. Shaalan (Eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems: ICETIS 2021* (Vol. 2, pp. 314–325). Springer. https://doi.org/10.1007/978-3-030-82616-1_28

Anwar, M. N., Ihsan, M., Hayat, Z., & Parvez, M. (2016). Teaching spoken English in Pakistan: Issues and options. *Journal of Educational Sciences & Research*, *3*(1), 55–60.

Bhutto, A. A., Jokhio, A. A., & Ahmad, A. (2023). A case study of teaching speaking skills in large classes: Problems and solutions. *Pakistan Journal of Humanities and Social Sciences*, *11*(2), 1953–1960. https://doi.org/10.52131/pjhss.2023.1102.0496

Curtis, A. (2021). What do we mean by under-resourced context? In K. M. Bailey & D. Christian (Eds.), *Research on teaching and learning English in under-resourced contexts* (pp. 14–28). Routledge. https://doi.org/10.4324/9781003057284-2

Duke, B. (2026). Artificial intelligence for inclusive and sustainable international education: A critical conceptual review. *Journal of International Students*, *16*(10), 167–198. https://doi/10.32674/96t36095

Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, *93*, 279–289. https://doi.org/10.1016/j.chb.2018.12.023

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, *57*(4), 542–570. https://doi.org/10.1111/ejed.12533

Hao, H. (2026). AI-driven analysis of English learners' perception, attitude, and motivation in online learning environments: A corpus-based study. *Interactive Learning Environments*, 1–15. https://doi.org/10.1080/10494820.2026.2649557

Huang, F., & Zou, B. (2024). English speaking with artificial intelligence (AI): The roles of enjoyment, willingness to communicate with AI, and innovativeness. *Computers in Human Behavior*, *159*, 108355. https://doi.org/10.1016/j.chb.2024.108355

Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: Impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skills. *Language Testing in Asia*, *12*(1), 40. https://doi.org/10.1186/s40468-022-00191-4

Jeon, J., Lee, S., & Choi, S. (2024). A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments*, *32*(8), 4613–4631. https://doi.org/10.1080/10494820.2023.2204343

Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, *206*, 104898. https://doi.org/10.1016/j.compedu.2023.104898

Kerly, A., Hall, P., & Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems*, *20*(2), 177–185. https://doi.org/10.1016/j.knosys.2006.11.014

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lichtman, K., & VanPatten, B. (2021). Was Krashen right? Forty years later. *Foreign Language Annals*, *54*(2), 283–305. https://doi.org/10.1111/flan.12552

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. Routledge. https://doi.org/10.4324/9780203813850

Moulavinafchi, A., Alhaideri, S. A. A., & Alfawwaz, S. N. M. (2026). The Impact of an AI-Powered Flipped Classroom on Grammar Competence and Language Anxiety in EFL Learners. *International Journal of Instruction*, *19*(1), 105–130. https://doi.org/10.29333/iji.2026.1916a

Morovat, E. (2014). Effects of Reflectivity/Impulsivity on IELTS Candidates' Band Scores in the Speaking Module of the Test. *Procedia - Social and Behavioral Sciences*, *98*, 1232–1239. https://doi.org/10.1016/j.sbspro.2014.03.538

Majeed, A. J., & Kareem, F. (2025). Teaching difficulties in overcrowded classrooms. *Journal of Humanities*, *32*(2), 1414–1443. https://doi.org/10.33855/0905-032-002-021

Poehner, M. E., & Lantolf, J. P. (2024). *Sociocultural Theory and Second Language Developmental Education* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009189422

Qin, L., & Derakhshan, A. (2026). The impact of robot-assisted language learning (RALL) on EFL students' classroom engagement and willingness to attend classes (WTAC): A technology acceptance model (TAM) perspective. *Language Teaching Research*. https://doi.org/10.1177/13621688261441862

Rashid, S., Cunningham, U., Watson, K., & Howard, J. (2018). Revisiting the digital divide(s): Technology-enhanced English language practices at a university in Pakistan. *Australian Journal of Applied Linguistics*, *1*(2), 64–87. https://doi.org/10.29140/ajal.v1n2.7

Rattanayat, W., & Boonsuk, Y. (2026). "My mind goes blank, and I just want to disappear": Dual-source anxiety and face dynamics in Thai collectivist EFL classrooms. *European Journal of Education*, *61*(3), Article e70723. https://doi.org/10.1111/ejed.70723

Santoso, D. R., Affandi, G. R., & Basthomi, Y. (2024). 'Getting stuck': A study of Indonesian EFL learners' self-efficacy, emotional intelligence, and speaking achievement. *Studies in English Language and Education*, *11*(1), 384–402. https://doi.org/10.24815/siele.v11i1.30969

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step toward second language learning. *Applied linguistics*, *16*(3), 371-391. https://doi.org/10.1093/applin/16.3.371

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge. https://doi.org/10.4324/9780203851357

Singh, R. B. (2024). Navigating English language education challenges in resource-limited contexts. *KMC Journal*, *6*(1), 135–152. https://doi.org/10.3126/kmcj.v6i1.62336

Xing, D., & Bolden, B. (2019). Exploring oral English learning motivation in Chinese international students with low oral English proficiency. *Journal of International Students*, *9*(3), 834–855. https://doi.org/10.32674/jis.v9i3.749

Yuan, L., & Liu, X. (2025). The effect of artificial intelligence tools on EFL learners' engagement, enjoyment, and motivation. *Computers in Human Behavior*, *162*, 108474. https://doi.org/10.1016/j.chb.2024.108474

Zargaran, Z. (2025). The effect of pending feedback on the IELTS speaking test result: Grammatical range and accuracy in focus. *Australian Journal of Applied Linguistics*, *8*(1), 102484. https://doi.org/10.29140/ajal.v8n1.102484

Author Bios

SARIA MAZHAR, MPhil, is a researcher at the University of Education, Lahore, Pakistan. Her research interests include applied linguistics, English language teaching, artificial intelligence in language education, and technology-mediated language learning. Email: sariamazhar890@gmail.com. ORCID: https://orcid.org/0009-0003-5055-5450

BUSHRA SANI, PhD, is an Assistant Professor in the Department of English at the University of Education, Lahore, Pakistan. She holds a PhD in Language and Communication Studies from the University of East Anglia, United Kingdom. Her research interests include sociolinguistics, applied linguistics, psycholinguistics, language acquisition, and language teaching and learning.
Email: bushra.sani@ue.edu.pk. ORCID: https://orcid.org/0009-0007-0009-8457

MUHAMMAD ADNAN SIAL, PhD, is an Assistant Professor in the Division of Management and Administrative Science at UE Business School, University of Education, Lahore, Pakistan. He holds a PhD in Administrative Sciences (Management) from the University of the Punjab, Lahore. His research interests include human resource management, knowledge management, and organizational behaviour, with recent work on socialization, tacit knowledge sharing, mobile technology, and public administration.
Email: adnan.sial@ue.edu.pk. ORCID: https://orcid.org/0000-0003-1751-8079