Tabaran Institute of Higher Education
International Journal of Language Testing
Vol. 15, No. 2; 2025, October

ISSN 2476-5880
https://www.ijlt.ir

The Effect of Artificial Intelligence (AI)-Mediated Speaking Assessment on Speaking Performance and Willingness to Communicate of Iraqi EFL Learners

Taif Abdulhussein Dakhil[1], Fatemeh Karimi[2]*, Riyadh Abbas Ubeid Al-Jashami[3], Zargham Ghabanchi[4]

ARTICLE INFO

Article History:
Received: November 2024
Accepted: December 2024

KEYWORDS
Artificial Intelligence (AI)-mediated speaking assessment
EFL learners
speaking performance
willingness to communicate

ABSTRACT

The present study aimed to find the effect of artificial intelligence (AI)-mediated speaking assessment on speaking performance and willingness to communicate of Iraqi EFL Learners. More specifically, the study sought to find whether AI-mediated speaking assessment enhance the speaking performance (grammar, vocabulary, pronunciation, intonation, and fluency) of Intermediate Iraqi EFL learners and whether AI-mediated speaking assessment enhanced their willingness to communicate in English. In so doing, 40 intermediate Iraqi EFL learners were randomly selected and assigned into experimental and control groups, each comprising 20 learners. The experimental group participants received ten 60-minute treatment sessions with ELSA Speech Analyzer, while the control group received no treatment. The speaking pre-test of both groups was run prior to the treatment, and all participants were post-tested at the end of the study. The Willingness to Communicate in a Foreign-Language Scale was also administered to groups prior to and after the treatment. A speaking assessment rubric, including vocabulary, grammar, intonation, pronunciation, and fluency, was used to assess the speaking performance of both groups. The findings demonstrated that AI-mediated speaking assessment enhanced the grammar, vocabulary, intonation, and fluency of the experimental group. However, the two groups did not differ in terms of pronunciation. Furthermore, this assessment tool enhanced the willingness to communicate with native speakers, the willingness to communicate with non-native speakers, and the willingness to communicate in the school context of this group. In general, the speaking assessment mediated by AI significantly enhanced the speaking performance and willingness to communicate of the learners. These findings might advance the current scholarly discourse on AI within the domains of language pedagogy and assessment.

[1] English Department, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran, Email: Taif55181@gmail.com
[2] English Department, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran, Email: fkarimi@khuisf.ac.ir
[3] Department of English language, College of Education, Sawa University, Almuthana, Iraq, Email: Dr.riyadh955@gmail.com
[4] Ferdowsi University of Mashhad, Email: ghabanchi@um.ac.ir

Cite this paper as: Abdulhussein Dakhil, T., Karimi, F., Abbas Ubeid Al-Jashami, R., & Ghabanchi, Z. (2025). The effect of Artificial Intelligence (AI)-mediated speaking assessment on speaking performance and willingness to communicate of Iraqi EFL learners. *International Journal of Language Testing, 15*(2), 1–18.
https://doi.org/10.22034/ijlt.2024.486564.1383

1. Introduction

Speaking constitutes an indispensable language skill that is utilized for the purpose of interpersonal communication, articulation of personal perspectives, and the manifestation of one's identity. In the contemporary context of globalization, proficiency in speaking is acknowledged as a fundamental requirement for international progress, access to advanced educational opportunities, and career prospects (Fulcher, 2015; Isaacs, 2016) and has emerged as a significant element in the majority of both international and local language proficiency assessments, attributable in part to the ascendancy of the communicative approach in language instruction and evaluation (Fulcher, 2000). Nevertheless, despite its paramount importance within the realms of language education and assessment, speaking has often been perceived as an abstract construct that poses considerable challenges in terms of conceptualization and reliable, valid assessment. This phenomenon may be ascribed to the fluid and contextually embedded characteristics of speaking or may also result from the diverse formats it can adopt (e.g., speech, paired dialogue, group format) and the varying circumstances in which speaking occurs, namely premeditated or unplanned (Luoma, 2004; Carter & McCarthy, 2017). In the assessment of speaking proficiency, a multitude of factors emerge that potentially influence the performance of test takers and, consequently, their resultant scores, encompassing task attributes, interactants' traits, rater biases, and the rating scale employed, among other variables (McNamara, 1996; Fulcher, 2015). Doosti and Ahmadi Safa (2021) also, found that training raters can enhance inter-rater reliability and fairness.

Within the domain of language assessment, significant scholarly inquiry and resources have been devoted to exploring speaking assessment and the introduction of technology. Now, artificial intelligence has partly changed the face of speaking assessment. As technology continues to evolve, artificial intelligence is becoming increasingly integrated into educational practices. There is a growing interest in understanding how AI tools can enhance language learning, particularly in areas like speaking assessment, where traditional methods may struggle to provide personalized feedback (Zou et al., 2024).

Artificial Intelligence (AI) has been extensively integrated into language teaching methodologies to augment speaking competencies (Kuddus, 2022). A variety of AI-driven instruments, including AI chatbots, intelligent personal assistants, and platforms based on automatic speech recognition (ASR), are accessible for the practice of L2 speaking (Bashori et al., 2021; Sha, 2009; Tai & Chen, 2022). Mobile applications harness processing features of natural languages and technologies for evaluating speech to facilitate L2 speaking practice (Chang et al., 2022). AI applications frequently incorporate algorithms to assess the voice designed for tasks such as oral reading and presentations (Sha, 2009). These applications present numerous advantages, including time efficiency for educators, enhanced instructional quality, and tailored feedback for learners (Xie et al., 2019). AI has the capability to furnish multidimensional feedback, encompassing performance ratings for speaking, color-coded highlighting for accuracy, and textual evaluations for practical suggestions (Ahn & Lee, 2016; Chiu et al., 2023; Ebadi & Ebadijalal, 2022). Although various claims have been made that AI has positively impacted learning and teaching (Alam, 2022; Sun et al., 2021), its application in assessment is limited. Furthermore, there is a lack of a coherent knowledge base of how and why it is applied in English language assessment (Alonzo et al., 2024).

Willingness to communicate is closely tied to speaking performance. The propensity of language learners to engage in discourse and, specifically, speaking skills within distinct contexts and among particular interlocutors delineates their willingness to communicate (Mehrgan, 2013). Dörnyei (2003) elucidated that mere competence in the second language (L2) may not suffice. It is essential for learners to possess the capability to communicate and the inclination to utilize the L2. Empirical studies indicate that learners' willingness to communicate increasingly influences their engagement in communicative interactions (Clément et al., 2003; Yashima et al., 2004). Consequently, MacIntyre et al. (1998) advocate for the recognition of willingness to communicate as a primary goal of second language acquisition, serving as a comprehensive framework to elucidate, explicate, and forecast communicative acts.

As stated, willingness to communicate is a crucial factor in language learning, influencing how often learners engage in speaking activities. By examining the impact of AI tools on learners' willingness to communicate, this study aims to identify ways to create more favorable conditions for language use, improving learners' engagement and, ultimately, their language proficiency.

The significance of such studies lies in their potential to influence teaching practices, curriculum design, educational policy, and further research in the field while also providing direct benefits to learners through improved language acquisition strategies. Specifically put, the findings of such studies might provide valuable insights for teachers and language instructors, particularly those working with EFL learners. The possible improvement in speaking performance and willingness to communicate suggests that integrating AI tools into language instruction and assessment can enhance learning outcomes. This can inform teaching methodologies and curricular design, encouraging educators to adopt technology-enhanced assessment practices.

Moreover, the evidence that AI-mediated assessments might enhance the speaking performance and willingness to communicate of EFL learners may encourage them to engage more actively with these technologies, fostering greater confidence and proficiency in their language skills because when it is time to speak, most Iraqi students cannot correctly produce the target language (Abdul Razak et al., 2018). Also, Hamza (2023) stated that the majority of Iraqi EFL learners are not fluent and cannot master speaking a foreign language.

2. Review of Literature

The assessment rubric employed by human raters is multi-faceted and specific to a level of proficiency. However, AI-driven automatic speaking assessment systems primarily concentrate on identifying explicit linguistic features for assessment purposes. These systems analyze a range of features, including vocabulary, fluency, etc., derived from the recordings of examinees' language production. Subsequently, these identified features are evaluated through the application of models inspired by neural networks (Wang et al., 2021) or automatic speech-scoring methodologies, such as multiple regression (Xi et al., 2012; Sharadgah & Sa'di, 2022). This approach of extracting and evaluating features is similarly adopted by online speaking programs generated by AI, like Speech Analyzer. Such programs frequently incorporate advanced scoring components to entice prospective users.

Recent studies have also explored spoken interactions mediated by AI and the utilization of technology to recognize speech for assessment and pedagogical functions automatically (Phillips et al., 2023; Jung Youn, 2023). The implementation of Intelligent Personal Assistants (IPAs) and Spoken Dialog Systems (SDS), like Alexa, has been recognized for their potential to augment interactive prospects and student engagement within educational environments. Notwithstanding these technological advancements, the reliability and validity of these tools as substitutes for human evaluators remain unverified.

Zou et al. (2023) undertook a comprehensive investigation into the influence of artificial intelligence applications on the speaking performance of EFL learners. The research employed a combination of questionnaires and interviews, revealing that the participants preferred interactive activities facilitated by AI technologies. Engagement in social interactions through networks within the context of AI markedly enhanced the speaking proficiency of the participants. The experimental group participants scored higher on post-tests and exhibited a heightened conviction regarding the advantages of AI applications. The incorporation of platforms such as WeChat yielded augmented communication possibilities with educators and peers, thereby fostering motivation for speaking practice and enhancing educational outcomes.

In sum, various technology-based resources, including AI-based tutoring technologies and virtual conversation partners, provide individualized feedback and tailored practice opportunities that address the specific needs of learners (Jung Youn, 2023). The consistent utilization of these resources contributes to increased speaking confidence and facilitates the development of proficient communication skills in the target language.

2.1. Speaking and Willingness to Communicate

The concept of willingness to communicate pertains to an individual's motivation and inclination to utilize the language and has been examined within diverse theoretical backgrounds (Zarrinabadi & Khodarahmi, 2017). The socio-psychological perspective underscores the significance of interpersonal elements such as language ego and motivation (MacIntyre et al., 2001; Wei & Xu, 2022). Conversely, the socio-cultural framework accentuates the influence of social and cultural norms

in forming willingness to communicate (Yu, 2021). The ecological model underscores the intricate interplay among personal, interpersonal, and contextual factors in second language (L2) interaction (Cao, 2011). Although the conceptualization of willingness to communicate in L2 differs by framework, it typically denotes an individual's willingness to communicate in their second language, shaped by factors such as motivation, ecological considerations, and the like.

Willingness to Communicate Framework (MacIntyre et al., 1998), which formed the theoretical underpinning of this study for willingness to communicate, is one of the foundational theories for willingness to communicate was proposed by MacIntyre and his colleagues (1998), who conceptualized willingness to communicate as a complex construct influenced by multiple factors, including personality traits, language proficiency, and social context. The framework suggests that the context in which communication occurs significantly impacts an individual's willingness to engage, leading to different categories of willingness to communicate. The framework also emphasizes the importance of situational context in influencing willingness to communicate. Factors such as the setting (e.g., classroom, social gatherings), the relationship with the interlocutor (e.g., native versus non-native speakers), and specific communication situations lead to varying levels of willingness to communicate (MacIntyre et al., 1998).

Willingness to communicate, which is influenced by variables such as personality character and social norms, demonstrates a significant correlation with proficiency in speaking (Zarrinabadi & Khodarahmi, 2017). Empirical evidence denotes a positive relationship between speaking proficiency in L2 and willingness to communicate, implying that enhanced proficiency fosters increased confidence in interaction (Khajavy et al., 2016). Confidence is paramount for learners as they strive to be involved in authentic language use and improve their proficiency by practicing within an individualized learning environment (Tai & Chen, 2022).

Ma et al. (2022) qualitatively investigated the willingness to communicate of a group of advanced Chinese EFL learners in an English medium instruction classroom to find the reason they are not willing to communicate in class despite their advanced language proficiency level. The data collected through observations and interviews demonstrated that educational, cultural, and individual factors played a role. The researchers called for further attention to factors other than proficiency when studying and examining the willingness to communicate construct.

Language acquisition applications and platforms, incorporating interactive exercises, speech recognition capabilities, and virtual communication mechanisms, have demonstrated significant potential in improving second language speaking proficiency and willingness to communicate (Ahn & Lee, 2016; Ebadi & Ebadijalal, 2022). Integrating artificial intelligence technology can further enhance individualized language learning resources and virtual assistants, thereby facilitating learners' speaking practice with augmented interactivity and efficacy (Chiu et al., 2023).

Regarding the link between AI tools and willingness to communicate, these tools provide tailored feedback on pronunciation, intonation, grammar, and vocabulary. By analyzing learners' responses, the AI can offer specific suggestions for improvement, thereby reducing anxiety and increasing willingness to communicate (Zhang et al., 2024). Systems that track learners' progress over time and highlight improvements and success can build confidence and reduce anxiety around speaking, thereby enhancing willingness to communicate. Furthermore, AI creates adaptive learning experiences that adjust based on learner performance and engagement levels, promoting a sustained willingness to communicate through progressively challenging speaking tasks.

Among the very few studies conducted on AL and willingness to communicate, Deng et al. (2024) investigated the effect of AI on the willingness to communicate among Chinese university students engaged in face-to-face (F2F) interactions with a native English speaker. Fourteen students were engaged in AI chatbot-assisted communication practice simulating human-like dialogues prior to F2F interactions with a native English-speaking teacher outside the traditional classroom setting. The results of the willingness to communicate survey indicated positive outcomes and content analysis findings reported that students exhibited shifts from negative emotional states to positive ones.

Zou et al. (2024) examined the perceptions of EFL learners regarding both the potential benefits and the constraints of *EAP Talk*, an artificial intelligence-driven speech assessment system, utilized for speaking practice. The data were gathered from 366 EFL learners at five distinct universities via the use of questionnaires and semi-structured interviews. The results indicated that *EAP Talk* considerably

improved speaking competencies, encompassing pronunciation, grammatical accuracy, organization of ideas, read-aloud proficiency, and presentation capabilities. Participants expressed their appreciation for the convenience, motivational factors, and institutional effectiveness associated with *EAP Talk*, which facilitated flexible and autonomous learning experiences.

Sayed et al. (2024) examined the impact of artificial intelligence-driven tools on the speaking competencies, psychological well-being, autonomy, and academic resilience of EFL learners. The study involved 28 upper-intermediate EFL students enrolled at an Ethiopian university. The findings indicated significant enhancements in speaking proficiency, psychological well-being, learner autonomy, and academic resilience. The findings imply that through the advancement of skill acquisition, provision of personalized feedback, and addressing the emotional and psychological requirements of students, AI systems such as ChatGPT possess the potential to revolutionize language assessment and instructional methodologies.

Dai and Wu (2023) and Xiao and Park (2021) used automatic speech recognition (ASR) technology to facilitate and assess the pronunciation learning of Chinese EFL learners. The findings of both studies demonstrated that the ASR technology was more successful in pronunciation assessment than human assessment, and the inclusion of this technology enhanced the pronunciation of the participants.

Zou et al. (2023) studied the efficacy of various automatic feedback mechanisms provided by artificial intelligence speech evaluation programs in facilitating the enhancement of speaking competencies of EFL learners. The findings indicated that a significant proportion of participants perceived an enhancement in their speaking abilities as a result of the feedback delivered by the AI speech evaluation program.

Liakin et al. (2014) investigated whether the participants better acquire the L2 French vowel /y/ pronunciation using ASR. In so doing, the participants were divided into ASR and non-ASR groups. The findings revealed the outperformance of the ASR group.

Anggraini (2022) utilized the ELSA Speech Analyzer to develop and implement lesson plans and pedagogical strategies aimed at assessing students' progress in fluency, grammatical accuracy, lexical precision, appropriateness, and overall comprehensibility throughout a pronunciation instruction course. The findings indicated that the enhancement of all specified attributes exhibited a consistent positive trajectory, with 85% of the students reporting increased motivation to engage in English practice.

Amidst the literature on EFL learners' willingness to communicate and speaking assessment, there is a research gap regarding the integration of AI into exploring these two factors. Hence, this study explores the following questions:

1. Does Artificial Intelligence (AI)-mediated speaking assessment enhance the speaking performance (grammar, vocabulary, pronunciation, intonation, and fluency) of Intermediate Iraqi EFL learners?
2. Does Artificial Intelligence (AI)-mediated speaking assessment enhance Intermediate Iraqi EFL learners' willingness to communicate in English?

3. Method

3.1. Participants and Setting

A total of 40 intermediate Iraqi EFL learners at the University of Kufa whose ages ranged between 23 and 39 (M = 31, SD = 4.62) were selected by convenience sampling, which, despite its drawbacks like bias that affect generalizability, enjoys benefits like spending less effort to select the participants compared to non-random sampling techniques (Shorten & Moorley, 2014). Fifteen students were male, and 25 were female. They were all Arabic native speakers, and none had lived or studied in an English-speaking country. It is noteworthy that the participants' only experience with technology was online courses during the COVID-19 pandemic.

3.2. Instrumentation

3.2.1. Oxford Quick Placement Test (OQPT). OQPT was administered to select intermediate EFL learners. The OQPT represents a versatile instrument for evaluating English language proficiency, comprising 60 multiple-choice questions that assess vocabulary (30 items) and grammar (30 items);

learners attaining scores between 0 and 10 are classified as beginners, while those with scores ranging from 11 to 17 are categorized as breakthrough learners; individuals who achieve scores from 18 to 29 are designated as elementary, pre-intermediate learners are identified with scores between 30 and 39; intermediate learners receive scores from 40 to 47; advanced learners are classified with scores from 48 to 54, and proficient learners are recognized with scores ranging from 55 to 60. The reliability of the test, estimated by Cronbach's alpha, was .7, and two TEFL university professors confirmed its validity.

3.2.2. ELSA Speech Analyzer. ELSA Speech Analyzer represents a novel AI-driven English language tool that attentively listens to the learner's speech and furnishes immediate feedback concerning their vocabulary pronunciation, grammar, intonation, and fluency. Additionally, it anticipates the prospective score a learner might achieve in prominent English-speaking assessments (Anguera et al., 2023). The pronunciation score is derived from the precision with which English phonemes are articulated in each word identified by the system. It offers corrective guidance on how to ameliorate each identified mispronunciation, as the frequency and severity of each mispronunciation are integral to the overall score. In assessing a user's intonation, the ELSA Speech Analyzer evaluates multiple parameters based on word prominence, pitch, and energy. Pitch and energy profiles are scrutinized for excess or inadequate variation. Fluency assessment encompasses the occurrence of hesitations, pauses, and pacing. The pacing score depends on the speaker's average word count per minute and variability. The pausing score compares the estimated optional and obligatory pauses within the transcript with the detected pauses. The frequency and characteristics of the identified hesitations determine the hesitations score. The ELSA Speech Analyzer computes the grammar score by appraising both grammatical range and accuracy, subsequently integrating these two sub-scores. The grammatical range assesses whether the speaker can employ diverse linguistic structures. The vocabulary score is grounded in the lexicon utilized during the recording, based on the Common European Framework of Reference (CEFR) levels for words and phrases. Furthermore, the ELSA Speech Analyzer recommends users incorporate more formal expressions or sophisticated alternatives to the words utilized (Anguera et al., 2023).

The researchers selected the ELSA Speech Analyzer for the present study because it provides learners with individualized, immediate feedback on their pronunciation and speaking skills. This helps learners identify specific areas of improvement, such as phonetics, intonation, and rhythm. Besides, this tool utilizes advanced AI algorithms to assess speech accuracy and fluency and allows for detailed analysis that can be more objective than traditional assessments conducted by human evaluators. It is noteworthy that the ELSA Speech Analyzer was used in previous studies on speaking (e.g., Azzahra et al., 2024).

3.2.3. Speaking Pre-test and Post-test. For the pre- and post-test of speaking, the experimental and control groups' participants were requested to respond to the question, "What are the advantages and disadvantages of social media?" The reliability coefficients of the pre- and post-tests were 0.8 and 0.77, respectively.

3.2.4. Speaking Assessment Rubrics. A structured speaking assessment rubric was developed based on the previous literature to evaluate the speaking proficiency of experimental and control participants (e.g., Girón-García & Llopis-Moreno, 2015; Anh et al., 2022). The reason for developing rubrics based on the literature was that the rubrics used for assessing the speaking performance of the participants needed to match what the ELSA Speech Analyzer assesses in speaking performance (i.e., vocabulary, grammar, intonation, pronunciation, and fluency).

This rubric assesses five key criteria: vocabulary, grammar, intonation, pronunciation, and fluency. Each criterion was rated on a scale of 1 to 5, with descriptors for each score level providing clarity on performance expectations (see Appendix A).

The rubric was pilot-tested with five intermediate Iraqi EFL learners prior to the main study. Feedback from assessors during the pilot phase led to refinements in the rubric wording, enhancing clarity and consistency. Furthermore, the rubric was reviewed by two experts in language assessment and EFL instruction to ensure that the criteria comprehensively measured key components of speaking proficiency.

3.2.5. Willingness to Communicate in a Foreign-Language Scale. This 22-item scale, which was developed and validated by Baghaei (2013), covers three sub-scales of (a) willingness to communicate with native-speakers (items 1–7), (b) willingness to communicate with foreign non-native speakers (items 8–14), and (c) willingness to communicate with classmates/instructors who learn and teach the foreign language, willingness to communicate in the school context (items 15–22). The responses are rated on a five-point Likert scale (0= strongly disagree to 4= strongly agree). The reliability coefficients of the scale reported by Baghaei (2013) and the present study were 0.9 and 0.83, respectively. Two TEFL university professors also checked the validity of the scale (See Appendix B).

3.3. Procedures

As mentioned above, 40 intermediate Iraqi EFL learners, homogenized by QOPT, participated in this study and were divided into experimental and control groups, each comprising 20 EFL learners. The experimental group participants received ten 60-minute treatment sessions with ELSA Speech Analyzer, while the control group received no treatment (i.e., doing routine classroom tasks). The experimental group participants used personal computers to access the ELSA Speech Analyzer. The speaking pre-test of both groups was run prior to the treatment, and all participants were post-tested at the end of the study. The willingness to communicate questionnaire was also administered to groups prior to and after the treatment.

In the first treatment session, the experimental group participants were given an overview of the ELSA Speech Analyzer and its features, including how it can provide real-time feedback on pronunciation, intonation, fluency, vocabulary, and grammar, and were engaged in a series of practice exercises using the ELSA Speech Analyzer to familiarize themselves with its interface and functionality. These exercises included reading passages, repeating sentences, or engaging in conversation prompts designed and prepared by the teacher. The rest of the nine sessions were devoted to speaking practice and receiving ELSA feedback. For example, participants described a favorite dish focusing on ELSA feedback on vocabulary usage or a recent event using at least five compound sentences and received ELSA feedback on grammar accuracy. Practicing reading sentences with varying intonations and pronunciation of different English words and a debate activity where learners express their viewpoints quickly (for fluency assessment) were also planned during the treatment sessions.

Progress was monitored throughout the intervention using the built-in features of the ELSA application, which allowed participants to view their improvement over time. The instructor also kept records of each participant's scores and feedback to personalize future sessions. Two independent raters assessed each speaking performance to mitigate potential bias. The raters were meticulously selected from those who had experience assessing speaking, especially for IELTS, and were familiar with ELSA. Besides, the assessment rubrics were given to them, and they were requested to access the participants based on the rubrics. Inter-rater reliability was calculated using Cohen's Kappa, resulting in a Kappa value of 0.85, indicating strong agreement between assessors.

4. Results

As stated above, the first research question sought to find whether AI-mediated speaking assessment enhanced the speaking performance (grammar, vocabulary, pronunciation, intonation, and fluency) of Intermediate Iraqi EFL learners. Multivariate analysis of variance (MANOVA) was run to compare the pre- and post-test scores of both groups in terms of grammar, vocabulary, pronunciation, intonation, and fluency. Multivariate analysis of variance (MANOVA) is an extension of the analysis of variance for use when there is more than one dependent variable (here, we mean vocabulary, grammar, intonation, pronunciation, and fluency). These dependent variables should be related in some way, or there should be some conceptual reason for considering them together.

The normality assumption of the analysis was checked by the Kolmogorov-Smirnoff and Shapiro-Wilk tests' statistics for the speaking performance components mentioned above, and the findings indicated the normal distribution of the data ($p > .05$).

Table 1
Descriptive Statistics of Speaking Performance of Experimental and Control Groups

	Group	Test	Mean	SD	N
grammar	Experimental	Pre-test	1.5	.51	20
		Post-test	3.7	.47	20
	Control	Pre-test	1.6	.48	20
		Post-test	1.85	.58	20
vocabulary	experimental	Pre-test	1.75	.63	20
		Post-test	3	.79	20
	Control	Pre-test	1.9	.64	20
		Post-test	2	.56	20
intonation	experimental	Pre-test	1.9	.55	20
		Post-test	2.7	.47	20
	Control	Pre-test	1.82	.48	20
		Post-test	1.8	.52	20
pronunciation	experimental	Pre-test	1.8	.52	20
		Post-test	1.65	.48	20
	Control	Pre-test	1.7	.57	20
		Post-test	1.8	.61	20
fluency	experimental	Pre-test	1.95	.6	20
		Post-test	3.45	.51	20
	Control	Pre-test	1.7	.47	20
		Post-test	2.05	.39	20

Table 1 shows that the experimental group obtained higher scores in the post-tests of grammar, vocabulary, pronunciation, intonation, and fluency compared to the control group. The following table shows if the two groups significantly differed from each other in this regard.

Table 2
Multivariate Test of Speaking Performance of Experimental and Control Groups

	Value	F	Hypothesis df	Error df	Sig.
Pillai's trace	.89	25.09	10	29	.00
Wilks' lambda	.1	25.09	10	29	.00
Hotelling's trace	8.65	25.09	10	29	.00
Roy's largest root	8.65	25.09	10	29	.00

The result of Wilk's Lambda F (10, 29) = 25.09, P= .00 indicates a statistically significant difference among the scores of speaking performance components of both groups (Table 2). The pairwise comparison results (Table 3) show the components whose difference was significant.

Table 3
Pairwise Comparisons of Speaking Performance of Experimental and Control Groups

						95% Confidence Interval for Difference	
Dependent Variable	(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
grammar post-test	Experimental	control	1.85*	.16	.00	1.51	2.19
vocabulary post-test	Experimental	control	1*	.21	.00	.55	1.44
intonation post-test	Experimental	control	.9*	.15	.00	.58	1.21
pronunciation post-test	Experimental	control	-.15	.17	.39	-.5	.2
fluency post-test	Experimental	control	1.4*	.14	.00	1.1	1.69

*. The mean difference is significant at the .05 level.

The pairwise comparisons table reveals that the difference between the mean scores of speaking performance components was significant comparing the two groups ($p < .05$). In other words, Artificial Intelligence (AI)-mediated speaking assessment enhanced grammar, vocabulary, intonation, and fluency. However, the two groups did not differ in terms of pronunciation.

The second research question aimed to find whether AI-mediated speaking assessment enhanced Intermediate Iraqi EFL learners' willingness to communicate in English. Multivariate analysis of variance (MANOVA) was run to compare the pre- and post-test scores of both groups in terms of willingness to communicate with native speakers, willingness to communicate with foreign non-native speakers, and willingness to communicate in the school context.

Table 4
Descriptive Statistics of Willingness to Communicate of Experimental and Control Groups

	Group	Test	Mean	SD	N
willingness to communicate with non-native speakers	experimental	Pre-test	16.7	1.34	20
		Post-test	24.3	1.65	20
	Control	Pre-test	17.85	.81	20
		Post-test	17.55	2.35	20
willingness to communicate in the school context	experimental	Pre-test	19.9	.91	20
		Post-test	27.4	1.46	20
	Control	Pre-test	19.8	.83	20
		Post-test	20.85	.87	20
willingness to communicate with native-speakers	experimental	Pre-test	13.7	1.55	20
		Post-test	23	2.31	20
	Control	Pre-test	14.25	1.68	20
		Post-test	15.45	.88	20

Table 4 shows that the experimental group obtained higher scores in the post-tests of willingness to communicate components than the control group. Table 5 shows if the difference between the two groups was significant in this regard.

Table 5
Multivariate Test of Willingness to Communicate of Experimental and Control Groups

	Value	F	Hypothesis df	Error df	Sig.
Pillai's trace	.93	73.49	6	33	.00
Wilks' lambda	.07	73.49	6	33	.00
Hotelling's trace	13.36	73.49	6	33	.00
Roy's largest root	13.36	73.49	6	33	.00

The result of Wilk's Lambda $F(6, 33) = 73.49$, $P = .00$ indicates a statistically significant difference among the scores of willingness to communicate components of both groups (Table 5). The pairwise comparison results (Table 6) show the components whose difference was significant.

Table 6
Pairwise Comparisons of Willingness to Communicate of Experimental and Control Groups

						95% Confidence Interval for Difference	
Dependent Variable	(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
willingness to communicate with native-speakers post-test	Experimental	control	7.55*	.55	.00	6.42	8.67

willingness to communicate with non-native speakers post-test	Experimental	control	6.75*	.64	.00	5.44	8.05
willingness to communicate in the school context post-test	Experimental	control	6.55*	.38	.00	5.77	7.32

*. The mean difference is significant at the .05 level.

The pairwise comparisons table reveals that the difference between the mean scores of willingness to communicate components was significant comparing the two groups ($p < .05$). In other words, AI-mediated speaking assessment enhanced willingness to communicate with native speakers, willingness to communicate with non-native speakers, and willingness to communicate in the school context.

5. Discussion

The present study aimed to find the effect of artificial intelligence (AI)-mediated speaking assessment on speaking performance and willingness to communicate of Iraqi EFL Learners. The first research question sought to find whether AI-mediated speaking assessment enhances the speaking performance (grammar, vocabulary, pronunciation, intonation, and fluency) of intermediate Iraqi EFL learners. The findings demonstrated that AI-mediated speaking assessment enhanced grammar, vocabulary, intonation, and fluency. However, the two groups did not differ in terms of pronunciation. The second research question aimed to uncover whether AI-mediated speaking assessment enhances Intermediate Iraqi EFL learners' willingness to communicate in English, and the findings revealed that AI-mediated speaking assessment enhanced willingness to communicate with native speakers, willingness to communicate with non-native speakers, and willingness to communicate in the school context.

The findings revealed the beneficial effects of the ELSA Speech Analyzer in enhancing the speaking performance of the present study participants. In explicating the findings, it can be stated that the speech evaluation driven by AI programs has been recognized as advantageous for advancing speaking competencies among second or foreign-language learners (Dai & Wu, 2023; Xiao & Park, 2021). Dai and Wu (2023) and Xiao and Park (2021) reported the beneficial effect of a speech recognition system in enhancing the pronunciation of Chinese EFL learners. Furthermore, it is regarded as engaging, accessible, and conducive to learners' enhancement in fluency, intonation, and pronunciation (Kan & Ito, 2020; Tai & Chen, 2023). The observed enhancement in speaking abilities, as it was also found in the present study, due to AI may be ascribed to the instantaneous feedback rendered by AI speech evaluation systems. Feedback is paramount in foreign language acquisition due to its affirmative effects on educational practices (Li, 2013). For instance, corrective feedback during speaking exercises facilitates learners to recognize their spoken grammatical errors and the disparities between their verbalizations and established linguistic norms, thereby increasing their motivation to amend their expressions (Penning de Vries et al., 2020). The participants of this study also received feedback on their speaking performance from the ELSA Speech Analyzer, which might further justify their outperformance compared to the control group.

The findings of the present study are somewhat supported by the literature. For instance, Anggraini (2022) used the ELSA Speech Analyzer for the development and implementation of lesson plans and pedagogical strategies and the results revealed students' increased motivation in learning English. Besides, Zou et al.'s (2023) utilization of AI-supported automatic feedback mechanisms enhanced the speaking performance of EFL learners.

As the findings revealed, no significant improvement was found in terms of pronunciation. In explaining this finding, it can be stated that notwithstanding their merits, programs predicated on artificial intelligence may encounter constraints regarding pronunciation teaching. Although these tools provide comprehensive exercises and prompt feedback, they lack components pertinent to phonetic elucidation, such as explicating the employment of the vocal system for particular phonemes or

delineating the distinctions between target phonemes and those of the learner's native language (Liakin et al., 2014). This might justify why no significant improvement was observed in the pronunciation of participants. Additionally, ten training sessions might not be long enough to enhance the pronunciation of Iraqi EFL learners.

Regarding the beneficial effect of Artificial Intelligence (AI)-mediated speaking assessment on willingness to communicate, the advantageous effects of AI-mediated interactions on the willingness to communicate of English learners can be ascribed to the distinctive attributes of the technology itself, which offers immediate and customized feedback that is specifically aligned with the learners' individual requirements and their respective levels of English language proficiency. In the words of Ma et al. (2022), such tailored feedback may prove particularly advantageous for learners who exhibit reluctance to engage in communication within a classroom environment, as it assists them in recognizing areas necessitating enhancement and fostering their self-efficacy.

Consistent with the findings of the present study, Shafiee Rad (2024) studied the effects of artificial intelligence on English speaking proficiency, willingness to communicate, and learner perceptions utilizing an experimental framework. The results revealed that the experimental group demonstrated markedly superior willingness to communicate, L2 speaking proficiency, and more favorable perceptions when compared to the control group. Additionally, Kim and Su (2024) documented substantial enhancements in willingness to communicate, diminished anxiety levels, and increased communication confidence among students in the experimental group who were trained with AI chatbots.

6. Conclusion

This research has yielded empirical evidence indicating that the speaking assessment mediated by AI significantly enhances the speaking competencies and willingness to communicate of English learners ($p<.05$). The findings advance the current scholarly discourse on AI within the domains of language pedagogy and assessment by scrutinizing the efficacy of interactions mediated by AI in augmenting the speaking abilities and willingness to communicate of EFL learners. More specifically, the results imply that interactions mediated by AI offer an effective and stimulating modality for EFL learners to refine their speaking skills and elevate their willingness to communicate ($p<.05$).

The implications of these findings are salient for EFL learners, educators, teacher trainers, and education researchers. Firstly, in light of the outcomes of this study, EFL learners are encouraged to utilize AI-mediated speaking assessment instruments to partake in interactive speaking exercises and enhance their speaking capabilities, as these tools offer dynamic platforms for conversation practice, grammatical instruction, and vocabulary enhancement. Furthermore, the tailored feedback pertaining to speaking skills provided by AI-mediated assessment caters to individual learning requirements and cultivates a supportive atmosphere conducive to the advancement of speaking competencies. The low-stress environment fostered by AI tools permits EFL learners to engage in speaking practice devoid of apprehension regarding evaluation, which may subsequently promote greater fluency and self-assurance in speaking. Additionally, EFL learners may leverage AI-mediated speaking assessment tools beyond the confines of traditional classroom settings to facilitate ongoing speaking practice in informal contexts, thereby empowering them to transcend conventional constraints and address challenges such as time limitations and large class sizes.

The implications extend to policymakers in education, particularly where EFL is taught. The positive impact of AI on language learning could inform policy decisions regarding funding, resources, and support for technology integration in language education. Moreover, teacher training programs can include modules on integrating AI and speech analysis tools into language teaching since educators need to be equipped with the skills to effectively incorporate technology into their pedagogy to enhance speaking skills and facilitate assessment.

EFL instructors can incorporate speaking assessment mediated by AI into their speaking curricula, thereby augmenting opportunities for learners to engage in self-evaluation of their speaking competencies. Furthermore, it is advisable for EFL instructors and teacher educators to equip EFL learners and instructors, respectively, with the necessary training to maximize the advantages of such technological tools.

Despite its advantages, while ELSA aims to provide personalized feedback on pronunciation, its effectiveness may vary for different English accents. Users with regional accents might not receive accurate assessments or tailored guidance. Also, ELSA primarily focuses on English. For learners whose native language is not supported or for those learning other languages, it may not be beneficial.

Regarding the limitations inherent in this study, it is noteworthy that the sample size was relatively small and may not have accurately represented the entirety of the EFL learner population. Consequently, future investigations would benefit from employing a larger sample size and replicating the current research to enhance the generalizability of the results. Additionally, the duration of the study was confined to a relatively brief timeframe (ten sessions). As such, the long-term ramifications of AI speaking assessments on speaking performance and willingness to communicate were not examined. Subsequent research endeavors should investigate the effects of prolonged utilization of these tools. Lastly, the speaking assessment was solely conducted utilizing ELSA Speech Analyzer; thus, exploring alternative AI-mediated instruments may yield other outcomes. Therefore, future researchers can compare the ELSA Speech Analyzer and similar tools.

Acknowledgments
We would like to thank all the students who participate in this study.

Declaration of Conflicting Interests
The authors have no conflict of interest.

Funding
The authors received no funds for this study.

Declaration of Applying AI
No AI-assisted technology was used in the writing process.

References

Abdul Razak, N., Saeed, H. K. & Alakrash, H. (2018). Pedagogical issues of using ICT applications in Iraq. *Asia-Pacific Journal of Information Technology & Multimedia, 2,* 79–87. https://doi.org/10.17576/apjitm-2018-0702(02)-07

Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, *47*(4), 778–786. https://doi.org/10.1111/bjet.12354

Alam, A. (2022). Employing adaptive learning and intelligent tutoring robots for virtual classrooms and smart campuses: Reforming education in the age of artificial intelligence. In R. N. Shaw, S. Das, V. Piuri, & M. Bianchini (Eds.), *Advanced computing and intelligent technologies* (pp. 43–60). Singapore.

Alonzo, D., Abril, J. M. V., & Zin Oo, C. (2024). *The use of artificial intelligence in English language assessment: Empirical evidence and future directions*. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4695394

Anggraini, A. (2022). Improving students' pronunciation skills using ELSA speak application. *Journal of English Language and Pedagogy, 5*(1), 135–141. https://doi.org/10.33503/journey.v5i1.1840

Anguera, X., Proença, J., Gulordava, K., Tarján, B., Parslow, N., Dobrovolskii, V., ... & Girard, R. (2023). ELSA Speech Analyzer: English communication assessment of spontaneous speech. In Y. Boytsiy, A. Stent, & N.O. Pelsma (Eds.), *Proceedings of the 9th Workshop on Speech and Language Technology in Education (SLaTE)* (pp. 95–96). Association for Computational Linguistics.

Anh, K. H., Dong, M. H., & Trang, N. H. (2022). An investigation into English-majored students' self-assessment of their speaking performance. *International Journal of Instruction, 15*(3), 191–208. https://doi.org/10.29333/iji.2022.15311a

Azzahra, N., Widiastuti, P. T., Sopyani, N., & Luthfiyyah, R. (2024). A study on English students' perception regarding the AI implementation on listening and speaking skills. *Indonesian*

Journal of Innovation and Applied Sciences (IJIAS), *4*(2), 113–120. https://doi.org/10.47540/ijias.v4i2.1466

Baghaei, P. (2013). Development and psychometric evaluation of a multidimensional scale of willingness to communicate in a foreign language. *European Journal of Psychology of Education*, *28*, 1087–1103. https://doi.org/10.1007/s10212-012-0157-y

Bashori, M., Van Hout, R., Strik, H., & Cucchiarini, C. (2021). Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. *System*, *99*,102496. https://doi.org/10.1016/j.system.2021.102496.

Cao, Y. (2011). Investigating situational willingness to communicate within second language classrooms from an ecological perspective. *System*, *39*(4), 468–479. https://doi.org/10.1016/j.system.2011.10.016.

Carter, R., & McCarthy, M. (2017). Spoken grammar: Where are we and where are we going? *Applied Linguistics*, *38*(1), 1–20. https://doi.org/10.1093/applin/amu080

Chang, Y., Lee, S., Wong, S. F., & Jeong, S. P. (2022). AI-powered learning application use and gratification: An integrative model. *Information Technology & People*, *35*(7), 2115–2139. https://doi.org/10.1108/ITP-09-2020-0632.

Chiu, T. K., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2023). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments*, 1–17. https://doi.org/10.1080/10494820.2023.2172044.

Clément, R., Baker, S. C., & MacIntyre, P. D. (2003). Willingness to communicate in a second language: The effects of context, norms, and vitality. *Journal of Language and Social Psychology*, *22*(2), 190–209. https://doi.org/10.1177/0261927X03022002003

Dai, Y., & Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, *36*(5–6), 861–884. https://doi.org/10.1080/09588221.2021.1952272

Deng, Y., Wen, K., Dusza, D. G., & Huang, H. W. (2024, March). AI-supported authentic communication with native speakers: Exploring EFL learners' willingness to communicate and emotional changes. In J. Mitrovic, J. Santiago, & H. Pasumarthi (Eds.), *Proceedings of the 2024 International Conference on Innovation in Artificial Intelligence* (pp. 59–64). Japan. https://doi.org/10.1145/3655497.3655530

Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. *Language Learning*, *53*, 3–32. https://doi.org/10.1111/1467-9922.53222.

Doosti, M., & Ahmadi Safa, M. (2021). Fairness in oral language assessment: Training raters and considering examinees' Expectations. *International Journal of Language Testing*, *11*(2), 64-90.

Ebadi, S., & Ebadijalal, M. (2022). The effect of Google Expeditions virtual reality on EFL learners' willingness to communicate and oral proficiency. *Computer Assisted Language Learning*, *35*(8), 1975–2000. https://doi.org/10.1080/09588221.2020.1854311

Fulcher, G. (2000). The 'communicative' legacy in language testing. *System*, *28*, 483–497. https://doi.org/10.1016/S0346-251X(00)00033-6

Fulcher, G. (2015). Assessing second language speaking. *Language Teaching*, *48*(2), 198–216. https://doi.org/10.1017/S0261444814000391

Girón-García, C., & Llopis-Moreno, C. (2015). Designing oral-based rubrics for oral language testing with undergraduate Spanish students in ESP contexts. *The Journal of Language Learning and Teaching*, *5*(2), 86–107.

Hamza, H. K. (2023). Barriers to mastery EFL communication skill in Iraqi governmental schools. *Journal of Language Studies*, *6*(2), 158–169. http://dx.doi.org/10.25130/jls.6.4.2.14

Isaacs, T. (2016). Assessing speaking. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 131–146). De Gruyter.

Jung Youn, S. (2023). Test design and validity evidence of interactive speaking assessment in the era of emerging technologies. *Language Testing*, *40*(1), 54–60. https://doi.org/10.1177/02655322221126606

Kan, M. S., & Ito, A. (2020). Language cognition and pronunciation training using applications. *Future Internet, 12*(3), 42–51. https://doi.org/10.3390/fi12030042

Khajavy, G. H., Ghonsooly, B., Hosseini Fatemi, A., & Choi, C. W. (2016). Willingness to communicate in English: A microsystem model in the Iranian EFL classroom context. *TESOL Quarterly*, *50*(1), 154–180. https://doi.org/10.1002/tesq.204.

Kim, A., & Su, Y. (2024). How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System*, *122*, 103256. https://doi.org/10.1016/j.system.2024.103256

Kuddus, K. (2022). Artificial intelligence in language learning: Practices and prospects. In A. Mire, S. Malik, & A. K. Tyagi (Eds.), *Advanced analytics and deep learning models* (pp. 3–18). Wiley.

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*(3), 634–654. https://doi.org/10.1111/j.1540-4781.2013.12030.x

Liakin, D., Cardoso, W., & Liakina, N. (2014). Learning L2 pronunciation with a mobile speech recognizer: French/y/. *CALICO Journal*, *32*(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

Ma, Y., Yu, S., Reynolds, B. L., & Jiang, L. (2022). A qualitative investigation of Chinese students' willingness to communicate in English in the graduate school EMI classroom. *English Teaching & Learning*, *46*(1), 77–98. https://doi.org/10.1007/s42321-021-00087-1

MacIntyre, P. D., Baker, S. C., Clément, R., & Conrod, S. (2001). Willingness to communicate, social support, and language-learning orientations of immersion students. *Studies in Second Language Acquisition*, *23*(3), 369–388. https://doi.org/10.1017/S0272263101003035.

MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, *82*(4), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

McNamara, T. (1996). *Measuring second language proficiency*. Longman.

Mehrgan, K. (2013). Willingness to communicate in second language acquisition: A case study from a socio-affective perspective. *Journal of Comparative Literature and Culture*, *2*(4), 172–175.

Penning de Vries, B. W., Cucchiarini, C., Strik, H., & Van Hout, R. (2020). Spoken grammar practice in CALL: The effect of corrective feedback and education level in adult L2 learning. *Language Teaching Research*, *24*(5), 714–735. https://doi.org/10.1177/1362168818819027

Phillips, T. M., Saleh, A., & Ozogul, G. (2023). An AI toolkit to support teacher reflection. *International Journal of Artificial Intelligence in Education*, *33*(3), 635–658. https://doi.org/10.1007/s40593-022-00295-1

Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: a probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, *14*(1), 49–70. https://doi.org/10.1186/s40468-024-00321-0

Sha, G. (2009). AI-based chatterbots and spoken English teaching: A critical analysis. *Computer Assisted Language Learning*, *22* (3), 269–281. https://doi.org/10.1080/09588220902920284.

Shafiee Rad, H. (2024). Revolutionizing L2 speaking proficiency, willingness to communicate, and perceptions through artificial intelligence: A case of Speeko application. *Innovation in Language Learning and Teaching*, *39* (3), 364–379. https://doi.org/10.1080/17501229.2024.2309539

Sharadgah, T. A., & Sa'di, R. A. (2022). A systematic review of research on the use of artificial intelligence in English language teaching and learning (2015-2021): What are the current effects? *Journal of Information Technology Education: Research*, *21*, 337–352. https://doi.org/10.28945/4999

Shorten, A., & Moorley, C. (2014). Selecting the sample. *Evidence Based Nursing*, *17*(2), 32–33. https://doi.org/10.1136/eb-2014-101747

Sun, Z., Anbarasan, M., & Praveen Kumar, D. (2021). Design of online intelligent English teaching platform based on artificial intelligence techniques. *Computational Intelligence, 37*(3), 1166–1180. https://doi.org/https://doi.org/10.1111/coin.12351

Tai, T. Y., Chen, H. H. J., & Todd, G. (2022). The impact of a virtual reality app on adolescent EFL learners' vocabulary learning. *Computer Assisted Language Learning*, *35*(4), 892–917. https://doi.org/10.1080/09588221.2020.1752735

Tai, T. Y., & Chen, H. H. J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, *31*(3), 1485–1502. https://doi.org/10.1080/10494820.2020.1841801

Wang, C. L., Wang, Y., Zeng, Z. Y., Lin, C. Y., & Yu, Q. L. (2021). Research on logistics distribution vehicle scheduling based on heuristic genetic algorithm. *Complexity*, *2021*(1), 8275714. https://doi.org/10.1155/2021/8275714

Wei, X., & Xu, Q. (2022). Predictors of willingness to communicate in a second language (L2 WTC): Toward an integrated L2 WTC model from the socio psychological perspective. *Foreign Language Annals*, *55*(1), 258–282. https://doi.org/10.1111/flan.12595.

Xi, X., Higgins, D., Zechner, K., & Williamson, D. (2012). A comparison of two scoring methods for an automated speech scoring system. *Language Testing*, *29*(3), 371–394. https://doi.org/10.1177/0265532211425673

Xiao, W., & Park, M. (2021). Using automatic speech recognition to facilitate English pronunciation assessment and learning in an EFL context: Pronunciation error diagnosis and pedagogical implications. *International Journal of Computer-Assisted Language Learning and Teaching*, 11(3), 74–91. https://doi.org/10.4018/IJCALLT.2021070105

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education*, *140*, 103599. https://doi.org/10.1016/j.compedu.2019.103599.

Yashima, T., Zenuk Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning*, *54*(1), 119–152. https://doi.org/10.1111/j.1467-9922.2004.00250.x

Yu, B. (2021). Research note: The role of willingness to communicate in cross-cultural Adaptation. *Intercultural Education 32*(2), 246–257. https://doi.org/10.1080/14675986.2021.1857115.

Zarrinabadi, N., & Khodarahmi, E. (2017). L2 willingness to communicate and perceived accent strength: A qualitative inquiry. *Journal of Intercultural Communication Research*, *46*(2), 173–187. https://doi.org/10.1080/17475759.2017.1301981.

Zhang, C., Meng, Y., & Ma, X. (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System*, *121*, 103259. https://doi.org/10.1016/j.system.2024.103259

Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *Sage Open*, *13*(3), 1–9. https://doi.org/10.1177/21582440231193818

Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. *System, 126*, 103497. https://doi.org/10.1016/j.system.2024.103497

Appendix A.
Speaking Assessment Rubrics

Criteria	1 (Poor)	2 (Fair)	3 (Good)	4 (Very Good)	5 (Excellent)
Vocabulary	Limited vocabulary, frequent misuse of basic words	Limited vocabulary, occasional use of appropriate words but often incorrect	Adequate range of vocabulary, using some relevant terms	Good range of vocabulary, using appropriate terms mostly correctly	Rich and varied vocabulary, using precise and contextually relevant terms effortlessly
Grammar	Frequent grammatical errors, meaning often unclear	Many grammatical errors, some confusing sentences	Some grammatical errors, overall clarity of meaning	Few grammatical errors, using mostly correct sentence structures	Excellent control of grammar, structurally sound and error-free sentences
Intonation	Monotone, lacking variety, not conveying meaning effectively	Minimal intonation, occasionally conveying meaning, but often flat	Generally good intonation, some variation, conveying emotions appropriately	Good use of intonation, effective variation enhancing communication	Excellent intonation, effectively emphasizing meaning and engaging the listener
Pronunciation	Frequent mispronunciations that hinder understanding.	Some mispronunciations occasionally hinder understanding	Generally clear pronunciation, some errors that do not affect understanding	Mostly clear pronunciation, rare errors, easily understandable speech,	Clear, precise pronunciation; consistently understandable and engaging speech
Fluency	Very hesitant, frequent pauses and interruptions, struggling to convey ideas	Hesitant speech, many pauses, incomplete or unclear ideas conveyed	Moderate fluency, some hesitations but generally conveying ideas	Good fluency, few pauses, conveying ideas clearly and coherently	Excellent fluency, smooth, confident speech with a natural flow of ideas

Appendix B.
Willingness to Communicate in a Foreign-Language Scale

	Please choose the one that best describes your idea.	Strongly Agree	Agree	No idea	Disagree	Strongly disagree
1	If I encountered some native speakers of English (British, American, Canadian, Australian) in the street, restaurant, hotel etc. I hope an opportunity would arise and they would talk to me.					
2	If I encountered some native speakers of English (British, American, Canadian, Australian) in the street, restaurant, hotel etc. I would find an excuse and would talk to them.					
3	If I encountered some native speakers of English (British, American, Canadian, Australian) who are facing problems in my country because of not knowing our language, I take advantage of this opportunity and would talk to them.					
4	I am willing to accompany some native speakers of English (British, American, Canadian, Australian) and be their tour guide for a day free of charge.					
5	I am willing to talk with native speakers of English (British, American, Canadian, Australian).					
6	Native speakers of English (British, American, Canadian, Australian) have interesting experiences that I would like to share.					
7	If someone introduced me to a native-speaker of English (British, American, Canadian, Australian) I would like to try my abilities in communicating with him/her in English.					
8	If I encountered some non-native speakers of English (Japanese, Pakistani, French, etc.) in the street, restaurant, hotel etc. I hope an opportunity would arise and they would talk to me.					
9	If I encountered some non-native speakers of English (Japanese, Pakistani, French, etc.) in the street, restaurant, hotel etc. I would find an excuse and would talk to them.					
10	If I encountered some non-native speakers of English (Japanese, Pakistani, French, etc.) who are facing problems in my country because of not knowing our					

	language I take advantage of this opportunity and would talk to them.
11	I am willing to accompany some non-native speakers of English (Japanese, Pakistani, French, etc.) and be their tour guide for a day free of charge.
12	I am willing to talk with non-native speakers of English (Japanese, Pakistani, French, etc.).
13	Non-native speakers of English (Japanese, Pakistani, French, etc.) have interesting experiences that I would like to share.
14	If someone introduced me to a non-native speaker of English (Japanese, Pakistani, French, etc.), I would like to try my abilities in communicating with him/her in English.
15	In order to practice my English, I am willing to talk in English with my classmates outside the class.
16	I am willing to ask questions in English in the classes at the university.
17	I am willing to talk and express my opinions in English in the class when all my classmates are listening to me.
18	I am willing to have pair and group activities in the class so that I can talk in English with my classmates.
19	In order to practice my English I am willing to talk in English with my professors outside the class.
20	I am willing to give a presentation in English in front of my classmates.
21	In group work activities in the class when the group is composed of my friends, I am willing to speak in English.
22	In group work activities in the class when the group is NOT composed of my friends, I am willing to speak in English.