International Journal of Instruction *January 2027 ● Vol.20, No.1*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp. 601-618

Article submission code: 20260422135301
Received: 22/04/2026 Revision: 06/08/2026
Accepted: 15/08/2026 OnlineFirst: 01/10/2026

AI-Supported Mobile Learning for EFL Speaking: A Quasi-Experimental Study

Chengli Li
Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia, *p131127@siswa.ukm.edu.my*

Harwati Hashim
Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia, *harwati@ukm.edu.my*

Supyan Hussin
Institute of Ethnic Studies, Universiti Kebangsaan Malaysia, Malaysia, *supyan@ukm.edu.my*

Mobile-Assisted Language Learning (MALL) supported by artificial intelligence offers new opportunities to extend speaking practice beyond the classroom through automated feedback and interactive environments. However, empirical evidence on its effectiveness in higher education remains limited. This study examined the impact of an AI-supported mobile learning module on the speaking skills of non-English majors in a Chinese private college. A quasi-experimental pretest–posttest control group design was employed with 83 first-year students (experimental group: n = 40; control group: n = 43). Over 12 weeks, the experimental group engaged in mobile-based speaking activities, including shadowing, pronunciation evaluation, AI-mediated interaction, and English Corner discussions, while the control group received traditional instruction. Speaking performance was assessed using IELTS-based tasks with automated scoring. Although paired-samples t-tests indicated significant improvement within the experimental group, the between-group comparison of gain scores was not statistically significant. The experimental group (M = 0.3125, SD = 0.4189) showed only a slightly greater gain than the control group (M = 0.2209, SD = 0.4126), t(81) = −1.003, p = .319. These findings suggest that while the AI-supported mobile learning module may have supported speaking development, its effectiveness was not demonstrated to be superior to traditional instruction within the 12-week intervention. Longer interventions and stronger pedagogical integration are recommended.

Keywords: artificial intelligence, mobile-assisted language learning, speech recognition technology, EFL speaking skills, quasi-experimental study

Citation: Li, C., Hashim, H., & Hussin, S. (2027). AI-supported mobile learning for EFL speaking: A quasi-experimental study. *International Journal of Instruction*, *20*(1), 601-618.

INTRODUCTION

Speaking as the Most Challenging Skill in Chinese EFL Context

Speaking has long been recognized as one of the most demanding skills for English as a Foreign Language (EFL) learners, particularly among non-English majors in China (Amoah & Yeboah, 2021). Although many students demonstrate competence in reading and writing, they often struggle to transfer their knowledge into fluent spoken communication (Richards, 2008). This imbalance is rooted in traditional English curricula that prioritize grammar instruction and test preparation over interactive oral practice, resulting in limited opportunities for meaningful speaking engagement during class time. Such instructional patterns—characterised by a focus on receptive and form-focused tasks rather than communicative or output-oriented activities—constrain opportunities for students to negotiate meaning and practise spoken English (Mingyan et al., 2025). In such contexts, learners spend far more instructional hours on receptive skills than on active oral engagement, which significantly impedes the development of communicative competence and speaking proficiency (Zhou et al., 2024).

Moreover, speaking anxiety—a form of foreign language anxiety—further compounds this challenge. Horwitz et al. (1986) foundational work identified anxiety related to speaking tasks as a significant barrier to performance, and recent empirical evidence suggests that Chinese tertiary learners continue to experience high levels of apprehension when required to speak in English (Lin & Wang, 2025). This anxiety stems from fear of making mistakes, concern about negative evaluation, and a classroom culture that rarely encourages risk-taking in oral communication. Psychological factors such as lack of confidence and communication apprehension have been shown to inhibit learners' willingness to engage in speaking activities (Amoah & Yeboah, 2021).

Compounding these issues is the scarcity of authentic interactive environments where learners can practise English in meaningful social contexts. Without access to native or fluent interlocutors, many students have few chances to negotiate meaning or sustain extended discourse, further restricting their oral skill acquisition. Collectively, these instructional and affective barriers contribute to the persistent underdevelopment of speaking competence among non-English majors in Chinese higher education, highlighting a critical need for targeted pedagogical interventions.

Mobile-Assisted Language Learning and Oral Production

Mobile-Assisted Language Learning (MALL) has emerged as a promising pedagogical approach that leverages ubiquitous mobile technologies to support language learning beyond the traditional classroom (Kukulska-Hulme, 2020). One of the core affordances of MALL is its capacity to deliver fragmented and context-independent practice opportunities, enabling learners to engage with language tasks during brief periods throughout the day, thereby extending opportunities for language use beyond the classroom (Sari et al., 2020; Stockwell, 2010). This accessibility is particularly beneficial for speaking practice, as learners can use mobile devices to record and replay their oral production at any time and place, facilitating repeated rehearsal that is often infeasible in limited classroom settings (Kukulska-Hulme, 2020).

In addition to flexibility and convenience, MALL environments increasingly integrate speech recognition and automated feedback mechanisms, providing immediate corrective input that supports self-regulated improvement (Burston, 2015; Ramalingam, 2022). Empirical studies on Mobile-Assisted Language Learning (MALL) suggest that mobile-based speaking activities and feedback mechanisms can support improvements in learners' oral performance, including pronunciation accuracy, speaking fluency, and communicative confidence (Benlaghrissi & Ouahidi, 2024; Li, 2024; Metruk, 2024). Such features create a form of "private rehearsal space" for oral practice, where learners can experiment with oral output without the social pressure of classroom performance, thereby mitigating speaking anxiety and fostering greater willingness to practise (Chen, 2024; Dong et al., 2022).

AI Speech Recognition and Automated Feedback

Recent advances in artificial intelligence (AI) have significantly expanded the pedagogical potential of mobile language learning environments, particularly through automatic speech recognition (ASR) and AI-driven feedback systems (Dai & Wu, 2023). These technologies enable automated pronunciation evaluation by analysing spoken input and providing immediate feedback on phonetic accuracy and segmental features, thereby enhancing speaking performance and engagement (Robillos, 2023; Sun, 2023). Such feedback mechanisms can raise learners' awareness of pronunciation errors and supply individualized corrective guidance that would be difficult to provide consistently in large language classes (Bashori et al., 2024).

Beyond pronunciation training, contemporary AI-supported speaking systems increasingly incorporate analytic functions that evaluate multiple aspects of oral production, including speech rate, pausing behaviour, and fluency patterns. These multidimensional feedback mechanisms facilitate self-regulated learning by allowing learners to monitor their own speaking performance and adjust their output accordingly (Inceoglu et al., 2024). In addition, AI technologies have enabled the development of conversational agents and interactive speaking platforms that simulate real-time dialogues with language learners. Such systems provide opportunities for extended spoken interaction and communicative practice that go beyond traditional drill-based exercises (Cui & Zhang, 2025). Empirical studies further suggest that AI-supported speaking practice can enhance learners' speaking performance and engagement by providing flexible and individualized practice environments (Zou et al., 2024).

Theoretical Foundations

From a theoretical perspective, the pedagogical potential of AI-supported mobile speaking practice can be interpreted through several complementary frameworks in second language acquisition research. Recent studies on technology-enhanced language learning suggest that digital speaking environments can facilitate meaningful oral production and interaction when grounded in established SLA theories (Godwin-Jones, 2023). First, Swain's Output Hypothesis posits that language production plays a critical role in promoting linguistic development because learners must "push" their output beyond their current competence, thereby noticing gaps in their interlanguage and engaging in deeper syntactic and lexical processing (Swain, 1985). Mobile-assisted

speaking tasks that require repeated oral production, imitation, and reformulation create structured opportunities for such pushed output, particularly when supported by automated corrective feedback (Sun, 2023). Second, Long's Interaction Hypothesis emphasizes the importance of interactional feedback and negotiation of meaning in facilitating language acquisition (Long, 1996). AI-mediated conversational systems, while not fully equivalent to human interlocutors, can simulate aspects of interactional modification by providing immediate responses and prompting learners to adjust their speech, thereby supporting interaction-driven language development (Cui & Zhang, 2025). Finally, Vygotsky's Sociocultural Theory underscores the mediating role of social interaction and scaffolding in cognitive development (Vygotsky, 1978). Within technology-enhanced learning environments, digital tools and AI systems can function as mediational artefacts that scaffold learners' speaking practice and support their development within the Zone of Proximal Development (ZPD) (Zhang & Zou, 2022). These theoretical perspectives provide a coherent foundation for understanding how AI-supported MALL environments may facilitate both individual speech automatization and socially mediated communicative competence (Zawacki-Richter et al., 2019).

Although previous research has demonstrated the potential of MALL and AI-supported technologies to enhance various aspects of language learning, the existing evidence remains inconclusive with respect to speaking development. While many studies have reported positive outcomes, they have primarily examined isolated technological features, such as automatic speech recognition, pronunciation feedback, or conversational agents, rather than evaluating the effectiveness of a structured AI-supported mobile learning module implemented over an extended instructional period. Moreover, relatively few studies have employed quasi-experimental designs to compare learning outcomes between intervention and control groups in tertiary EFL contexts. Consequently, it remains unclear whether the integration of AI-supported mobile learning can produce measurable improvements in learners' speaking performance beyond those achieved through conventional classroom instruction. To address these gaps, the present study investigates the impact of an AI-supported mobile learning module on the speaking performance of non-English major students in a Chinese tertiary EFL context.

Research Questions

Building upon the theoretical and technological foundations discussed above, the present study aims to examine the pedagogical value of an AI-supported mobile-assisted language learning module (hereinafter referred to as the Module) for improving speaking competence among non-English majors in a Chinese tertiary context. Given the persistent underdevelopment of oral proficiency and the emerging affordances of AI-integrated mobile systems, it is essential to evaluate both instructional effectiveness and learner experience in a structured manner. Accordingly, this study addresses the following research questions:

RQ1: How effective is the Module in improving EFL speaking skills of non-English majors?

RQ2: What challenges do students encounter during the implementation of the Module for speaking practice?

Taken together, these two questions enable a more comprehensive evaluation of the instructional module by integrating outcome-based evidence with learner experience. To address these research questions systematically, the following section describes the research design, intervention procedures, instruments, and data analysis methods employed in this study.

METHOD

Research Design

This study adopted a quasi-experimental pretest–posttest control group design to examine the effectiveness of the Module in improving EFL speaking competence among non-English majors. A quasi-experimental approach was selected because intact classes were used as research units, and random assignment at the individual level was not feasible within the institutional setting. Such a design is widely recognized as appropriate for classroom-based educational research where ecological validity must be preserved (Creswell, 2017; Fraenkel, 1990).

A total of 83 students participated in the study, all of whom were fully informed about the research procedures and signed informed consent forms prior to participation. The control group consisted of 43 students, while the experimental group included 40 students. The intervention was conducted over a 12-week period from April to July 2025. Ethical approval for the study was obtained from the institutional ethics committee, which reviewed, stamped, and formally authorized the implementation of the experiment.

Two intact first-year classes with comparable English proficiency levels were designated as the experimental and control groups, respectively. Because the classes had already been formed by the university before the study commenced, random assignment of individual students was not feasible. To minimize selection bias, the two classes were selected from the same academic program, followed the same curriculum, used the same textbook, and had comparable class sizes. Both groups were taught by the same instructor throughout the 12-week intervention to ensure consistency in teaching content, instructional pace, and assessment procedures. Baseline equivalence was further examined using pretest speaking scores prior to the intervention.

Prior to the intervention, both groups completed a standardized speaking pretest to establish baseline equivalence in pronunciation accuracy, fluency, and overall speaking performance. Pretesting procedures are essential in quasi-experimental designs to ensure initial group comparability and strengthen internal validity (Ary, 2019). The intervention lasted 12 weeks, corresponding to one academic semester.

During the intervention period, the experimental group participated in the Module as a supplementary training component outside regular class hours. Students engaged in structured mobile speaking tasks, including recorded practice, AI-based pronunciation scoring, fluency analysis, and interactive conversational simulations. The module emphasized repeated oral production, automated feedback, and self-regulated practice, which are considered key features of technology-enhanced language learning environments (Kukulska-Hulme, 2020; Stockwell, 2010).

In contrast, the control group received traditional instruction only, which primarily consisted of textbook-based classroom teaching, teacher-led explanations, and limited in-class speaking activities. Such instructional approaches are typical of conventional EFL classrooms, where speaking opportunities are often constrained by time and curriculum requirements (Richards, 2008).

At the end of the 12-week period, both groups completed a posttest equivalent in format and difficulty to the pretest. The use of parallel pretest and posttest measures is a standard procedure for assessing learning gains in intervention studies and allows for the identification of statistically significant improvements attributable to the treatment (Dörnyei, 2007).

This design ensured methodological rigor while maintaining authentic classroom conditions, thereby enhancing both the internal validity and ecological validity of the study (Creswell, 2017).

Intervention: AI-supported Mobile-Assisted Speaking Module

The intervention was implemented through an AI-supported mobile-assisted speaking module delivered over a 12-week period as a supplementary out-of-class speaking enhancement program. The Module was conducted via the Daily English Listening App, which integrates automatic speech recognition (ASR), AI-based pronunciation scoring, and simulated conversational interaction. The design of the Module systematically combined structured input exposure guided oral production, AI-mediated interaction, and socially scaffolded communication to promote measurable gains in pronunciation accuracy and speaking fluency.

The first component emphasized structured listening and imitation practice using materials from Voice of America Special English (VOA Learning English), which provides linguistically graded news content with controlled lexical load and moderated speech rate. Learners engaged in daily listening, dictation, and shadowing tasks. Shadowing required students to repeat the audio in synchrony with the original recording, facilitating phonological alignment and prosodic awareness. This component was theoretically grounded in Krashen's Input Hypothesis, which underscores the role of comprehensible input in acquisition (Krashen, 1985), and was further supported by Swain's Output Hypothesis, which highlights the importance of pushed output in promoting linguistic noticing and restructuring (Swain, 1985). Recent empirical studies have confirmed that mobile-based shadowing and repeated oral production significantly enhance pronunciation and speech rhythm among EFL learners (Hamada, 2025; Whitworth & Rose, 2025). Furthermore, AI-enabled pronunciation scoring has been shown to increase learners' self-monitoring capacity and foster autonomous phonological refinement (Liu et al., 2025; Thi-Nhu Ngo et al., 2024).

The second component involved AI-supported conversational practice. Through the app's conversational interface, students engaged in topic-based spoken interaction aligned with the daily VOA themes. The AI conversational partner provided immediate feedback on pronunciation accuracy, speech rate, pause frequency, and lexical appropriateness. This design aligns with Long's Interaction Hypothesis, which posits that negotiation of meaning and feedback facilitate language development (Long, 1996).

Emerging research indicates that AI conversational agents can effectively simulate interactional adjustments and provide individualized corrective feedback, contributing to measurable improvements in fluency and speaking confidence (Ding & Muhyiddin B Yusof, 2025; Li & Kim, 2024). Importantly, AI-mediated interaction also reduces performance anxiety by creating a low-stakes rehearsal environment, thereby encouraging greater risk-taking in oral production (Ding & Muhyiddin B Yusof, 2025; Wang et al., 2024).

The third component extended individual practice into socially mediated communication through a weekly guided English Corner session. Students participated in small-group discussions based on collaboratively selected news topics, applying vocabulary and expressions practiced during mobile sessions. The instructor provided scaffolding and delayed corrective feedback to consolidate learning outcomes. This collaborative dimension was grounded in Vygotsky's Sociocultural Theory, emphasizing the mediating role of social interaction within the Zone of Proximal Development (Vygotsky, 1978). Recent classroom-based research confirms that combining AI-supported individual practice with peer interaction enhances both linguistic accuracy and communicative competence more effectively than technology-only approaches (Wang et al., 2024; Wu et al., 2025).

To ensure intervention fidelity, weekly learning tasks were standardized for all participants in the experimental group. Students were required to complete the assigned mobile speaking activities each week, and their learning records, including task completion and speaking practice logs. Weekly English Corner sessions were conducted according to the same instructional procedures, and the instructor followed a predefined implementation schedule throughout the intervention period.

Overall, the Module integrated comprehensible input, pushed output, interactional feedback, and socially scaffolded communication within a coherent mobile-supported framework. By aligning technological affordances with established second language acquisition theories and recent AI-based pedagogical research, the Module aimed to provide a theoretically grounded and empirically informed pathway for systematic speaking development.

Instruments

To address the research questions, two primary instruments were employed: a standardized speaking test for measuring learning outcomes and an open-ended questionnaire for exploring learner-reported challenges during the intervention.

The first instrument was a pretest–posttest speaking assessment administered to both the experimental and control groups. The speaking tasks were adapted from authentic IELTS speaking test prompts provided through the Xiaozhan IELTS platform, a widely used online preparation system in China. Participants were required to complete individual oral responses following the standard IELTS Speaking format. Their responses were audio-recorded and uploaded to the platform, which generated automated scores based on the official IELTS assessment criteria, including fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation. The IELTS speaking band descriptors are internationally recognized for their construct

validity and analytic rigor in assessing communicative competence (British Council, 2023). This framework aligns with widely accepted principles of communicative language assessment and analytic rating scales in second language testing (Fulcher, 2015).

In addition, the incorporation of automated scoring is supported by recent research demonstrating that AI-based speech assessment systems can achieve acceptable reliability and strong correlations with human ratings in evaluating pronunciation and fluency (*Automated speaking assessment: Using language technologies to score spontaneous speech*, 2020). Automated scoring further enhances objectivity and consistency by minimizing rater bias and variability, which are common concerns in speaking assessment (Chapelle & Voss, 2016). To ensure comparability, equivalent test formats and difficulty levels were used for both the pretest and posttest, and all recordings were completed under standardized conditions.

The automated scoring system applies standardized IELTS analytic rating criteria consistently across all responses, thereby reducing potential scorer variability. In the present study, identical scoring procedures were applied to both the pretest and posttest under the same testing conditions. In addition, the internal consistency of the speaking scores was examined prior to analysis and demonstrated acceptable reliability.

The second instrument was an open-ended question designed to explore learners' perceived challenges during the implementation of the Module. At the conclusion of the 12-week intervention, students in the experimental group were asked to describe the main difficulties they encountered while using the Module. Although responses were generally brief, they provided supplementary qualitative information regarding students' perceived challenges during the intervention, such as limited vocabulary, difficulty organizing ideas, hesitation during spontaneous speech, and adaptation to AI-mediated interaction. Open-ended elicitation allows participants to express concerns in their own words and is particularly suitable for exploratory inquiry in educational technology research (Creswell, 2017). The responses were thematically coded and categorized to identify recurring patterns, thereby complementing the quantitative findings and offering contextual explanations for performance outcomes.

Together, these two instruments provided complementary quantitative and qualitative evidence to support the interpretation of the findings.

Data Analysis

To comprehensively address the research questions, both quantitative and qualitative data analyses were conducted. The quantitative analysis examined the effectiveness of the Module, whereas the qualitative analysis explored learner-reported challenges during the intervention.

For the quantitative strand, descriptive statistics were first computed using IBM SPSS Statistics (Version 31) to summarize the mean scores, standard deviations, and score distributions of the pretest and posttest speaking assessments for both groups. Descriptive analysis provides an essential overview of central tendency and dispersion before inferential testing (Field, 2024). To evaluate the instructional impact of the

Module, gain scores were calculated by subtracting pretest scores from posttest scores for each participant. Independent-samples t-tests were then conducted to compare gain scores between the experimental and control groups, while paired-samples t-tests were performed to examine within-group improvements over time. Gain score comparison is widely applied in quasi-experimental educational research to determine intervention effects when intact groups are used (Dörnyei, 2007).

In addition to statistical significance testing, effect sizes were calculated using Cohen's d to estimate the magnitude of the intervention effect. Reporting effect size is strongly recommended in applied linguistics and educational technology research, as it provides practical significance beyond p-values (Plonsky & Oswald, 2014). Following conventional benchmarks, effect sizes were interpreted as small (0.2), medium (0.5), or large (0.8). Reliability analysis was also conducted to ensure the internal consistency and stability of the speaking scores. Cronbach's alpha coefficients were calculated to examine the internal consistency of the speaking scores before the inferential analyses were conducted (Chapelle & Voss, 2016). Recent methodological research highlights the importance of reporting statistical significance alongside effect sizes and reliability indices to enhance transparency and interpretability in technology-enhanced language learning studies (Norris et al., 2015; Plonsky & Oswald, 2014).

For the qualitative strand addressing RQ2, a thematic analysis approach was employed to identify recurring patterns in students' open-ended responses. Thematic analysis is suitable for exploratory investigations aiming to categorize common experiences across participants (Braun & Clarke, 2006). The coding procedure followed three stages. First, all responses were transcribed and read repeatedly to achieve data familiarization. Second, initial codes were generated inductively to capture salient difficulty-related expressions (e.g., "limited vocabulary," "difficulty organizing ideas," "nervousness during speaking"). Third, similar codes were grouped into broader thematic categories reflecting learner-level challenges. The coding process and frequency calculations were conducted using Microsoft Excel. To enhance credibility, coding consistency was checked through iterative comparison of categories and refinement of theme labels. The frequency of each theme was calculated to provide a descriptive overview of dominant challenges. This integration of basic quantification within qualitative analysis aligns with recent recommendations for mixed-methods rigor in applied linguistics research (*The Routledge Handbook of Research Methods in Applied Linguistics*, 2019).

The qualitative findings were used to supplement and contextualize the quantitative results by providing descriptive information about learners' perceived challenges during the intervention.

FINDINGS

Baseline Equivalence

Prior to the intervention, Levene's test confirmed that the assumption of homogeneity of variance was not violated, $F(1, 81) = 0.005$, $p = .943$. An independent-samples t-test further revealed no statistically significant difference between the experimental group ($M = 6.18$, $SD = 0.90$) and the control group ($M = 6.08$, $SD = 0.95$) in pretest speaking

scores, t(81) = -0.46, p = .646. The effect size was negligible (Cohen's d = -0.10), indicating that the two groups were statistically equivalent at baseline (see Table 1).

Table 1
Baseline speaking scores of the control and experimental groups

Group	N	Mean	SD
Control	43	6.08	0.95
Experimental	40	6.18	0.90

Within-Group Improvements

Paired-samples t-tests were conducted to examine pretest–posttest changes in speaking performance within each group.

For the control group, the mean speaking score increased slightly from the pretest (M = 6.081, SD = 0.951) to the posttest (M = 6.302, SD = 0.920). The correlation between pretest and posttest scores was strong and statistically significant (r = .903, p < .001), indicating high consistency between the two measurements. The paired-samples t-test showed that the improvement was statistically significant, t(42) = −3.511, p = .001. The mean difference between the two tests was −0.2209 (SD = 0.4126), with a 95% confidence interval ranging from −0.3479 to −0.0939. The effect size was moderate (Cohen's d = −0.535), suggesting a modest improvement in speaking performance among students in the control group.

For the experimental group, the mean speaking score increased from the pretest (M = 6.175, SD = 0.896) to the posttest (M = 6.488, SD = 0.788). The correlation between the pretest and posttest scores was also strong and statistically significant (r = .884, p < .001). The paired-samples t-test indicated a statistically significant improvement in speaking performance, t(39) = −4.718, p < .001. The mean difference was −0.3125 (SD = 0.4189), with a 95% confidence interval ranging from −0.4465 to −0.1785. The effect size was relatively large compared with the control group (Cohen's d = −0.746), indicating a stronger improvement in speaking performance for students in the experimental group (see Table 2).

Table 2
Pretest and posttest speaking scores within each group

Group	Test	N	Mean	SD
Control	Pretest	43	6.081	0.951
Control	Posttest	43	6.302	0.920
Experimental	Pretest	40	6.175	0.896
Experimental	Posttest	40	6.488	0.788

Overall, both groups demonstrated statistically significant improvements in speaking scores from pretest to posttest. Descriptively, the experimental group showed a larger mean gain and effect size than the control group. The statistical significance of this between-group difference is examined in the following section.

Between-Group Gain Score Comparison

To further examine the effectiveness of the Module, gain scores were calculated by subtracting pretest scores from posttest scores for each participant. An independent-samples t-test was conducted to compare the improvement in speaking performance between the experimental group and the control group.

The results showed that the experimental group ($M = 0.3125$, $SD = 0.4189$) achieved a slightly higher mean gain in speaking scores than the control group ($M = 0.2209$, $SD = 0.4126$). However, the difference between the two groups was not statistically significant, $t(81) = -1.003$, $p = .319$. The 95% confidence interval for the mean difference ranged from −0.273 to 0.090, indicating that the observed difference could plausibly occur by chance.

The effect size was small, with Cohen's $d = -0.22$, indicating a limited difference in gain scores between the two groups. (see Table 3).

Table 3
Comparison of speaking gain scores between the control and experimental groups

Group	N	Mean Gain	SD
Control	43	0.221	0.413
Experimental	40	0.313	0.419

Overall, both groups showed significant improvements in speaking performance from pretest to posttest. However, no statistically significant difference was found between the two groups in gain scores.

Qualitative Findings: Learner-Reported Challenges

Thematic analysis of the open-ended responses revealed several recurring challenges encountered by students during the implementation of the AI-supported mobile speaking module. Four primary themes were identified: (1) limited lexical resources, (2) difficulty organizing ideas during extended speech, (3) hesitation and reduced fluency in spontaneous production, and (4) initial unfamiliarity with AI-based interaction.

Among these, limited lexical resources was the most frequently reported difficulty. Many participants indicated that although the system supported pronunciation practice through automated feedback, producing appropriate vocabulary in real time remained challenging. Students also reported difficulty organizing ideas during extended speaking tasks. This was often reflected in pauses, repetitions, and fragmented responses during longer utterances. Hesitation and reduced fluency in spontaneous production were also commonly reported. Some learners noted difficulty maintaining continuous speech, particularly in tasks requiring immediate responses. In addition, several participants mentioned initial unfamiliarity with AI-based interaction. This issue appeared to decrease over time as learners became more accustomed to the system.

Despite these challenges, participants generally did not report technical barriers that prevented them from engaging with the Module.

DISCUSSION

The findings showed that both the experimental group and the control group achieved significant improvements in speaking performance between the pretest and posttest. Although the experimental group obtained a slightly larger mean gain than the control group, the between-group difference was not statistically significant. Therefore, the findings do not provide sufficient evidence to conclude that the Module was more effective than traditional instruction within the 12-week intervention period. Rather, the results suggest that structured AI-supported mobile speaking practice may serve as a useful supplement to classroom instruction, although its additional contribution beyond conventional teaching was not statistically confirmed in this study.

The results align with previous research highlighting the pedagogical benefits of Mobile-Assisted Language Learning for developing productive language skills. As discussed in the literature, mobile learning environments allow learners to engage in fragmented and flexible practice opportunities that extend beyond classroom constraints (Kukulska-Hulme & Shield, 2008; Ramalingam, 2023). The present study supports this argument by demonstrating that regular mobile-based speaking practice, including daily listening, imitation, and oral recording, can facilitate incremental improvements in speaking performance. Consistent with earlier findings, the ability to repeatedly record and rehearse speech may have enabled learners to engage in more sustained oral practice than would typically be possible during limited classroom speaking activities (Kukulska-Hulme, 2020).

At the same time, the absence of a statistically significant between-group difference suggests that the benefits of mobile-supported practice may not be sufficiently large to produce measurable advantages over regular classroom instruction within a single semester. It is also possible that students in both groups benefited from the same classroom teaching and accumulated speaking practice over time, thereby reducing the observable difference between the two instructional conditions.

Another important factor contributing to the effectiveness of the intervention may be the integration of AI-based speech recognition and automated feedback. Previous studies have suggested that automated pronunciation scoring and fluency analysis can provide learners with immediate and objective feedback on their spoken performance (Godwin-Jones, 2023). In the present study, the Daily English Listening platform allowed students to receive instant evaluation of their pronunciation accuracy and fluency, which may have increased their awareness of phonological features and encouraged self-regulated learning. In addition, the AI conversational partner enabled learners to participate in simulated dialogues, creating opportunities for spontaneous language production and meaning negotiation. Such interactional affordances have been shown to promote oral fluency and communicative confidence in AI-assisted language learning environments (Cui & Zhang, 2025; Ding & Muhyiddin B Yusof, 2025).

Nevertheless, given the absence of a statistically significant between-group difference, the contribution of the AI components should be viewed as complementary rather than superior to traditional instruction. Their primary value may lie in providing additional

opportunities for practice, immediate feedback, and learner autonomy, rather than producing substantial short-term gains in speaking performance.

The learner-reported challenges identified in the qualitative findings also help explain why the intervention did not produce a statistically significant advantage over traditional instruction. Many participants reported difficulties related to limited vocabulary, organizing ideas during extended speech, and maintaining fluency in spontaneous communication. These challenges are associated with broader aspects of communicative competence that are not easily addressed through pronunciation feedback or repeated oral practice alone. Although the AI-supported Module provided opportunities for frequent speaking practice, learners may still have required more explicit instructional support to develop lexical resources and discourse organization. In addition, several students reported initial unfamiliarity with AI-mediated interaction, suggesting that part of the intervention period may have been devoted to adapting to the learning environment rather than maximizing language development. These learner-level challenges may therefore have reduced the measurable impact of the intervention during the 12-week period.

From a theoretical perspective, these findings can be more fully understood by linking the observed learning outcomes to established frameworks in second language acquisition, which together help explain how the Module facilitated speaking development. First, the results support Swain's Output Hypothesis, which emphasizes the role of language production in promoting linguistic development (Swain, 1985). The speaking tasks embedded in the Module required students to actively produce oral language through shadowing, recording, and conversational interaction, thereby pushing learners to notice gaps in their linguistic knowledge. Second, the interactive component of the AI conversational system reflects principles of the Interaction Hypothesis, which highlights the importance of interactional feedback in facilitating language acquisition (Long, 1996). Although the AI partner does not fully replicate human interaction, it provides immediate responses that prompt learners to adjust their output. Finally, the weekly English Corner activities demonstrate the relevance of Vygotsky's Sociocultural Theory, as collaborative discussions and instructor feedback provided scaffolding that supported learners' development within their Zone of Proximal Development (Vygotsky, 1978).

Taken together, these theoretical perspectives provide a plausible explanation for the improvements observed across the intervention. However, the lack of a statistically significant between-group difference suggests that these learning mechanisms may also operate, to some extent, within well-designed classroom instruction.

CONCLUSION

This study investigated the effectiveness of an AI-supported mobile-assisted language learning module in improving the speaking skills of non-English major students in a Chinese private college. Using a quasi-experimental pretest–posttest control group design, the study compared students receiving traditional instruction with those participating in a 12-week mobile-assisted speaking program. The results indicated that although both groups demonstrated significant improvements in speaking performance

over the course of the intervention. Although the experimental group obtained a slightly higher mean gain than the control group, the between-group difference was not statistically significant. Therefore, the findings do not provide sufficient evidence to conclude that the AI-supported Module was more effective than traditional instruction. Instead, they suggest that integrating structured mobile speaking activities with AI-supported feedback may serve as a complementary approach to supporting speaking development. From a pedagogical perspective, the findings highlight the potential of mobile-assisted learning as a complementary approach to conventional classroom instruction. Mobile-based speaking practice can extend learning beyond classroom constraints and provide learners with additional opportunities for repeated oral production. AI-supported feedback may encourage learners to monitor aspects of their pronunciation and fluency, thereby supporting self-regulated learning. When combined with classroom instruction and collaborative activities, such as English Corner discussions, these features may enrich rather than replace conventional teaching practices. Nevertheless, these potential benefits should be interpreted as complementary to, rather than superior to, conventional classroom instruction.

Several limitations should be noted. The study was conducted in a single institution with a relatively small sample, which may limit the generalizability of the findings. In addition, the 12-week intervention may not have been sufficient to produce clear between-group differences, and the use of brief open-ended responses may not fully capture the complexity of learners' experiences. Future research is therefore recommended to adopt longitudinal designs with extended intervention periods to examine the sustained impact of AI-supported mobile learning on speaking development. In addition, larger and more diverse samples across different institutional contexts would help enhance the external validity of the findings. Further studies could also incorporate in-depth qualitative methods, such as interviews or learning diaries, to better understand how learners engage with AI-supported speaking tools over time. Finally, future investigations may explore how different configurations of instructional design — such as the balance between individual practice, AI interaction, and teacher guidance—affect learning outcomes in mobile-assisted language learning environments.

Funding: This study was not funded.

Conflict of Interest: The authors declare no conflict of interest.

Authenticity: The authors confirm that this manuscript is original, has not been published previously, and is not under consideration for publication elsewhere. All sources are duly cited and referenced.

Artificial Intelligence Statement: During the drafting process, AI tools such as Grammarly and ChatGPT were used only for language improvement and idea refinement.

REFERENCES

Amoah, S., & Yeboah, J. (2021). The Speaking Difficulties of Chinese EFL Learners and Their Motivation Towards Speaking the English language. *Journal of Language and Linguistic Studies*, *17*(1), 56-69. https://doi.org/10.52462/jlls.4

Ary, D., Jacobs, L., Irvine, C., & Walker, D. (2019). *Introduction to research in education* (10th ed.). Cengage Learning.

Zechner, K., & Evanini, K. (Eds.). (2020). *Automated speaking assessment: Using language technologies to score spontaneous speech*. Routledge.

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). I Can Speak: Improving English Pronunciation Through Automatic Speech Recognition-Based Language Learning Systems. *Innovation in Language Learning and Teaching*, *18*(5), 443-461. https://doi.org/10.1080/17501229.2024.2315101

Benlaghrissi, H., & Ouahidi, L. M. (2024). The Impact of Mobile-Assisted Project-Based Learning on Developing EFL Students' Speaking Skills. *Smart Learning Environments*, *11*(1), Article 18. https://doi.org/10.1186/s40561-024-00303-y

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

British Council, I. I. A., & Cambridge Assessment English. (2023). *IELTS speaking band descriptors*. Retrieved 04 January, 2026 from https://www.ielts.org

Burston, J. (2015). Twenty Years of MALL Project Implementation: a Meta-Analysis of Learning Outcomes. *ReCALL*, *27*(1), 4-20. https://doi.org/10.1017/S0958344014000159

Chapelle, C. A., & Voss, E. (2016). 20 Years of Technology and Language Assessment in Language Learning & Technology. *Language Learning & Technology*, *20*(2), 116-128. http://llt.msu.edu/issues/june2016/chapellevoss.pdf

Chen, Y. C. (2024). Effects of Technology-Enhanced Language Learning on Reducing EFL Learners' Public Speaking Anxiety. *Computer Assisted Language Learning*, *37*(4), 789-813. https://doi.org/10.1080/09588221.2022.2055083

Creswell, J. W., & Creswell, J. D. (2017). *Research design: qualitative, quantitative, and mixed methods approaches*. Sage publications.

Cui, W., & Zhang, J. (2025). AI Applications in Language Learning: Improving Speaking Fluency and Pronunciation Using Chatbots and Speech Feedback. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2025.2546634

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

Ding, D., & Muhyiddin B Yusof, A. (2025). Investigating the Role of AI-Powered Conversation Bots in Enhancing L2 Speaking Skills and Reducing Speaking Anxiety: a Mixed Methods Study. *Humanities and Social Sciences Communications*, *12*(1), 1223. https://doi.org/10.1057/s41599-025-05550-z

Dong, L., Jamal Mohammed, S., Ahmed Abdel-Al Ibrahim, K., & Rezai, A. (2022). Fostering EFL Learners' Motivation, Anxiety, and Self-efficacy Through Computer-Assisted Language Learning- and Mobile-Assisted Language Learning-Based

Instructions. *Frontiers in Psychology*, *13*, Article 899557. https://doi.org/10.3389/fpsyg.2022.899557

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th ed.). Sage.

Fraenkel, J. R., & Wallen, N. E. (1990). *How to design and evaluate research in education*. McGraw-Hill.

Fulcher, G. (2015). *Re-examining language testing: a philosophical and social inquiry*. Routledge.

Godwin-Jones, R. (2023). Emerging Spaces for Language Learning: AI bots, Ambient Intelligence, and the Metaverse. *Language Learning & Technology*, *27*(2), 6-27. https://hdl.handle.net/10125/73501

Hamada, Y. (2025). Developing a New Shadowing Procedure for Listening Vocabulary: Selective Shadowing. *Innovation in Language Learning and Teaching*, *19*(2), 122-135. https://doi.org/10.1080/17501229.2024.2340567

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign Language Classroom Anxiety. *The Modern Language Journal*, *70*(2), 125-132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Inceoglu, S., Chen, W. H., & Lim, H. (2024). Monitoring student behavior in autonomous automatic speech recognition-based pronunciation practice. *System*, *124*, Article 103387. https://doi.org/10.1016/j.system.2024.103387

Krashen, S. D. (1985). *The input hypothesis: issues and implications*. Longman.

Kukulska-Hulme, A. (2020). Mobile-Assisted Language Learning. *The Encyclopedia of Applied Linguistics,* pp. 1-9. https://doi.org/https://doi.org/10.1002/9781405198431.wbeal0768.pub2

Kukulska-Hulme, A., & Shield, L. (2008). An Overview of Mobile Assisted Language Learning: from Content Delivery to Supported Collaboration and Interaction. *ReCALL*, *20*(3), 271-289. https://doi.org/10.1017/S0958344008000335

Li, L., & Kim, M. (2024). It Is like a Friend to Me: Critical Usage of Automated Feedback Systems by Self-Regulating English Learners in Higher Education. *Australasian Journal of Educational Technology*, *40*(1), 1-18. https://doi.org/10.14742/ajet.8821

Li, R. (2024). Effects of Mobile-Assisted Language Learning on Foreign Language Learners' Speaking Skill Development. *Language Learning and Technology*, *28*(1), 1-26. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85182228330&partnerID=40&md5=a5c5edfadbe1040c0b215a011122caac

Lin, J., & Wang, Y. (2025). Exploring Chinese University Students' Foreign Language Enjoyment, Engagement and Willingness to Communicate in EFL Speaking Classes. *Humanities and Social Sciences Communications*, *12*(1), Article 650.

https://doi.org/10.1057/s41599-025-04948-z

Liu, Y., binti Ab Rahman, F., & binti 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

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. R. T. Bhatia (Ed.), *Handbook of Second Language Acquisition*. Academic Press.

Metruk, R. (2024). Mobile-Assisted Language Learning and Pronunciation Instruction: a Systematic Literature Review. *Education and Information Technologies*, *29*(13), 16255-16282. https://doi.org/10.1007/s10639-024-12453-0

Mingyan, M., Noordin, N., & Razali, A. B. (2025). Improving EFL Speaking Performance among Undergraduate Students with an AI-Powered Mobile App in After-Class Assignments: an Empirical Investigation. *Humanities and Social Sciences Communications*, *12*(1), Article 370. https://doi.org/10.1057/s41599-025-04688-0

Norris, J. M., Plonsky, L., Ross, S. J., & Schoonen, R. (2015). Guidelines for Reporting Quantitative Methods and Results in Primary Research. *Language Learning*, *65*(2), 470-476. https://doi.org/https://doi.org/10.1111/lang.12104

Plonsky, L., & Oswald, F. L. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning*, *64*(4), 878-912. https://doi.org/10.1111/lang.12079

Ramalingam, S. Y., M. M.； Hashim, H. (2022). Blended Learning Strategies for Sustainable English as a Second Language Education: a Systematic Review. *Sustainability*, *14*(13), Article 8051. https://doi.org/https://doi.org/10.3390/su14138051

Ramalingam, S. Y., M. M.； Hashim, H. (2023). ESL Learners' Qualitative Perspective on Learner-to-Instructor Interaction in Blended Environment. *Arab World English Journal*(9), 234–248. https://doi.org/https://doi.org/10.24093/awej/call9.16

Richards, J. C. (2008). *Teaching listening and speaking* (Vol. 35). Cambridge University Press.

Robillos, R. J. (2023). Improving Students' Speaking Performance and Communication Engagement through Technology-Mediated Pedagogical Approach. *International Journal of Instruction*, *16*(1), 551-572. https://doi.org/10.29333/iji.2023.16131a

Rose, H. (Ed.). (2019). *The Routledge handbook of research methods in applied linguistics*. Routledge.

Sari, A. B. P., Dardjito, H., & Azizah, D. M. (2020). EFL Students' Improvement through the Reflective YouTube Video Project. *International Journal of Instruction*, *13*(4), 393-408. https://doi.org/10.29333/iji.2020.13425a

Stockwell, G. (2010). Using Mobile Phones for Vocabulary Activities: Examining the Effect of the Platform. *Language Learning and Technology*, *14*(2), 95-110. http://llt.msu.edu/vol14num2/stockwell.pdf

Sun, W. (2023). The Impact of Automatic Speech Recognition Technology on Second

Language Pronunciation and Speaking Skills of EFL Learners: a Mixed Methods Investigation. *Frontiers in Psychology*, *14*, Article 1210187. https://doi.org/10.3389/fpsyg.2023.1210187

Swain, M. (1985). Communicative Competence: Some Roles of Comprehensible Input and Comprehensible Output in Its Development. *Input in Second Language Acquisition*, *15*, 165-179.

Thi-Nhu Ngo, T., Hao-Jan Chen, H., & Kuo-Wei Lai, K. (2024). The Effectiveness of Automatic Speech Recognition in ESL/EFL Pronunciation: a Meta-Analysis. *ReCALL*, *36*(1), 4-21. https://doi.org/10.1017/S0958344023000113

Vygotsky, L. S., & Cole, M. (1978). *Mind in Society : the Development of Higher Psychological Processes*. Harvard University Press.

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The Impact of Different Conversational Generative AI Chatbots on EFL Learners: an Analysis of Willingness to Communicate, Foreign Language Speaking Anxiety, and Self-Perceived Communicative Competence. *System*, *127*, 103533. https://doi.org/https://doi.org/10.1016/j.system.2024.103533

Whitworth, B., & Rose, H. (2025). A Systematic Review of Research on the Use of Shadowing for Second Language Pronunciation Teaching. *Research Synthesis in Applied Linguistics*, *1*(2), 239-269. https://doi.org/10.1080/29984475.2025.2546827

Wu, T.-T., Hapsari, I. P., & Huang, Y.-M. (2025). Effects of Incorporating AI Chatbots into Think–Pair–Share Activities on EFL Speaking Anxiety, Language Enjoyment, and Speaking Performance. *Computer Assisted Language Learning*, 1-39. https://doi.org/10.1080/09588221.2025.2478271

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where Are the Educators?. *International Journal of Educational Technology in Higher Education*, *16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, R., & Zou, D. (2022). Types, Purposes, and Effectiveness of State-of-the-Art Technologies for Second and Foreign Language Learning. *Computer Assisted Language Learning*, *35*(4), 696-742. https://doi.org/10.1080/09588221.2020.1744666

Zhou, R., Samad, A., & Perinpasingam, T. (2024). A Systematic Review of Cross-Cultural Communicative Competence in EFL Teaching: Insights from China. *Humanities and Social Sciences Communications*, *11*(1), Article 1750. https://doi.org/10.1057/s41599-024-04071-5

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*, Article 103497. https://doi.org/10.1016/j.system.2024.103497