Vol. 15, No. 04, pp. 2760-2780, 2025 DOI: 10.23960/jpp.v15i4.pp2760-2780

Jurnal Pendidikan Progresif

e-ISSN: 2550-1313 | p-ISSN: 2087-9849

http://jpp.fkip.unila.ac.id/index.php/jpp/

Bridging the Affective-Linguistic Gap: A Mixed-Methods Exploration of AI-Assisted Speaking Practice and Willingness to Communicate

Ratna Yunita[1,*], Candra Febrilyantri[2], Aya Hayasaki[3], & Muh Imdaadun Ni'am[4]

[1]Department of English Education, UIN Kiai Ageng Muhammad Besari Ponorogo, Indonesia
[2]Department of Sharia Accounting, UIN Kiai Ageng Muhammad Besari Ponorogo, Indonesia
[3]Center for Secondary and Higher Education Linkage, Institute for Comprehensive Education, Kagoshima University, Japan
[4]Department of Sharia Economics, UIN Kiai Ageng Muhammad Besari Ponorogo, Indonesia

*Corresponding email: ratnayunita@uinponorogo.ac.id

Received: 15 November 2025 Accepted: 20 December 2025 Published: 29 December 2025

Abstract: Bridging the Affective-Linguistic Gap: A Mixed-Methods Exploration of AI-Assisted Speaking Practice and Willingness to Communicate. Objectives: This study examined the effectiveness of AI-assisted learning in enhancing students' Willingness to Communicate (WTC) and speaking proficiency in an Indonesian higher education context. It also explored students' perceptions of AI-based tools' support for communicative confidence and readiness during speaking practice. **Methods:** A convergent mixed-methods design was employed with 60 undergraduate students enrolled in a Business English course. The experimental group received six AI-assisted speaking sessions (100 minutes each) using the SmallTalk2Me platform, while the control group received conventional instruction. WTC and speaking proficiency were measured using parallel pre- and post-tests, and qualitative data were collected through open-ended reflection surveys. Quantitative data were analyzed using descriptive statistics and two-way mixed ANOVA, and qualitative data were analyzed thematically. **Findings:** Quantitative results showed a significant main effect of time, indicating that students in both groups improved their speaking proficiency and communicative readiness across the semester. However, neither the group effect nor the interaction effect reached significance, suggesting that AI-assisted practice did not produce statistically greater gains than conventional teaching. Qualitative findings, however, revealed perceived affective benefits among AI users, including reduced speaking anxiety, increased confidence, and appreciation of a low-pressure practice environment with instant feedback and opportunities for repetition. Some students also reported challenges, including occasional misrecognition in AI feedback, dependence on stable internet access, and interactions that felt less natural than human communication. **Conclusion:** The findings indicate that AI-assisted learning did not yield statistically greater gains in speaking proficiency or WTC than conventional instruction. Although qualitative data suggest that some learners perceived changes in their affective experiences during AI-assisted speaking practice, these perceived advantages did not translate into statistically superior outcomes. Overall, the results indicate that the pedagogical impact of AI-assisted learning remains limited within the scope of the present study.

Keywords: AI-assisted learning, higher education, language learning, speaking proficiency, willingness to communicate.

To cite this article:

Yunita, R., Febrilyantri, C., Hayasaki, A., & Ni'am, M. I. (2025). Bridging the Affective-Linguistic Gap: A Mixed-Methods Exploration of AI-Assisted Speaking Practice and Willingness to Communicate. *Jurnal Pendidikan Progresif*, *15*(4), 2760-2780. doi: 10.23960/jpp.v15i4.pp2760-2780.

INTRODUCTION

Developing students' oral communication skills remains a central goal in English as a Foreign Language (EFL) instruction. However, speaking remains one of the most challenging skills for learners in many Asian contexts, including Indonesia. A substantial body of research has documented persistent issues such as speaking anxiety, low communicative confidence, fear of negative evaluation, and limited opportunities for authentic oral interaction, all of which constrain learners' willingness to engage in communicative tasks. These affective and contextual barriers are closely aligned with the construct of Willingness to Communicate (WTC), defined as an individual's readiness to initiate communication under particular situational conditions (MacIntyre et al., 1998). In Indonesia specifically, numerous studies demonstrate that university students experience communication apprehension and reticence in classroom speaking tasks, resulting in limited WTC and limited development of speaking proficiency (Astuti, 2016; Zulfikar et al., 2019; Nazri, 2025; Ambawani et al., 2025). These constraints are often more pronounced in English for Specific Purposes (ESP) contexts such as Economics and Business where students must perform discipline-specific communicative tasks despite having fewer opportunities for oral interaction.

Recent advances in artificial intelligence (AI) have opened new possibilities for addressing these long-standing challenges. With the rise of automatic speech recognition (ASR), conversational agents, and intelligent tutoring systems, language learners now have access to tools that provide individualized feedback, expanded practice opportunities, and reduced psychological pressure compared to traditional classroom environments (Bashori et al., 2021; Fathi et al., 2024; Zhang et al., 2024). AI-assisted speaking environments have demonstrated promising effects on affective variables, such as confidence, enjoyment, and reduced anxiety, which are central predictors of WTC and oral performance (Greenier et al., 2021; Dewaele, 2021). Systematic reviews similarly indicate that AI-based learning platforms offer low-pressure, personalized, and repeatable speaking experiences conducive to autonomous learning, although challenges remain concerning emotional responsiveness and feedback accuracy (Wiboolyasarin et al., 2025). Together, these developments suggest that AI-mediated learning environments may serve as psychologically supportive spaces that foster both affective readiness and communicative competence.

Despite these advances, important gaps persist within the intersection of WTC theory, AI-assisted learning, and ESP instruction. First, although research increasingly investigates AI-based tools for improving pronunciation, fluency, or accuracy, studies rarely integrate AI-mediated speaking practice with the theoretical layers of WTC, particularly its affective antecedents such as perceived competence, communication anxiety, and situational confidence. As Peng (2025) emphasizes, WTC is shaped by dynamic psychological processes rather than merely linguistic performance factors. However, most AI-focused studies conceptualize technology as a performance enhancer, overlooking the psychological mechanisms that may explain why learners become more willing (or unwilling) to communicate when supported by AI. This theoretical gap limits our understanding of how AI tools interact with the affective foundations of spoken communication.

Second, AI-supported speaking practice in ESP contexts remains underexplored. ESP learners in Economics, Business, and related fields face unique communicative pressures, including discipline-specific discourse expectations, professional terminology, and culturally informed interactional norms. Prior research shows that ESP learners often experience higher anxiety during professional communication tasks and require specialized support that differs from general EFL learners (Dafouz & López-Serrano, 2026). However, most existing AI-speech studies have been conducted with general EFL populations (e.g., Bashori et al., 2021; Zhang et al., 2024), leaving unanswered questions about how AI-mediated practice influences WTC and speaking proficiency in discipline-specific settings. Considering the rapid expansion of AI in higher education, investigating its pedagogical potential within ESP programs represents a timely and necessary advancement.

Third, most existing research on WTC or AI-assisted speaking adopts a quantitative or performance-focused design, with fewer studies integrating qualitative insights to explain the underlying processes that shape learner outcomes (e.g., Fathi et al., 2024; Wiboolyasarin et al., 2025). Scholars have recently called for mixed-methods approaches that illuminate the affective, behavioral, and experiential mechanisms through which AI tools support L2 speaking development. Without such integration, the field risks overlooking critical learner-centered perspectives, especially regarding how AI influences self-confidence, anxiety management, self-regulation, and readiness for spontaneous communication.

The present study advances international scholarship on WTC and AI-assisted language learning through four main theoretical contributions. First, it extends MacIntyre et al.'s (1998) WTC model into AI-mediated speaking environments by examining how AI features—such as instant feedback, repetition, and non-judgmental interaction—shape the psychological antecedents of WTC. This represents an initial extension, as research has rarely explored AI as an affective catalyst rather than solely as a linguistic tool. Second, the study bridges the affective–linguistic gap in existing AI research by suggesting a tentative developmental sequence in which AI-assisted practice first supports affective gains (e.g., anxiety reduction, confidence building, perceived competence), which may subsequently facilitate gradual improvements in speaking proficiency. This suggests a potential mediational mechanism that is not yet well established in the literature. Third, the study contributes to ESP scholarship by examining WTC within an Economics and Business context. In this area, communicative demands differ significantly from those in general EFL settings but remain underrepresented in AI-supported research. Fourth, the study employs a convergent mixed-methods design that integrates quantitative performance measures with qualitative evidence of affective and experiential phenomena. This methodological contribution aligns with recent calls for deeper, theoretically grounded analyses of WTC dynamics in technology-enhanced environments.

Guided by these identified gaps and theoretical contributions, the study examines the effectiveness of AI-assisted learning in enhancing students' WTC and speaking proficiency within an Indonesian ESP (Economics and Business) context. Through a mixed-methods approach, the study analyzes quantitative improvements in WTC and speaking performance as well as qualitative insights into students' perceptions of AI-based practice, affective experiences, and challenges encountered during AI-supported speaking activities. Taken collectively, these perspectives aim to provide a comprehensive understanding of how AI tools influence both the psychological and linguistic dimensions of oral communication.

Three research questions were formulated to guide the inquiry. The first research question (RQ 1) examines whether AI-assisted learning improves students' WTC and speaking proficiency compared with traditional instruction. Two quantitative hypotheses accompany this question: H1a predicts that both experimental and control groups will show improvement over time, while H1b hypothesizes that the AI-assisted group will experience greater gains. The second research question (RQ2) explores how students perceive the role of AI tools in supporting confidence, reducing communication anxiety, and enhancing readiness to speak. The third research question (RQ3) examines the challenges students face when using AI for speaking practice, including technical limitations, issues with automated feedback, and the emotional nuances of interacting with AI systems.

By addressing these research questions, the present study aims to advance theoretical understanding of WTC in AI-mediated learning environments, contribute practical insights for integrating AI into ESP speaking courses, and provide a methodologically robust framework for examining the affective mechanisms underlying learners' oral communication development.

■ METHOD

Participants

The participants of this study were 60 undergraduate students enrolled in the Faculty of Islamic Economics and Business at Universitas Islam Negeri Kiai Ageng Muhammad Besari, Ponorogo. They were drawn from intact classes representing three programs: Islamic Zakat and Waqf Management, Islamic Business Management, and Islamic Banking, to reflect the actual distribution of students in the faculty. Two intact classes were purposively selected to form the comparison groups, with 30 students assigned to the experimental group and 30 to the control group. The use of intact classes is consistent with the logic of quasi-experimental research in higher education, where random reassignment is often impractical and may disrupt academic schedules.

An additional demographic variable included in this study was participants' initial English proficiency. This was measured using a pre-test of speaking proficiency based on an IELTS-aligned rubric. Pre-test scores ranged from 5.0 to 8.0, indicating that most students began the study at the intermediate to upper-intermediate level. The distributions of scores were comparable across groups, with the control group scoring between 6.0 and 8.0 and the experimental group between 5.0 and 8.0, providing a relatively balanced baseline before the intervention.

Inclusion criteria required that students complete all components of the study, including the pre-tests and post-tests for both WTC and speaking proficiency, as well as the qualitative reflection survey. All participants voluntarily agreed to participate in the study, provided written consent, and were assured that their participation would not affect their academic evaluation. The sample included both male and female students. It represented varying specializations within Islamic Economics and Business programs, increasing the ecological validity and representativeness of the findings for Islamic higher education contexts.

Research Design and Procedures

This study adopted a quasi-experimental pretest–posttest control-group design within a convergent mixed-methods framework. The design was intended to evaluate the effects of AI-assisted learning on students' willingness to communicate and speaking proficiency over the course of one academic semester (January–June 2025). The independent variable was the mode of instruction, contrasting an AI-assisted intervention and traditional teacher-led instruction. The dependent variables were students' WTC and their English-speaking proficiency.

The experimental group received AI-supported speaking practice using SmallTalk2Me, an adaptive platform that provides instant feedback on fluency, accuracy, pronunciation, and vocabulary. Speaking prompts were automatically generated by SmallTalk2Me and focused on Business English topics, including marketing decisions, workplace communication, supply chain, and customer service challenges. All participants received the same prompts for both the pre-test and post-test to ensure consistency.

The intervention lasted six meetings, every 100 minutes, structured as follows:

1. 10–15 minutes: class opening and orientation on the Business English topic
2. 30–40 minutes: speaking practice using SmallTalk2Me
3. 20–25 minutes: analysis of AI feedback (pronunciation, fluency, grammar)
4. 25 minutes: extended practice through roleplay or simulation based on AI feedback.

The control group received conventional instruction through role-plays, pair discussions, presentations, and teacher-delivered feedback, without AI support. Both groups covered the same English for Specific Academic Purposes (ESAP) content relevant to Islamic Economics and Business, ensuring curricular equivalence while isolating differences in feedback modality.

Data collection occurred in two strands. In the quantitative strand, students completed WTC and speaking pre-tests in Week 1 and equivalent post-tests in Week 14. The speaking test followed an IELTS-style structure comprising 11 prompts across personal, narrative, and abstract discussion tasks. The post-test used parallel prompts to preserve measurement consistency while minimizing recall. In the qualitative strand, students completed a structured Google Forms survey with seven open-ended questions concerning their learning experiences, perceptions of AI-assisted or traditional instruction, encountered challenges, and suggestions for improvement. The use of Google Forms replaced the initial plan for interviews due to practical advantages, such as full participation, reduced scheduling conflicts, anonymity, and reduced interviewer bias. Quantitative and qualitative data were collected concurrently, analyzed separately, and then integrated during interpretation following the logic of a convergent mixed-methods design.

Instruments

Two primary instruments were used in the quantitative component: the WTC scale and an analytic speaking proficiency measure generated through the SmallTalk2Me platform.

The WTC scale was adapted from the widely used instruments developed by McCroskey and grounded in MacIntyre's WTC framework. The questionnaire consisted of 20 items administered in Bahasa Indonesia, each representing a different communicative situation in which students rated their willingness to communicate in English. Students indicated their responses on a 0–100 probability scale (0 = Never, 100 = Always), allowing fine-grained measurement of their communicative readiness across varying interpersonal and public-speaking contexts. Sample items include statements such as "giving a presentation to a group of strangers" (Item 3), "speaking in a small group with acquaintances" (Item 15), and "speaking in a large meeting with friends" (Item 6), representing a range of interpersonal and public-speaking contexts. The scale's reliability was examined in this study, yielding a Cronbach's á of .856, indicating strong internal consistency.

Speaking proficiency was assessed using SmallTalk2Me's automated analytic scoring system, which evaluates fluency, accuracy, vocabulary use, and pronunciation through machine-learning–based speech analysis. All pre-tests and post-tests for both groups were scored using the same platform, ensuring complete consistency. Because the platform produces a single composite speaking score, internal consistency indices such as Cronbach's á cannot be computed. Nevertheless, prior validation work has demonstrated that SmallTalk2Me yields stable and discriminative performance across learner proficiency levels. For instance, in an activity theory–based investigation, Ebadi et al. (2025) reported that the platform's construct validity had been examined using a group-differential method, indicating its capacity to differentiate proficiency levels among EFL learners.

SmallTalk2Me uses the official IELTS Speaking Band Descriptors, fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation, to generate CEFR-aligned analytic scores (IELTS Partners, 2025; SmallTalk2Me, 2025). A recent study by Maulidianti et al. (2024) further supports the platform's practical validity, demonstrating that automated scores align with expected proficiency gains and provide pedagogically meaningful feedback for non-native speakers. While the present study did not include human expert ratings to examine convergent validity, the use of a standardized scoring algorithm, documented construct validation procedures, and evidence from prior EFL research support the appropriateness of SmallTalk2Me as an assessment tool in this context. Future studies may incorporate independent human ratings to further evaluate the system's performance across diverse second-language (L2) accents.

It is important to acknowledge that using SmallTalk2Me as both the intervention tool for the experimental group and the speaking assessment instrument for both groups may raise concerns about potential measurement-familiarity effects. All speaking assessments were conducted using identical prompts, scoring algorithms, and procedures across groups, ensuring measurement equivalence. Moreover, the automated scoring system relies on standardized acoustic, lexical, and temporal features rather than task-specific memorization, reducing the likelihood that observed differences reflect test familiarity alone. Nevertheless, students in the experimental group may have developed greater procedural familiarity with the platform interface, which is therefore treated as a methodological limitation and further considered in the discussion when interpreting the non-significant interaction effects.

The AI-based intervention provided real-time feedback on pronunciation, fluency, grammar, and overall speaking performance. Students in the experimental group used the tool during in-class speaking activities and were free to practice independently outside scheduled class time.

Students in the control group participated in conventional speaking instruction during a 100-minute class session, following a standard lesson structure of an introduction, guided practice, and speaking tasks. The instructor provided feedback targeting the same linguistic dimensions assessed by the AI tool—fluency, accuracy, vocabulary use, and pronunciation. Due to time constraints and the need to accommodate all learners, this feedback was delivered selectively rather than individually after every performance. Immediate feedback was typically given following each speaking activity, with the instructor addressing recurring errors observed across multiple students and offering model corrections for frequent or shared issues. Individual feedback was provided when feasible, particularly when a student repeated prominent errors, although comprehensive 1:1 feedback for every learner was not possible within the allotted class time. Thus, while both groups received feedback covering equivalent linguistic aspects, the experimental group benefited from continuous, instant, and scalable AI feedback. In contrast, the control group relied on human instructor feedback embedded within standard classroom interaction.

Data Analysis

Quantitative data were analyzed using SPSS. Descriptive statistics were calculated to summarize pre- and post-test performance for both WTC and speaking proficiency. Prior to inferential tests, assumptions of normality and homogeneity of variance were examined using Kolmogorov–Smirnov, Shapiro–Wilk, and Levene's tests. Because the data satisfied these assumptions, two-way mixed ANOVA (Group × Time) was used to evaluate the effects of instructional mode and test time, as well as their interaction. Effect sizes (partial ç²) were reported to assess the practical significance of findings.

Qualitative data were analyzed through thematic analysis following Braun and Clarke's six-phase model. Responses were coded inductively to identify emerging patterns related to learning experiences, perceived effectiveness of feedback mechanisms, anxiety reduction, challenges, and personal preferences. Themes were then compared with quantitative results to identify areas of convergence, complementarity, or divergence, consistent with the principles of convergent mixed-methods design.

Ethical Considerations

Ethical approval for the study was obtained from the university's research ethics committee prior to data collection. All participants signed informed consent forms and were informed that participation was voluntary and unrelated to course grading. No personal identifiers were collected, and all responses were anonymized prior to analysis. Participants were informed that they could withdraw at any stage without penalty. All data were stored securely and used exclusively for research purposes in accordance with institutional and international ethical guidelines.

RESULT AND DISCUSSION

This section presents the findings on the effects of AI-assisted learning on students' WTC and speaking proficiency. Quantitative and qualitative results are integrated to explain how the intervention shaped linguistic performance, affective readiness, and learners' experiences with the AI tool.

Effectiveness of AI-Assisted Learning in Improving Students' Willingness to Communicate (WTC) and Speaking Proficiency

Both groups showed improvement in WTC over time, but the pattern of change reveals meaningful nuances regarding how students engaged with the intervention. Descriptive statistics for the pre- and post-test scores are presented in Table 1, while Figure 1 illustrates the distributional patterns across groups. The boxplot shows that students in the experimental group experienced a more pronounced upward shift in their WTC distribution. The post-test median rose more than in the control group, and the score spread suggests that several learners experienced substantial individual gains. In contrast, the control group's median increased only slightly, and the distribution remained comparatively stable, indicating more modest improvement.

Table 1. Mean scores of the students' WTC (pre-test and post-test)

Group	Pre-Test	Post-Test	Gain
Experimental	56.40	66.87	+10.47
Control	51.10	62.50	+11.40

These descriptive patterns align with the statistical analysis. The two-way mixed ANOVA showed a significant main effect of time, $F(1,116) = 14.642$, $p < .001$, ç² = .11, demonstrating that students in both groups became more willing to communicate over the semester. However, neither the main effect of group ($p = .093$) nor the interaction effect ($p = .871$) reached significance. Thus, although the experimental group showed a more favorable distributional shift, the AI-assisted intervention did not yield statistically greater gains than traditional instruction at the group level.

Understanding this discrepancy requires careful consideration of the characteristics of WTC and the learning context. WTC is a situational and affectively mediated construct, shaped by moment-to-moment fluctuations in confidence, anxiety, and perceived competence (MacIntyre et al., 1998; MacIntyre, 2007). At the same time, the AI tool may have created a psychologically safe space that encouraged individual learners to take risks, leading to affective benefits that often manifest unevenly across participants. As a result, improvements may appear at the individual level visible in the widened upper quartile yet remain insufficiently large or consistent to produce statistically significant group-level differences within a short intervention period.

Contextual factors common in Indonesian university settings may also help explain the lack of differential gains. Cultural tendencies toward reticence, fear of negative evaluation, and limited exposure to authentic English communication (Anggrisia & Robah, 2023; Astuti, 2016; Zulfikar et al., 2019) can inhibit learners' willingness to speak regardless of the instructional method. Such deeply rooted sociocultural influences typically require long-term, sustained intervention to overcome. Consequently, while AI-assisted practice offered practical benefits such as reduced anxiety and increased comfort these advantages may not have been strong or widespread enough to yield statistically distinguishable outcomes over just six instructional sessions.

Despite the lack of significant interaction effects, qualitative data reveal meaningful insights

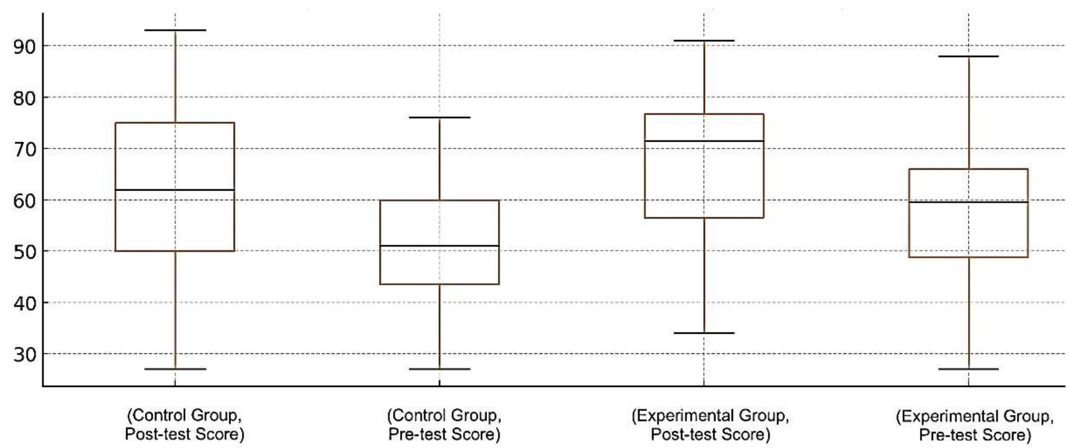

Figure 1. Boxplot of WTC scores for control and experimental groups (pre- and post-test)

Table 2. Results of the two-way mixed ANOVA test for WTC scores

Tests of Between-Subjects Effects					
Dependent Variable: WTC Results					
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Corrected Model	4293.500[a]	3	1431.167	5.843	.001
Intercept	420793.633	1	420793.633	1718.077	.000
Class	700.833	1	700.833	2.861	.093
PrePos	3586.133	1	3586.133	14.642	.000
Class * PrePost	6.533	1	6.533	.027	.871
Error	28410.867	116	244.921		
Total	453498.000	120			
Corrected Total	32704.367	119			

a. R Squared = .131 (Adjusted R Squared = .109)'

Source	F	Sig.	Partial η²	Interpretation
Class	2.861	.093	.02	Small effect
PrePost	14.642	.000*	.11	Moderate effect
Class* PrePost	.027	.871	.00	No effect

Note. *$p < .05$ indicates significance. Partial η^2 = effect size measure (small ≈ .01, moderate ≈ .06, large ≈ .14).*

into how students experienced the AI-assisted practice. Many students described the AI environment as *"a safe place to practice," "non-judgmental,"* and *"more comfortable than speaking in front of classmates."* These perceptions align with the WTC framework (MacIntyre et al., 1998), which posits that reduced anxiety and increased perceived competence facilitate immediate willingness to communicate. Students also highlighted the value of AI's instant corrective feedback, noting that being able to *"fix mistakes before practicing again"* made them feel more prepared for in-class speaking.

However, these perceived advantages must be interpreted with caution. Although students reported feeling more confident and supported, such affective benefits did not translate into statistically significant group-level gains. This discrepancy suggests that the advantages of AI-assisted practice may operate primarily at the individual, micro-affective level, improving comfort and confidence for some learners but not producing large or consistent enough improvements across the whole group. Moreover, because WTC is highly situational and influenced by fluctuating psychological states, short-term gains in perceived safety or confidence may not accumulate into robust, measurable changes within a six-session intervention. In this sense, the qualitative responses offer valuable insight into learner experience, but they do not necessarily predict broad, statistically detectable outcomes.

These findings resonate with several digital learning studies. Khalik (2025) reported that chatbot-assisted preparation reduces speaking anxiety, while Panggua et al. (2025) found that AI-enhanced speaking tools foster a nonjudgmental environment that increases learner autonomy and confidence. The parallels are not merely descriptive; they point toward a shared psychological mechanism. Across studies, AI tools appear to reduce perceived social-evaluative threat, enabling learners to experiment with language without fear of embarrassment. This reduction in threat lowers affective filters and increases learners' sense of agency, which explains why students consistently feel more prepared and more confident when using AI tools, even when such improvements do not always manifest as significant statistical differences in group-level performance.

Furthermore, qualitative responses in this study also revealed a set of negative or ambivalent experiences, which may help explain the limited quantitative effects. Some students reported feeling pressured by continuous automated feedback, uncertainty about the accuracy of AI-generated suggestions, and a sense of "*speaking to a machine*," which over time reduced their motivation. These mixed reactions indicate that AI-assisted environments may simultaneously lower anxiety for some learners while introducing new cognitive or emotional challenges for others. Such variability in learner experience may further contribute to the absence of large, statistically detectable differences between groups.

Parallel patterns appeared in the speaking proficiency outcomes. Although the experimental group narrowed the initial performance gap and demonstrated meaningful individual progress, the two-way mixed ANOVA again showed a significant main effect of time, $F(1,116) = 5.852$, $p = .017$, $ç^2 = .05$, but nonsignificant group ($p = .115$) and interaction effects ($p = .449$). These results indicate that speaking proficiency improved across the semester for both groups, yet the AI-assisted practice did not produce statistically superior gains at the group level. This pattern suggests that while AI tools may enhance perceived preparedness and lower affective barriers, such benefits may not translate into uniform improvements in linguistic performance within a relatively short intervention period. Instead, the underlying psychological mechanisms, such as reduced social pressure, increased autonomy, and, for some learners, occasional frustration or uncertainty, likely contributed to varied individual trajectories, thereby limiting the emergence of statistically significant between-group differences.

Figure 2 presents the distribution of pre- and post-test speaking scores across groups. The experimental group's post-test distribution shows a modest upward shift and reduced variance, suggesting improvements in overall speaking proficiency for several learners. Meanwhile, the control group's distribution remained relatively stable, with minimal changes in spread or median values. Taken together, these descriptive trends mirror the WTC results: meaningful individual improvements occurred in the AI group, but the magnitude and consistency of these gains were insufficient to yield significant between-group differences.

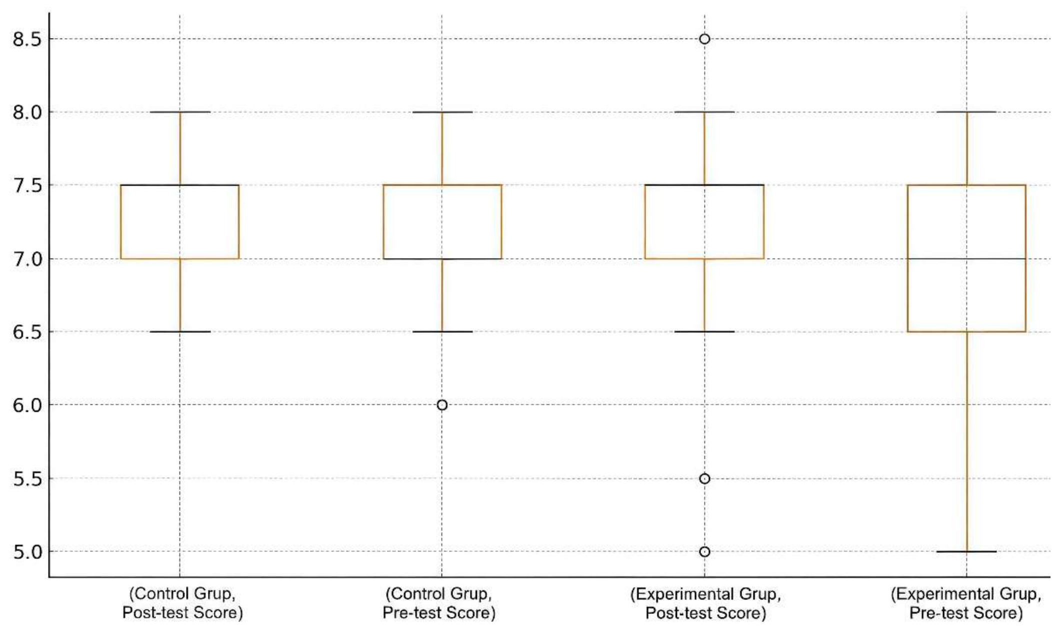

Figure 2. Boxplot of WTC scores for control and experimental groups (pre- and post-test)

Table 3. Results of the two-way mixed ANOVA test for speaking scores

Tests of Between-Subjects Effects					
Dependent Variable: Speaking Results					
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Corrected Model	390.625[a]	3	130.208	2.986	.034
Intercept	608475.208	1	608475.208	13951.532	.000
Class	110.208	1	110.208	2.527	.115

PrePost	255.208	1	255.208	5.852	.017
Class * PrePost	25.208	1	25.208	.578	.449
Error	5059.167	116	43.614		
Total	613925.000	120			
Corrected Total	5449.792	119			
a. R Squared = .072 (Adjusted R Squared = .048)					

Source	F	Sig.	Partial η^2	Interpretation
Class	2.527	.115	.02	Small effect
PrePost	5.852	.017*	.05	Moderate effect
Class*PrePost	.578	.449	.01	Negligible effect

Note. *p < .05 indicates significance. Partial ç^2 = effect size measure (small H” .01, moderate H” .06, large H” .14).*

Qualitative reflections further enrich and complicate the interpretation of these findings. Many students reported increased fluency, clarity, and response speed after practicing with AI. Examples included: *“I can repeat the speaking tasks until I am satisfied. This makes me more fluent,”* and *“The AI corrected my pronunciation immediately, so I learned faster than waiting for feedback in class.”* Other students noted heightened readiness for in-class participation: *“I feel more ready to speak in class now because I practiced with AI before coming to class.”* These perceptions are consistent with research showing that AI-assisted tools support fluency development and reduce hesitation through repeated practice and instant feedback (Bashori et al., 2021; Shen et al., 2025; Zhang et al., 2024).

However, these positive perceptions must be reconciled with the nonsignificant quantitative findings. The qualitative evidence suggests that AI-assisted practice primarily strengthened affective factors, such as confidence, reduced anxiety, and perceived communicative safety, which serve as essential foundations for speaking development but do not necessarily translate into immediate, uniform linguistic gains across learners. Affective gains tend to emerge quickly, whereas measurable linguistic gains (e.g., accuracy, pronunciation, lexical control) often require longer, sustained exposure before becoming statistically detectable. This layered progression helps explain why learners perceived substantial benefit even though group-level proficiency scores improved at similar rates across conditions.

One critical factor that may help explain the absence of statistically significant interaction effects lies in the limitations of the AI-based assessment instrument itself. As reflected in students’ qualitative feedback, SmallTalk2Me occasionally struggled to accurately recognize Indonesian-accented English, leading learners to perceive feedback as inconsistent or inaccurate (Del Río et al., 2023; Nakatumba-Nabende et al., 2025). Such misrecognition may have led to corrective feedback that did not fully align with students’ actual communicative intent or performance quality.

Inaccurate or unstable feedback can undermine the pedagogical advantages of AI-assisted learning. Rather than reinforcing improvement, inconsistent scoring may increase cognitive load, generate frustration, and reduce learners’ trust in the system (Alshehri, 2025; Safar & Anggraheni, 2024). For some students, repeated exposure to questionable feedback led to hesitation and decreased motivation, potentially neutralizing the theoretical benefits of instant and individualized correction. These instrument-related constraints provide a plausible explanation for why the affective advantages reported qualitatively, such as reduced anxiety and increased confidence, did not translate into statistically significant group-level gains in speaking proficiency. This familiarity effect may partially explain why no statistically significant interaction effect emerged, as procedural familiarity does not necessarily translate into broader communicative competence.

Another important consideration is that because the AI system evaluates speech primarily through acoustic and algorithmic parameters, it is less sensitive to accent variation and phonological deviation, which may lead to misrecognition and unstable feedback compared to human evaluators (Del Río et al., 2023). As a result, learners may have received feedback that prioritized formal accuracy over communicative effectiveness, limiting the development of higher-level speaking competence. When considered together, these limitations suggest that the non-significant interaction effects observed in the ANOVA should be interpreted not as evidence of AI ineffectiveness, but as an indication that current AI feedback systems may not yet fully operationalize their theoretical advantages in diverse EFL contexts.

Alongside these limitations, students expressed negative or ambivalent reactions that may have tempered the overall impact of the AI tool. Some users reported feeling pressured by continuous automated feedback, doubting the accuracy of AI suggestions, or becoming less motivated over time because “speaking to a machine” felt unnatural. Such reactions introduce variability in engagement and may dilute the potential for consistent group-level effects. These patterns mirror broader findings in AI-assisted language learning, where positive affective effects are frequently accompanied by challenges such as cognitive load (Safar & Anggraheni, 2024), mixed perceptions of feedback accuracy and fluctuating trust (Alshehri, 2025), complex emotional responses and uneven engagement (Dong et al., 2026), and occasional decreases in intrinsic motivation over time (Ahmed et al., 2025). Together, these studies suggest that while AI tools can effectively reduce anxiety and offer supportive practice environments, they may simultaneously introduce new emotional or cognitive challenges that produce uneven learning trajectories across students.

These results also align with dynamics commonly observed in Islamic higher education contexts, where large class sizes, limited speaking opportunities, and cultural tendencies toward reticence shape participation patterns. Anggrisia (2023) found that shyness and anxiety strongly constrain oral participation, while Atifnigar et al. (2022) identified class size as a significant barrier to spoken interaction. Technology-enhanced tasks have been shown to mitigate some of these barriers; for example, Ridayani and Purwanto (2024) reported that multimedia-supported role-play offers a lower-pressure environment for practicing speaking. The current findings echo these patterns: AI tools provide a psychologically safer space but cannot, within a short intervention, overcome broader structural and sociocultural constraints.

To synthesize the findings, Table 4 integrates quantitative outcomes with qualitative excerpts, demonstrating how numerical patterns align with learners’ experiences while also explaining the variability across participants.

This triangulated evidence strengthens the conclusion that the AI-supported intervention exerted its most robust influence in the affective domain, which in turn created the conditions for gradual improvement in speaking proficiency. Across learners, reductions in anxiety, enhanced perceived communicative safety, and increased confidence emerged as key psychological shifts

Table 4. Triangulation of quantitative and qualitative findings

Variable	Quantitative Findings	Qualitative Evidence (Representative Excerpts)	Integrated Interpretation
WTC	Significant improvement over time, $F(1,116) = 14.642$, $p < .001$, $\eta^2 = .11$. Higher post-test median in the experimental group, but nonsignificant group and interaction effects.	*"I feel less nervous because the AI gives feedback without judging me."* *"Practicing with AI makes me more confident to speak in class."* *"I can try again and again until I am ready—no one laughs at me."* *"Sometimes I still feel anxious when speaking spontaneously, even after AI practice."*	AI consistently reduced anxiety and increased perceived communicative safety, strengthening situational WTC. However, gains varied across learners some experienced large affective benefits, while others continued to struggle with spontaneous communication. This variability explains why individual improvements did not translate into statistically significant group-level differences.
Speaking Proficiency	Significant improvement over time, $F(1,116) = 5.852$, $p = .017$, $\eta^2 = .05$. Both groups improved similarly; no significant group or interaction effects.	*"Repeating the tasks made my speaking smoother and faster."* *"The AI corrected my pronunciation immediately, so I learned faster."* *"Sometimes I did not trust the feedback—it felt different from my lecturer's explanation."* *"Speaking to a machine feels unnatural, so my motivation went down after a while."* *"I became more aware of grammar and vocabulary mistakes because AI pointed them out."*	AI tools supported improvements in fluency, pronunciation, and self-monitoring, but the short intervention and mixed learner engagement limited measurable group-level effects. Positive affective changes (confidence, reduced anxiety) emerged quickly, while linguistic accuracy and lexical control require sustained practice to show stronger statistical differences. Negative or ambivalent reactions (feedback pressure, trust issues, reduced motivation) introduced further variability, tempering the overall impact.

that preceded and facilitated linguistic growth. However, the magnitude and stability of these affective gains varied across individuals, reflecting the inherently dynamic and situational nature of both WTC and speaking behavior. Students who experienced a strong sense of safety and autonomy tended to show noticeable improvements in fluency, accuracy, and responsiveness. In contrast, those who struggled with technological uncertainty, feedback pressure, or reduced motivation progressed more slowly.

These layered patterns point to a mechanism of indirect influence: AI tools do not automatically enhance proficiency; rather, they shape emotional readiness and self-regulatory beliefs, which then open the space for linguistic development. This helps explain why affective improvements emerged more quickly and consistently than changes in measurable performance—and why group-level differences in proficiency remained nonsignificant despite strong qualitative endorsements of the AI environment. The literature similarly indicates that AI-assisted environments primarily enhance enjoyment, engagement, and confidence, which subsequently mediate learning outcomes (Dong et al., 2026; Safar & Anggraheni, 2024; Ahmed et al., 2025). These mediating effects generally require extended, sustained practice before producing statistically detectable gains in accuracy, pronunciation control, or lexical sophistication.

The pattern observed here also aligns with broader findings in Indonesian higher education, where large class sizes, limited communicative exposure, and sociocultural reticence toward public speaking create substantial barriers to oral participation (Anggrisia, 2023). In such contexts, improvements in affective readiness are particularly consequential, as they help learners overcome longstanding psychological constraints. AI tools, by providing private, repeatable, and nonjudgmental practice, appear to partially counteract these challenges. However, learners' trust in the technology moderates their impact, tolerance for automated feedback, and the cognitive effort required to navigate the tool. Notably, recent studies in Indonesian higher education also show that AI-based learning analytics can strengthen personalized learning pathways and provide targeted support (Alifah & Hidayat, 2025), while sustained engagement with AI tools is shaped by perceived usefulness and self-efficacy (Fauzi et al., 2025). These dynamics were evident in participants' qualitative responses in this study.

Taken together, the results suggest that the primary pedagogical value of AI-assisted speaking practice lies not in producing immediate, uniform proficiency gains but in reshaping the affective and motivational foundations that support longer-term language development. As affective readiness stabilizes and learners learn to self-regulate within AI-supported environments, more substantial and measurable linguistic improvements may gradually emerge. In this sense, the present findings do not indicate a limitation of AI tools; instead, they highlight the importance of duration, affective mediation, and learner variability in determining the trajectory of AI-assisted language learning.

Students' Perceptions of AI in Supporting Confidence, Reducing Anxiety, and Enhancing Readiness to Communicate

The qualitative findings further illuminate how AI-assisted learning shaped students' affective experiences and their readiness to communicate in English. As summarized in Table 5, four overarching themes emerged from the thematic analysis: increased confidence, reduced anxiety, enhanced readiness to communicate, and supportive AI features.

Table 5. Thematic map of students' perceptions of AI-assisted speaking practice

Theme	Summary Description	Representative Excerpts	Interpretation
Increased Confidence	Students reported feeling safer to speak, more self-assured, and less afraid of making mistakes when practicing with AI.	*"I feel more confident because I can practice until I am ready."* *"AI makes me brave to speak because no one is judging me."*	AI provided a psychologically safe environment, reducing self-monitoring and fear of negative evaluation. This early confidence gain appears to be an initial affective shift that supports subsequent improvements in WTC and speaking performance.

Reduced Anxiety	AI created a low-pressure, non-judgmental environment that lowered speaking anxiety and fear of evaluation.	*"Speaking to the AI feels less stressful than speaking to people."* *"I don't panic because the AI doesn't judge my mistakes."*	Lower anxiety reduced cognitive load, allowing learners to participate more freely. This aligns with the WTC framework (MacIntyre et al., 1998) and research showing that AI-mediated practice reduces affective barriers.
Enhanced Readiness to Communicate	Students felt more prepared and willing to speak in class after rehearsing with AI.	*"After practicing with AI, I know what I want to say in class."* *"The AI helped me organize my ideas before speaking."*	AI-supported rehearsal improved message planning and conceptualization, strengthening pre-task readiness. This readiness bridges affective gains and observable improvements in communicative engagement.
Supportive AI Features	Instant feedback, unlimited practice, and topic-based simulations helped students monitor and improve performance.	*"The feedback told me exactly what to fix."* *"I like that I can practice as many times as I want."* *"Sometimes the AI gives too many corrections at once it feels too much for me."* *"When the feedback keeps popping up, I feel a bit pressured."*	Continuous individualized feedback enhanced self-regulation and metalinguistic awareness (Zimmerman, 2002; Shen et al., 2025). However, the intensity of automated feedback occasionally felt overwhelming, demonstrating the dual nature of AI-mediated evaluation (Alshehri, 2025; Safar & Anggraheni, 2024).

A dominant theme in the reflections was the increase in students' confidence after practicing with the AI tool. Many indicated that the private, judgment-free setting allowed them to experiment with language without fear of making mistakes. One student expressed that the tool made them "*feel more confident speaking because I could practice repeatedly until I felt ready.*" Such comments suggest that the AI environment supported psychological safety, enabling students to speak more freely than in face-to-face classroom interactions.

Closely related to confidence gains was the theme of reduced anxiety. Students emphasized that speaking to an AI system felt less intimidating than speaking to peers or instructors. Several students reported that the absence of immediate social evaluation "*made speaking feel less stressful,*" helping them break long-standing patterns of hesitation. This finding is consistent with research showing that AI-mediated practice reduces social-evaluative threat and lowers emotional load during speaking tasks (Dong et al., 2026).

Students also described a higher level of readiness to communicate following repeated AI practice. One student explained, "*After practicing with the AI, I felt more ready to speak in class because I already knew what I wanted to say.*" This suggests that AI-assisted rehearsal facilitated pre-task planning: a process known to improve fluency, conceptual clarity, and turn-taking readiness in subsequent communication.

A fourth theme highlighted the advantages of supportive AI features such as instant feedback, real-time scoring, and topic-based simulations. Students appreciated the transparency of performance metrics and the ability to repeat tasks until satisfied. One student noted, *"The feedback told me exactly what to fix,"* while another valued the opportunity for unlimited practice. This continuous, individualized feedback strengthened learners' self-regulation and metalinguistic awareness, consistent with theoretical models of self-regulated learning (Zimmerman, 2002) and recent findings on AI-mediated improvements in speaking (Shen et al., 2025).

However, the thematic analysis also revealed that the highly intensive nature of automated feedback sometimes felt demanding. Some learners expressed that *"the AI gives too many corrections at once. It feels too much,"* indicating that frequent corrective prompts could create momentary pressure. This aligns with studies showing that automated corrective feedback can both facilitate learning and induce cognitive load or feedback fatigue (Alshehri, 2025; Safar & Anggraheni, 2024).

Taken together, these qualitative insights deepen the quantitative findings by showing that learners' improvements in WTC and speaking proficiency were grounded in evolving affective states, particularly confidence, reduced anxiety, and communicative preparedness. Rather than functioning solely as a linguistic tutor, the AI tool appears to have shaped the psychological precursors of communicative performance, demonstrating that affective readiness forms a crucial layer in development of speaking ability.

Challenges Experienced by Students When Using AI Tools

Although students generally expressed positive perceptions of AI-assisted learning, several challenges emerged that shaped their overall experience with the technology. These challenges fell into four broad categories: technical constraints, difficulties interpreting automated feedback, learning-related concerns, and affective, instructional limitations. Together, these issues highlight the areas in which AI-mediated learning still requires pedagogical and infrastructural support.

Table 6. Summary of challenges reported by students

Category of Challenge	Description
Technical Challenges	- Unstable internet connection - Limited device compatibility - Audio/recording interruptions
Feedback Interpretation Issues	- Misrecognition of Indonesian-accented English - Inconsistent automated scoring - Difficulty interpreting AI feedback
Learning Concerns	- Over-reliance on AI for correction - Difficulty integrating AI practice with class tasks
Affective–Instructional Challenges	- Lack of emotional nuance in AI responses - Desire for teacher clarification - Need for AI-literacy guidance

The findings show that students generally reported positive experiences with AI-assisted learning, noting that the tool provided an engaging, motivating, and supportive environment for speaking practice. Learners valued features such as pronunciation scoring, fluency tracking, and topic-based simulations, which made speaking practice more interactive and allowed them to monitor their progress over time. The ability to practice privately, repeatedly, and at their own pace contributed to increased comfort and autonomy, echoing research showing that AI tools can foster greater engagement and personalized oral skill development, especially through real-time scoring and adaptive feedback (Shen et al., 2025; Zhang et al., 2024; Wiboolyasarin et al., 2025).

Despite these benefits, students encountered several challenges that at times limited the effectiveness of AI-assisted practice. The most common issues involved technical constraints, including unstable internet connectivity, limited device compatibility, and occasional audio recording disruptions. These technological barriers have been well documented: for example, research on ASR in language learning highlights misrecognition issues with non-native accents (Liu et al., 2025), while randomized trials found that system errors from AI speech recognition tools can trigger anxiety and disrupt flow (Xiao, 2025). In addition, integrating ASR with automated feedback systems has shown potential to improve speaking competence, but also introduces technical anxiety when recognition fails (Li et al., 2025).

A second set of challenges concerned the interpretation of AI-generated feedback. Learners reported that the system occasionally misrecognized Indonesian-accented English, resulting in feedback that did not reflect their actual speech. Others observed that automated scoring fluctuated significantly across attempts, raising concerns about the reliability of evaluations. Such issues reflect broader concerns about accent bias and imperfect speech-recognition accuracy in current AI systems (Del Río et al., 2023; Nakatumba-Nabende et al., 2025). For some students, these inconsistencies made it difficult to determine how to apply the feedback to improve specific aspects of their speaking performance.

Another concern involved learning-related issues, particularly the potential for over-reliance on AI. Some students expressed worry that depending too heavily on automated corrections might undermine the development of independent self-monitoring skills. This aligns with recent findings in AI literacy research, which emphasize the need for learners to critically evaluate AI-generated output rather than accept it uncritically (Beæiroviæ et al., 2025). Additionally, some students reported difficulty integrating AI-based practice with classroom activities, suggesting that scaffolding and alignment with self-regulated learning strategies could strengthen continuous learning conditions (Wang et al., 2025).

A final set of challenges related to affective–instructional dimensions. Although students appreciated the efficiency and immediacy of AI feedback, some noted limitations in emotional nuance, stating that the AI lacked the supportive interpersonal qualities of face-to-face interaction. Several learners expressed a desire for teacher confirmation, explaining that while AI feedback was useful, "*I still need to check with the lecturer to be sure the feedback is accurate.*" These concerns resonate with Zhang's (2024) findings, which argue that AI can enhance affective support but cannot replicate the relational, empathetic, and context-sensitive dimensions of human instruction.

In summary, while students' experiences with AI tools were largely positive particularly regarding motivation, confidence, and autonomous practice challenges related to technology, feedback interpretation, accent recognition, potential over-reliance, and the absence of human emotional support underscore the need for careful pedagogical planning. Providing structured orientation sessions, AI-literacy training, and stronger integration between AI tasks and classroom instruction will be essential for ensuring that AI-assisted learning is implemented effectively and sustainably within speaking courses.

CONCLUSION

This study suggests that AI-assisted learning can support certain aspects of university students' oral communication development, particularly affective readiness for speaking in contexts where speaking practice is limited, and anxiety is prevalent. Quantitative findings showed that, while students in both the AI-assisted and conventional instruction groups improved in speaking proficiency and WTC over time, the AI-assisted intervention did not yield statistically greater gains at the group level.

Qualitative evidence suggests that AI-supported practice contributed to improved affective conditions for some learners, including increased confidence, reduced anxiety, and greater readiness to communicate. However, these affective benefits were insufficient to yield statistically significant interaction effects during the intervention. Students also perceived the AI platform as a supportive environment for autonomous speaking practice. However, challenges related to feedback interpretation, technical constraints, and integration with classroom learning limited its overall impact.

Despite these positive outcomes, the study is not without limitations. The sample was limited to 60 students from a single Islamic higher education institution, which may restrict the generalizability of the findings to broader EFL populations. The use of intact classes within a quasi-experimental design also limited random assignment, thereby making it impossible to fully rule out pre-existing group differences. Furthermore, the intervention relied on a single AI platform (SmallTalk2Me), and thus, the results may reflect characteristics specific to this tool rather than AI-supported learning more broadly. The relatively short duration of the intervention provided limited time for long-term proficiency development, and the qualitative perceptions were based on self-report, which may be influenced by social desirability or reflective bias. These limitations should be addressed in future research through multi-site sampling, more extended intervention periods, and the use of multiple AI tools or triangulated data sources.

Overall, this study highlights both the possibilities and the current limitations of AI-assisted speaking practice in higher education, underscoring the need for cautious interpretation and further empirical investigation.

■ REFERENCES

Ahmed, I., Ghafoor, A., Liliuara, D., & Akyuningrum, V. Q. (2025). The impact of AI learning tools on ESL learners' motivation and success: A systematic literature review. *English Learning Innovation, 6*(1). https://doi.org/10.22219/englie.v6i1.38961

Alifah, N., & Hidayat, A. R. (2025). Effectiveness of an artificial intelligence-based learning analytics tool in supporting personalized learning in higher education. *Jurnal Pendidikan Progresif, 15*(1), 74–84. https://doi.org/10.23960/jpp.v15i1.pp74-84

Alshehri, A. (2025). AI's effectiveness in language testing and feedback provision. *Social Sciences & Humanities Open, 12*, 101892. https://doi.org/10.1016/j.ssaho.2025.101892

Ambawani, S., Astasari, I., & Rukiati, E. (2025). Overcoming barriers to EFL speaking proficiency: A multidimensional analysis of language learning challenges. *Journal of English in Academic and Professional Communication, 11*(1), 23–36. https://doi.org/10.25047/jeapco.v11i1.5769

Anggrisia, N. F., & Robah, A. (2023). Exploring challenges and strategies in English speaking among Indonesian university students: A case study of AKM University. *Englisia Journal, 11*(1), 1–18. https://doi.org/10.22373/ej.v11i1.19156

Astuti, S. P. (2016). Exploring reticence in Indonesian EFL classrooms: "Silence is gold"? *TEFLIN Journal, 27*(1), 1–22. https://doi.org/10.15639/teflinjournal.

v27i1/1-22

Atifnigar, H., Bawar, H., Momand, M., & Abdul Hamid, S. A. (2022). Oral participation practices in classroom among university students in Afghanistan. *International Journal of Evaluation and Research in Education, 11*(1), 409–416. https://doi.org/10.11591/ijere.v11i1.21865

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

Beæiroviæ, S., Polz, E., & Tinkel, I. (2025). Exploring students' AI literacy and its effects on their AI output quality, self-efficacy, and academic performance. *Smart Learning Environments, 12*, Article 29. https://doi.org/10.1186/s40561-025-00384-3

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

Dafouz, E., & López-Serrano, S. (2026). Unlocking the full potential of English-medium education: Business students' disciplinary literacies. *English for Specific Purposes, 81*, 197–211. https://doi.org/10.1016/j.esp.2025.11.001

Del Río, M., Miller, C., Profant, J., Drexler-Fox, J., McNamara, Q., Bhandari, N., Delworth, N., Pirkin, I., Jetté, M., Chandra, S., Ha, P., & Westerman, R. (2023). Accents in speech recognition through the lens of a World Englishes evaluation set. *Research in Language, 21*(3), 29–55. https://doi.org/10.18778/1731-7533.21.3.02

Dewaele, J.-M. (2021). The emotional rollercoaster ride of foreign language learners and teachers. In M. Simons & T. F. H. Smits (Eds.), *Language Education and Emotions* (pp. 205–220). Routledge.

Dong, S., Wang, H., & Dong, J. (2026). The complex interactions between foreign language emotions and engagement in AI-assisted EFL classrooms. *System, 137*, 103900. https://doi.org/10.1016/j.system.2025.103900

Ebadi, S., Velayati, S., Ramezanzadeh, A., & Salman, A. R. (2025). Exploring the impact of AI-powered speaking tasks on EFL learners' speaking performance and anxiety: An activity theory study. *Acta Psychologica, 259*, 105391. https://doi.org/10.1016/j.actpsy.2025.105391

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System, 121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Fauzi, C., Rahmani, E. F., & Utimadini, N. J. (2025). Beyond technology acceptance: An interplay of self-efficacy, language proficiency, and ChatGPT adoption from a TAM perspective. *Jurnal Pendidikan Progresif, 15*(3), 1970–1988. https://doi.org/10.23960/jpp.v15i3.pp1970-1988

Greenier, V., Derakhshan, A., & Fathi, J. (2021). Emotion regulation, psychological well-being, and teacher engagement. *System, 97*, 102446. https://doi.org/10.1016/j.system.2020.102446

Hsu, H.-W. (2024). An examination of Automatic Speech Recognition (ASR)-based Computer-assisted Pronunciation Training for less proficient EFL students using the Technology Acceptance Model. *International Journal of Technology in Education, 7*(3), 456–473. https://doi.org/10.46328/ijte.681

IELTS Partners. (2025). *IELTS speaking band descriptors*. Cambridge Assessment

English. https://assets.cambridgeenglish.org/webinars/ielts-speaking-band-descriptors.pdf

Khalik, M. F. R. (2025). Reducing English speaking anxiety in classroom presentations: The role of chatbot-assisted preparation. *ELP (Journal of English Language Pedagogy), 10*(2), 106–119. https://doi.org/10.36665/elp.v10i2.1067

Liu, Y., Ab Rahman, F., & Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. *Cogent Education, 12*(1), 2466288. https://doi.org/10.1080/2331186X.2025.2466288

Li, W., Mohamad, M., & You, H. W. (2025). Integrating automatic speech recognition and automated writing evaluation to reduce speaking anxiety and enhance speaking competence among Chinese EFL learners. *Cogent Education, 12*(1), 2559161. https://doi.org/10.1080/2331186X.2025.2559161

MacIntyre, P. D. (2007). Willingness to communicate in the second language: Understanding the decision to speak as a volitional process. *The Modern Language Journal, 91*(4), 564–576. https://doi.org/10.1111/j.1540-4781.2007.00623.x

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

Maulidianti, I., Berliana, N., & Yuliana, Y. G. S. (2024). The effectiveness of using SmallTalk2Me AI in teaching speaking skills. *ELTALL: English Language Teaching, Applied Linguistics and Literature, 5*(2), 148–156. https://doi.org/10.21154/eltall.v5i2.9146

McCroskey, J. C. (1992). Reliability and validity of the willingness-to-communicate scale. *Communication Quarterly, 40*(1), 16–25. https://doi.org/10.1080/01463379209369817

Nakatumba-Nabende, J., Kagumire, S., Kantono, C., & Nabende, P. (2025). A systematic literature review on bias evaluation and mitigation in automatic speech recognition models for low-resource African languages. *ACM Computing Surveys, 58*(4), Article 105, 1–24. https://doi.org/10.1145/3769089

Nazri, M. A. (2025). Communication barriers in speaking: Challenges faced by Indonesian EFL learners. *Journal of English Language and Education, 10*(5), 134–143. https://doi.org/10.31004/jele.v10i5.1089

Panggua, S., Pongsapan, N., Ismail, H., Patandung, Y., & Tangirerung, J. R. (2025). AI-enhanced academic speaking skills: A qualitative investigation of digital tool integration in Indonesian EFL university context. *Forum for Linguistic Studies, 7*(8). https://doi.org/10.30564/fls.v7i8.10763

Peng, J.-E. (2025). *Willingness to Communicate in a Second Language.* Cambridge University Press. https://doi.org/10.1017/9781009451327

Ridayani, M., & Purwanto, B. (2024). Enhancing speaking skills through role play and multimedia technology. *Refleksi: Jurnal Penelitian Tindakan, 2*(1), 1–12. https://doi.org/10.70437/refleksi.v2i2.413

Safar, M., & Anggraheni, D. (2024). Language learning through AI chatbots: Effectiveness and cognitive load analysis. *Journal of Social Science Utilizing Technology, 2*(3), 430–445. https://doi.org/10.70177/jssut.v2i3.1346

Shen, J., Qian, B., Ma, Q., & Deris, F. D.

(2025). Exploring the impact of AI-assisted speaking practice on EFL learners' speaking proficiency and motivation: A mixed-method study. *International Journal of Academic Research in Progressive Education and Development, 14*(4), 43–60. https://doi.org/10.6007/IJARPED/v14-i4/26672

SmallTalk2Me. (2025). *SmallTalk2Me: AI-powered speaking assessment platform*. https://smalltalk2.me

Wang, K., Cui, W., & Yuan, X. (2025). Artificial Intelligence in Higher Education: The Impact of Need Satisfaction on Artificial Intelligence Literacy Mediated by Self-Regulated Learning Strategies. *Behavioral Sciences, 15*(2), Article 165. https://doi.org/10.3390/bs15020165

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. *Ampersand, 14*, 100224. https://doi.org/10.1016/j.amper.2025.100224

Xiao, Y. (2025). The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: A randomized controlled trial. *Humanities and Social Sciences Communications, 12*, 425. https://doi.org/10.1057/s41599-025-04672-8

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

Zhang, M. (2024). Enhancing self-regulation and learner engagement in L2 speaking: exploring the potential of intelligent personal assistants within a learning-oriented feedback framework. *BMC Psychology, 12,* 421. https://doi.org/10.1186/s40359-024-01917-0

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Zulfikar, T., Dahliana, S., & Sari, R. A. (2019). An exploration of English learners' anxiety in speaking classroom. *Studies in English Language and Education, 6*(2), 214–227. https://doi.org/10.24815/siele.v6i2.14652