



ARTICLE



<https://doi.org/10.1057/s41599-025-04688-0>

OPEN

Improving EFL speaking performance among undergraduate students with an AI-powered mobile app in after-class assignments: an empirical investigation

Ma Mingyan^{1,2}, Nooreen Noordin¹ & Abu Bakar Razali¹

English speaking represents one of the most challenging competencies for EFL learners, mainly due to the limited opportunities for authentic practice, especially within monolingual contexts like China. However, the ubiquitous personal mobile devices (smartphones) and the advent of AI-powered mobile apps equipped with automatic speech recognition, natural language processing, and text-to-speech present novel solutions to overcome these hurdles. This study investigates the effects of an AI-powered mobile application (Liulishuo) on Chinese undergraduate EFL students' speaking performance, using features such as automatic feedback, process-oriented monitoring, and tailored instructions. Employing a quasi-experimental design, this 10-week study was conducted at a Chinese university with two groups of participants. The control group ($n=31$) engaged with WeChat for after-class assignments, whereas the experimental group ($n=32$) supplemented WeChat use with Liulishuo, diversifying their educational tools. Results indicated that participants in the experimental group significantly outperformed those in the control group regarding overall speaking performance. More specifically, notable improvements were observed in pronunciation and fluency, while vocabulary and grammar enhancements were not statistically significant. These findings underscore the capacity of AI-powered mobile apps to enhance EFL speaking performance, signaling novel pathways for the evolution of language education.

¹ Department of Language and Humanities Education, Faculty of Educational Studies, Universiti Putra Malaysia (UPM), Serdang, Selangor, Malaysia. ² School of Foreign Language Department, Jilin Business and Technology College, Changchun City, Jilin Province, China. email: mamingyan12345@126.com; nooreen@upm.edu.my; abmr_bakar@upm.edu.my

Introduction

English as a global lingua franca is central to international communication, emphasizing the importance of EFL speaking proficiency for learners. For university students, mastering English speaking is crucial for academic achievements, career prospects, and personal development. In recent years, English language education has become a strategic priority in China, with about 93.8% of the population participating in English-learning initiatives (Wei & Su, 2012), reflecting the goal of integrating with the global community. However, many learners struggle to master EFL speaking skills (Ur, 2000). Despite at least six years of compulsory English education, many Chinese undergraduate students still encounter difficulties in reaching a level of English speaking proficiency comparable to their peers in other Asian countries, such as South Korea and Japan (Zhu (2022)), primarily due to the lack of exposure to authentic English-speaking context (Wang, Smyth, & Cheng, 2017; Amoah & Yeboah, 2021). Factors leading to these difficulties are constraints, such as limited in-class time and large class sizes, which make providing sufficient practice for each student on speaking performance impossible (Hafour, 2022; Pu & Chang, 2023).

The digital era, characterized by the pervasive presence of wireless networks and mobile devices (e.g., smartphones and tablets), has been seamlessly integrated into daily life and academic pursuits (Hwang, et al., 2022; Morgana, Kukulska-Hulme (2021)). This shift has significantly attracted the interest of language researchers and educators toward mobile-assisted language learning (MALL) (Kukulska-Hulme et al., 2018; Reinders & Pegrum, 2017), as it helps overcome foreign language learning difficulties by extending the physical and temporal boundaries of the classroom (Stockwell, 2022). Particularly with cutting-edge technologies like artificial intelligence (AI) in the MALL context, AI apps, equipped with automatic speech recognition, natural language processing, and text-to-speech technologies, offer targeted automatic feedback (Zou et al., 2023), process-oriented monitoring (Junaidi, 2020), and personalized instruction (El Shazly, 2021).

However, empirical research-based evidence on the impact of AI-powered tools on EFL speaking performance remains limited (Shortt et al., 2023; Zhou, 2021). Most existing research on the impact of AI technology in the MALL context overlooks the current mobile age, resulting in inconsistent outcomes (Nami, 2020; Kukulska-Hulme et al., 2020; Mihaylova et al., 2022). These studies often focus solely on comparing AI-powered MALL with traditional instruction without MALL. As Sharples et al. (2016) stated, smartphones with diverse mobile apps have already shaped new educational patterns and transformed how language is taught and learned both inside and outside the traditional classroom. Smartphones, equipped with communication support mobile applications such as WhatsApp, Telegram, and WeChat, have become more prevalent tools in MALL for facilitating exposure to the target language (Rajendran & Yunus, 2021) than AI-powered MALL tools. It is crucial to consider the existing effects of communication support mobile apps to fully appreciate the effectiveness of AI technology in English as a foreign language education.

Similarly, while AI applications show promise in enhancing EFL speaking skills, empirical research in actual classroom settings remains limited (Hwang et al., 2024; Chen et al., 2020). Yang and Kyun's (2022) review of 25 studies from 2007 to 2021 highlights a predominant focus on technology development rather than practical outcomes in real-world language learning. Even the limited existing research on this topic focuses on one or two specific sub-skills of speaking performance, such as pronunciation and/or fluency, rather than comparing all four sub-skills of speaking performance. Given that effective speaking

performance encompasses both linguistic knowledge and fluency (Ur, 2002; Ghafar, 2023), this lack of clarity underscores the need for a comprehensive approach that integrates all speaking sub-skills—vocabulary, grammar, pronunciation, and fluency—to fully understand how AI can enhance EFL proficiency and communicative effectiveness, particularly for learners in China (Zhou, 2021).

This study intends to bridge gaps regarding the impact of AI-powered mobile apps on EFL speaking performance, these gaps refer to the lack of comprehensive studies that evaluate all four speaking sub-skills simultaneously—vocabulary, grammar, pronunciation, and fluency—as well as the under-representation of AI apps compared to communication mobile apps, like WeChat. These limitations result in an incomplete understanding of how AI apps can improve overall EFL speaking performance. In this context, the independent variable in this study is the use of AI-powered mobile apps, specifically the Chinese AI app Liulishuo, while the dependent variable is the EFL speaking performance of Chinese undergraduate students, measured by overall speaking ability and four specific sub-skills—vocabulary, grammar, pronunciation, and fluency—based on IELTS speaking test criteria. This study focuses on the effects of the Liulishuo app within the existing mobile educational environments, where communication mobile apps serve as the baseline level of the MALL context. The study evaluates EFL speaking performance by examining both overall speaking abilities and the four specific sub-skills (i.e., vocabulary, grammar, pronunciation, and fluency), providing a comprehensive evaluation of speaking proficiency. The main research questions are as follows:

1. What are the effects of AI-powered mobile applications on enhancing the EFL speaking performance of Chinese undergraduate students?
2. What are the effects of AI-powered mobile applications on four EFL speaking sub-skills (i.e., vocabulary, grammar, pronunciation, and fluency) among Chinese undergraduate students?

Literature review

EFL speaking performance. EFL speaking performance is recognized as one of the most challenging skills among the four language competencies for EFL learners (Jao, Yeh, Huang, & Chen, 2022). This is due to the multifaceted nature of speaking skills. As Leong and Ahmadi (2017) emphasized, effective communication requires both fluency and accuracy. To convey meaning clearly and effectively, speaking performance must demonstrate accuracy in linguistic knowledge and fluency in delivery (Ghafar, 2023). Given the complexity of speaking, a comprehensive assessment must account for multiple dimensions, including pronunciation, grammar, vocabulary, and fluency, which together form the foundation of effective speaking performance (Suzuki & Kormos, 2020). Furthermore, speakers must simultaneously act as listeners, receivers, and processors while generating speech in real-time social contexts (Brown & Lee, 2015). In line with Vygotsky's (1987) sociocultural theory, language learning occurs primarily through social interaction, not individual effort. Since EFL speaking inherently involves social interactions, practice in authentic contexts is essential for enhancing speaking performance (Hwang et al., 2022; Sun et al., 2017). However, insufficient opportunities for learners to engage with the target language in real-world contexts remain a significant obstacle to their speaking proficiency (Hafour, 2022; Pu & Chang, 2023; Ahmadi et al., 2013).

As a monolingual country, China claims the world's largest English-learning demographic (Kang & Lin, 2019), underscoring the pivotal role of English-speaking proficiency in shaping the employment opportunities and overall achievements of its undergraduate students (Nam & Jiang, 2023). Yet, many Chinese undergraduate EFL students demonstrate lower English-speaking performance compared to their counterparts (Wang, Smyth, & Cheng, 2017). A significant factor behind this issue is the Chinese exam-oriented educational system, which fails to prioritize English speaking skills, focusing instead on listening, reading, and writing, especially in critical assessments like the National College Entrance Exam (Butler, Lee, & Peng, 2022). This oversight within exam-oriented education results in limited opportunities/exposure for oral practice within classroom settings, subsequently leading to poorer speaking skills among many students. In response to this challenge, mobile-assisted language learning (MALL) has emerged as a promising solution to enhance English speaking performance for each student.

Mobile-assisted language learning (MALL). The proliferation of wireless networks has led to the widespread use of mobile devices, such as smartphones and tablets, in daily lives (Bortoluzzi, Bertoldi, & Marenzi, 2021), offering significant advantages for foreign language education, particularly mobile-assisted language learning (MALL) (Li, Fan, & Wang, 2022; Kukulska-Hulme, 2019; Pegrum, 2019; Foroutan & Noordin, 2012). MALL refers to the use of mobile technologies in language learning, particularly where the portability of devices presents unique benefits (Kukulska-Hulme et al., 2018, p.2). In the context of China, the internet traffic volume attributed to mobile devices witnessed a significant surge, reaching 21.2 billion in 2017—a notable 158.2% increase compared to the previous year. This phenomenon is underscored by the fact that a vast majority of Chinese internet users, 97.5% to be precise, predominantly access the internet via smartphones (Kang & Lin, 2019). This trend is particularly evident among undergraduate students, for whom smartphones have become an almost indispensable tool (Chwo, Marek & Wu, 2018), thereby seamlessly facilitating the incorporation of mobile-assisted language learning (MALL) into their EFL education.

Kukulska-Hulme (2016; 2019) categorizes the assistance provided by MALL into two primary supports: communication support and mobile language learning support. She emphasizes that “technology would connect people to facilitate assistance, while in other cases assistance would be built into the design of materials, applications, tools or avatars” (p.130), which indicates two main mobile types of MALL: mobile communication support apps (e.g., ZOOM, WhatsApp, and WeChat), which function as social networking service (SNS) tools to enhance communication and provide support among users, thereby aiding foreign language learners in receiving assistance from others; and mobile language learning support apps, which employ advanced technologies (e.g., AI and VR) and are specifically designed to enhance the foreign language learning experience. To date, smartphones equipped with social networking service (SNS) apps have become the most prevalent tools in MALL for authentic language learning (Burston, 2015; Kukulska-Hulme et al., 2018; Zou et al., 2023). The accessibility, interactive settings, and vast linguistic content of these apps have led to their widespread use. SNS apps provide ample opportunities for EFL exposure by promoting learner autonomy (Okumura, 2022; Evita, Muniroh, & Suryati, 2021), building learning communities (Meyasa & Santosa, 2023; Peeters & Pretorius, 2020), enabling collaborative learning (Cai & Zhang, 2023; Yang, 2020), and immersing learners in authentic language environments (Khodabandeh, 2022). These platforms facilitate direct spoken exchanges, offer

authentic conversational experiences, and promote collaborative learning among peers, which collectively enhances EFL speaking proficiency.

However, as stated by Pegrum, Hockly and Dudeney (2022), the lack of instant guidance and personalized instruction in navigating digital resources can lead to “digital distraction” due to the overwhelming of information, raising concerns among researchers about the effectiveness of MALL. The distractions could stem from extraneous notifications and the inundation of data from various SNS apps, underscoring the urgent need for constant supervision, guidance, or feedback to navigate and alleviate these interruptions effectively. This issue emphasizes a pivotal challenge in utilizing SNS within the MALL context: balancing and leveraging these platforms’ communicative advantages and minimizing their potential to detract from focused language learning (Kukulska-Hulme et al., 2018; Stockwell, 2022). In response to the issues of the current MALL (SNS MALL), the deployment of sophisticated solutions, such as artificial intelligence (AI) technologies, are imperative (Kukulska-Hulme et al., 2020; Stockwell, 2022; Viberg, Kukulska-Hulme, & Peeters, 2023; Han & Lee, 2024).

AI-powered language learning mobile app. AI technology is defined as “computer systems that have been designed to interact with the world through capabilities (for example, visual perception and speech recognition) and intelligent behaviors (for example, assessing the available information and then taking the most sensible action to achieve a stated goal)” (Luckin et al., 2016, p. 14). There are two types of artificial intelligence (AI): general AI and narrow AI. While general AI embodies the ambitious concept of an intelligent agent that could theoretically understand and master a wide range of human behaviors and intellectual tasks, this comprehensive form of AI remains hypothetical and is not yet operational (Chen et al., 2020). In contrast, narrow AI refers to an intelligent agent designed to excel in specific, limited domains (Pegrum, 2019). The real-world application of narrow AI, particularly in education, has demonstrated its immediate utility. In the educational field, the ubiquity of smartphones and extensive wireless networks has made narrow AI technology readily accessible to learners and teachers via AI mobile apps (Kukulska-Hulme, 2019; 2020).

These AI mobile apps not only support communication functions linked to SNS apps but also feature automatic speech recognition, natural language processing, text-to-speech, and speech-to-text technologies. Such features address issues from conventional SNS apps use and enhance the effectiveness of MALL. AI-powered mobile apps facilitate personalized learning experiences (Kukulska-Hulme, 2019), automated feedback (Reinders & Stockwell, 2017), and adaptive content (Kukulska-Hulme et al., 2020), helping learners focus on their objectives, minimize distractions, and receive continuous support (Hwang et al., 2022; Hwang, Rahimi, & Fathi, 2024). Hwang et al. (2022) introduced an AI-powered mobile app (Smart UEnglish) designed to improve Chinese undergraduates’ EFL speaking skills through structured and free-flow conversations, emphasizing real-life conversational practice. The AI app was tailored for flexible, sustainable, and adaptive conversations to facilitate flexible and adaptive dialog (affordance of MALL). This led to notable improvements in speaking ability and vocabulary acquisition, as reported by participants who also enjoyed increased engagement and practical conversational experiences. In the same vein, Hwang, Rahimi, and Fathi (2024) found in their study that MALL with AI mobile language learning app (i.e., HE app) enhances EFL speaking skills by providing personalized, accessible, and context-rich learning. Functions of the mobile app, like

immediate feedback, real-world language exposure, and practice opportunities, are crucial for oral language improvement. These studies highlight that AI has the potential to develop the affordances of MALL—personalization, collaboration, and authenticity—more effectively than SNS (Kukulska-Hulme, 2024), thereby optimizing the impact of MALL on improving EFL speaking performance.

In the Chinese context, a national AI strategy for education was launched in 2017 as part of the Chinese Next Generation Artificial Intelligence Development Plan (Jing, 2018). This initiative seeks to position China as a worldwide hub for AI innovation by 2030. In line with this, the use of mobile English-learning applications has surged among university students, significantly contributing to the improvement of English speaking skills (Yang & Hu, 2023). In this context, popular Chinese AI-powered English-speaking mobile apps, such as Liulishuo, have attracted significant attention from Chinese researchers for their effects on EFL speaking performance (Green & O'Sullivan, 2019; Tai et al., 2020; Wei, Yang, & Duan, 2022).

While AI applications show great promise in enhancing EFL speaking skills, concrete empirical support in authentic classroom settings remains insufficient (Hwang et al., 2024; Chen et al., 2020). Yang and Kyun (2022) provided a systematic literature review on AI-supported language learning, examining 25 empirical research papers on AI-supported language learning published from 2007 to 2021. Their findings reveal that research has primarily concentrated on technology development and theoretical modeling, emphasizing less on investigating how these technologies affect language learning outcomes in natural classroom settings. This discrepancy underlines a significant gap in existing research, suggesting a need for more comprehensive research into the practical applications of AI mobile apps in education. This includes evaluating the effectiveness and practicality of these technologies.

Furthermore, most research tends to focus on overall speaking performance or specific sub-skills, such as pronunciation and fluency, rather than conducting a comparative analysis to determine which four sub-skills benefit most from AI MALL. For instance, Karim et al. (2023) reported that AI English-learning mobile apps can significantly enhance overall speaking performance. While the study identified vocabulary as a key factor influencing speaking abilities, it did not specifically evaluate the app's impact on this sub-skill. Similarly, Dennis (2024) utilized an AI-powered speech recognition program to improve EFL pronunciation and speaking skills. Both quantitative and qualitative results showed that the AI mobile app enhanced students' pronunciation and overall speaking skills. However, this study also fell short of providing a comprehensive understanding of the impact of AI mobile apps across all four speaking sub-skills. As highlighted by Zhai and Wibowo (2023) in their review study, the impact of AI on EFL learning remains in its early stages, requiring further research to develop a more comprehensive understanding.

Theoretical framework. Vygotskian Sociocultural Theory, developed by Vygotsky and his colleagues from the 1960s to the 1990s, highlights the importance of social interaction, cultural context, and mediation in learning. This theory has been extensively applied in education, especially in foreign language learning (Aljaafreh & Lantolf, 1994; Thorne, 2003; Lantolf et al., 2014). According to Lantolf (2000), Sociocultural Theory views human mental activity as occurring through interactions with cultural peers and mediated by cultural artifacts like tools and symbols. In this study, Vygotskian Sociocultural Theory (1978; 1987) provides the foundation for investigating how advanced mental

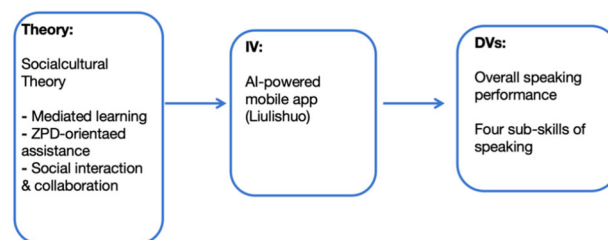


Fig. 1 Conceptual framework of the study.

functions, such as EFL speaking skills, are mediated by symbolic tools (mobile apps) and physical tools (smartphones) during purposeful activities. This mediation facilitates interaction and collaboration with teachers, peers, or authentic language contexts (Sharples et al. (2016); Lantolf, Poehner, & Thorne, 2020), enabling learners to internalize their language experiences and improve speaking performance.

To effectively use these mediating artifacts for enhancing EFL learning among teachers and students, it is crucial to integrate another key Vygotskian concept: the zone of proximal development (ZPD). Defined by Vygotsky (1978) as the difference between what learners can do alone and what they can achieve with guidance from more capable others, ZPD focuses on how teacher or peer support can facilitate students' learning (Abdullah et al., 2022). In the context of AI-powered MALL, the AI app continuously evaluates learners' speaking accuracy and fluency during practice. Based on this evaluation, the app dynamically offered personalized instructions and instant guidance, ensuring that students' practice remains within their ZPD. By continuous personalized feedback to the learner, AI facilitates a process of scaffolding that aligns with Vygotsky's principle of guided learning (Lantolf, 2000; Aljaafreh & Lantolf, 1994; Lantolf, Poehner, & Thorne, 2020).

Specifically, social interaction and collaboration, as elements of Vygotskian Sociocultural Theory, interact with the ZPD and mediated learning in distinct ways between the experimental and control groups in this study. For the experimental group, AI-powered mobile apps simulate interactive experiences by enabling students to engage in language tasks with immediate feedback and personalized guidance, fostering a novel form of learner-app collaboration. In contrast, the control group relies on teacher and peer assessment to encourage interaction and collaboration in the SNS mobile app. However, while the control group depends on teacher and peer interactions mediated by the SNS mobile app to provide general support, the experimental group benefits from AI-mediated scaffolding, which enhances the precision and immediacy of ZPD-oriented assistance. Therefore, AI-powered mobile apps, as mediating tools, have the potential to improve EFL learning achievements, particularly in speaking proficiency, by delivering instant targeted ZPD-oriented assistance to students (Kukulska-Hulme, 2019; 2024). These features improve EFL speaking proficiency by addressing specific sub-skills, including vocabulary, grammar, pronunciation, and fluency (Zou et al., 2023; Sabili et al., 2024). The conceptual framework is presented in Fig. 1, which illustrates the cause-effect relationship between the independent variable (AI-powered mobile apps) and the dependent variable (EFL speaking performance). This relationship is mediated by key theoretical constructs, including ZPD-oriented assistance, social interaction, and collaboration, which are rooted in Sociocultural Theory.

Research methodology

Research design. This investigation employed a quasi-experimental research design, utilizing a pretest-posttest design

with nonequivalent control groups, to explore the effectiveness of an AI-powered mobile app on the EFL speaking abilities of Chinese undergraduate students. Conducted in the spring of 2023, the research took place at a public university in Jilin Province, China. This institution is recognized as a second-tier university, noted for its slightly lower academic standards than first-tier universities. Students at this university typically engage in 90 min of English language instruction twice a week, with an additional 10–15 min spent on after-class assignments. Additionally, each student possesses at least one smartphone, a vital component of their daily academic and personal life, facilitated by continuous access to free Wi-Fi across the campus. To ensure the internal validity of this study, a baseline level was established using the SNS app as a common platform, along with uniform in-class instruction for both the experimental and control groups. This consistent use of the same SNS app (WeChat) ensures that any differences observed in the dependent variables can be attributed to the AI-powered mobile application rather than other variables. The distinction between the groups lies in their method of completing after-class speaking assignments.

Specifically, both the control and experimental groups received the same in-class instruction based on the textbook. For the control group, the teacher shared a short video (including about ten sentences) via the WeChat group. Students practised based on the submitted video recordings and received marks from the teacher via the WeChat group. For the experimental group, the teacher assigned a speaking topic related to the lesson, similar to the control group. Based on the topic provided by the teacher and their proficiency level determined by Liulishuo's initial test (six levels), students freely selected practice materials from Liulishuo's recommendation list for their speaking tasks. Liulishuo provided instant, personalized feedback on their speaking performance. Afterward, students shared their scores and performance details with the teacher through the WeChat group. The only differences between the groups were the speaking materials and feedback sources. The control group received standardized speaking materials and general feedback from the teacher, while the experimental group selected speaking materials aligned with their proficiency level and received personalized feedback from the AI app.

Participants. For this study, two first-grade classes of non-English majors were chosen, each led by instructors with comparable teaching experience. Both instructors had incorporated social networking service (SNS) mobile applications into their English instruction for more than 3 years. The research began after receiving the required approvals from the educational institution. Before collecting data, students were asked to provide their consent electronically.

A pretest was administered to evaluate the students' initial EFL speaking performance. Subsequently, one class was randomly designated as the experimental group and the other as the control group. The experimental group included 32 students (26 females and 6 males), whereas the control group comprised 31 (24 females and 7 males). All the participants were between 18 and 20 years old and identified as digital natives, highlighting their early exposure to and familiarity with digital technology. Both groups were provided with the same course content to maintain consistency in learning materials.

Instruments

English speaking performance. The evaluation of EFL speaking proficiency was conducted using the IELTS speaking test, renowned for its global recognition and direct relevance to real-life English communication scenarios. This assessment focuses on

Table 1 Pearson's correlation for inter-rater reliability between rater A and B on IELTS speaking test scores (pretest and posttest).		
	Pearson correlation (r)	Sig. (two-tailed)
Pretest scores (overall speaking)	0.98**	0.00
V	0.97**	0.00
G	0.98**	0.00
F	0.98**	0.00
P	0.98**	0.00
Posttest scores (overall speaking)	0.97**	0.00
V	0.97**	0.00
G	0.98**	0.00
F	0.99**	0.00
P	0.99**	0.00

V vocabulary, G grammar, P pronunciation, F fluency, **p < 0.01.

critical dimensions of speaking ability, encompassing vocabulary, grammar, pronunciation, and fluency, making it well-suited to address the research questions of this study. Participants were assessed twice during both a pretest and a posttest phase, with each session comprising distinct sets of authentic IELTS inquiries. According to the IELTS speaking test criteria, each of the four sub-skills—fluency, vocabulary, grammar, and pronunciation—is individually scored on a scale from 1 to 9. The final overall score is calculated as the average of these sub-skill scores and is reported in intervals of 0.5. To ensure the reliability and validity of the results, data were collected by four teachers. Two teachers, who served as English instructors for each group, conducted IELTS-speaking interviews with their students and recorded the sessions. The remaining two teachers, each with five years of IELTS teaching experience, reviewed the recordings and provided scores for each participant. Each participant's test lasted around 15 min, during which their responses were audio-recorded and appraised by two experienced evaluators to guarantee assessment uniformity, as reflected by a substantial inter-rater reliability index (Table 1). The appraisal adhered to the IELTS scoring guidelines, affirming the test's credibility in accurately measuring speaking skills.

AI-powered mobile application. For its comprehensive suite of AI functionalities offered without charge, the Liulishuo app was chosen. This app encompasses ten segments designed to bolster English speaking and listening competencies through educational and recreational modules. However, the study specifically focused on the educational segments related to speaking for the after-class assignments in the experimental group. Initially, the Liulishuo app offers users a free opportunity to evaluate their speaking performance twice a week across five dimensions, as depicted in Fig. 2. This allows students to ascertain their level of EFL speaking proficiency. Based on this evaluation, the AI tutor recommends speaking exercises tailored to the student's proficiency levels, or students may select exercises based on their preferences, as illustrated in Fig. 3. The exercises commence with students listening to a chosen recording topic, followed by their attempts to replicate 5–10 sentences. The AI tutor provides automatic feedback and corrective guidance for each sentence reproduced by the students, who can then share these results with their teachers and peers via the SNS app, as illustrated in Fig. 4. Upon completing all the sentences for a topic, students receive an overall score for their speaking performance (Fig. 5), and they are encouraged to practice repeatedly until satisfied with their final assessment

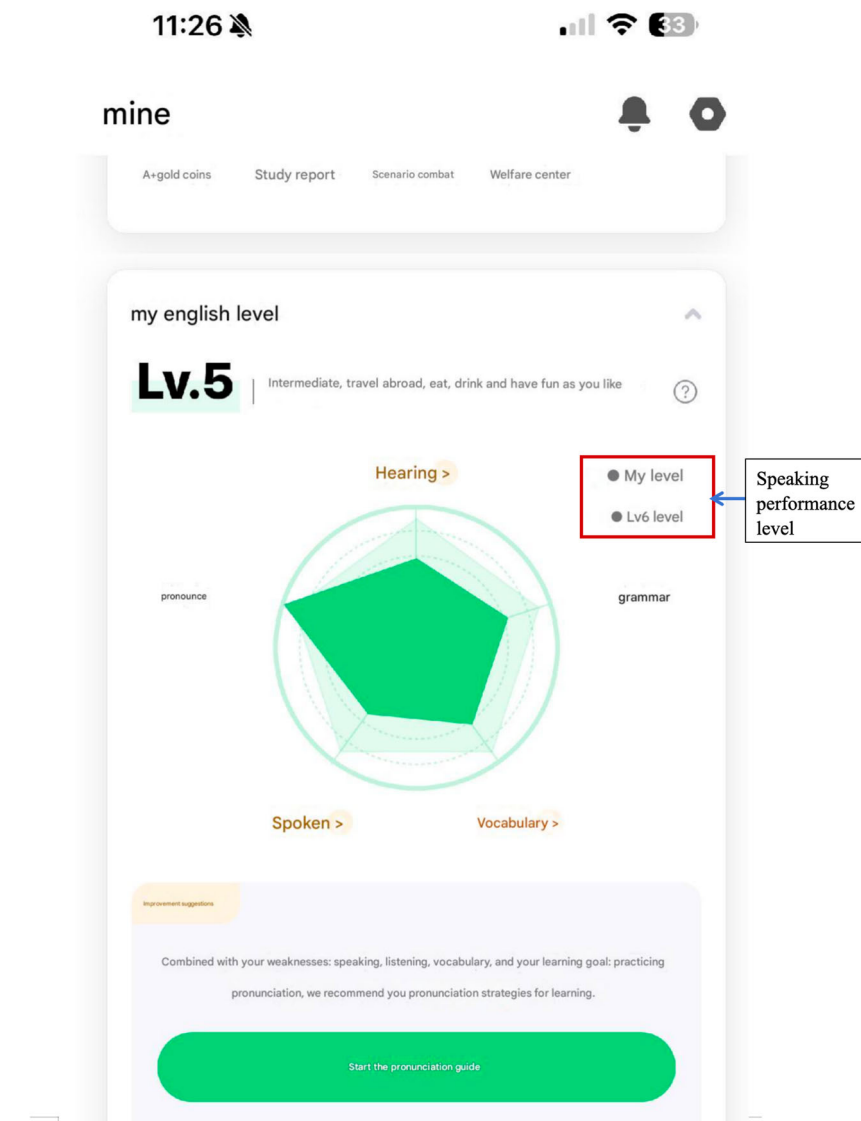


Fig. 2 Translation version of the interface for testing the initial English. Experimental group using the AI-powered mobile app to assess their initial English proficiency.

score. Subsequently, students receive a comprehensive analysis of their speaking performance, including their scores and the duration of their practice. This detailed feedback can be shared with peers or teachers via WeChat, showcasing their learning score or journey, as highlighted in Fig. 6.

Procedure. This study was conducted at the beginning of the second semester of the 2022–2023 academic year; see Fig. 7 for the detailed procedure. The intervention was integrated into the regular English course's after-class assignments for first-year non-English major undergraduates. It allocated at least 20 min weekly to activities focused on English speaking practice, following the university's guidelines for such assignments. The daily teaching method involved alternating weeks of vocabulary and grammar instruction with student presentations and shadowing oral practice in 45 min of regular class time in both experimental and control groups. The only difference between the two groups is how to finish after-class English-speaking assignments.

Specifically, in week 1, after the pretest of the English-speaking test, the experimental group had a 90-min training session on how to utilize different mobile apps for completing their after-

class assignments. The training included an introduction to the Liulishuo app and WeChat and instructions on using Liulishuo and WeChat for post-class assignments. While the control group used WeChat as part of their usual routine to complete their after-class speaking practice assignments. The 10-week intervention ran from week 2 to week 11 of the academic semester. For instance, in week 2, following class instruction, the control group's teacher assigned a two-minute video and text from Obama's school opening speech as an after-class assignment through the WeChat group. Students were instructed to study, listen to, and mimic the speech, then record their best rendition and share the recording with the class WeChat group. Subsequently, the teacher assigned scores to each student and encouraged peer assessments within the group chat. In the experimental group during the same week, after having the in-class instruction, the teacher directed students to log into the Liulishuo app, complete an English speaking proficiency test to get the English speaking initial level before the practice, and then ask the students to choose speaking practice materials that appealed to them and matched their speaking test level. During the practice, students can get the scores, instance feedback and personalized guidance from the Liulishuo app. Upon completing

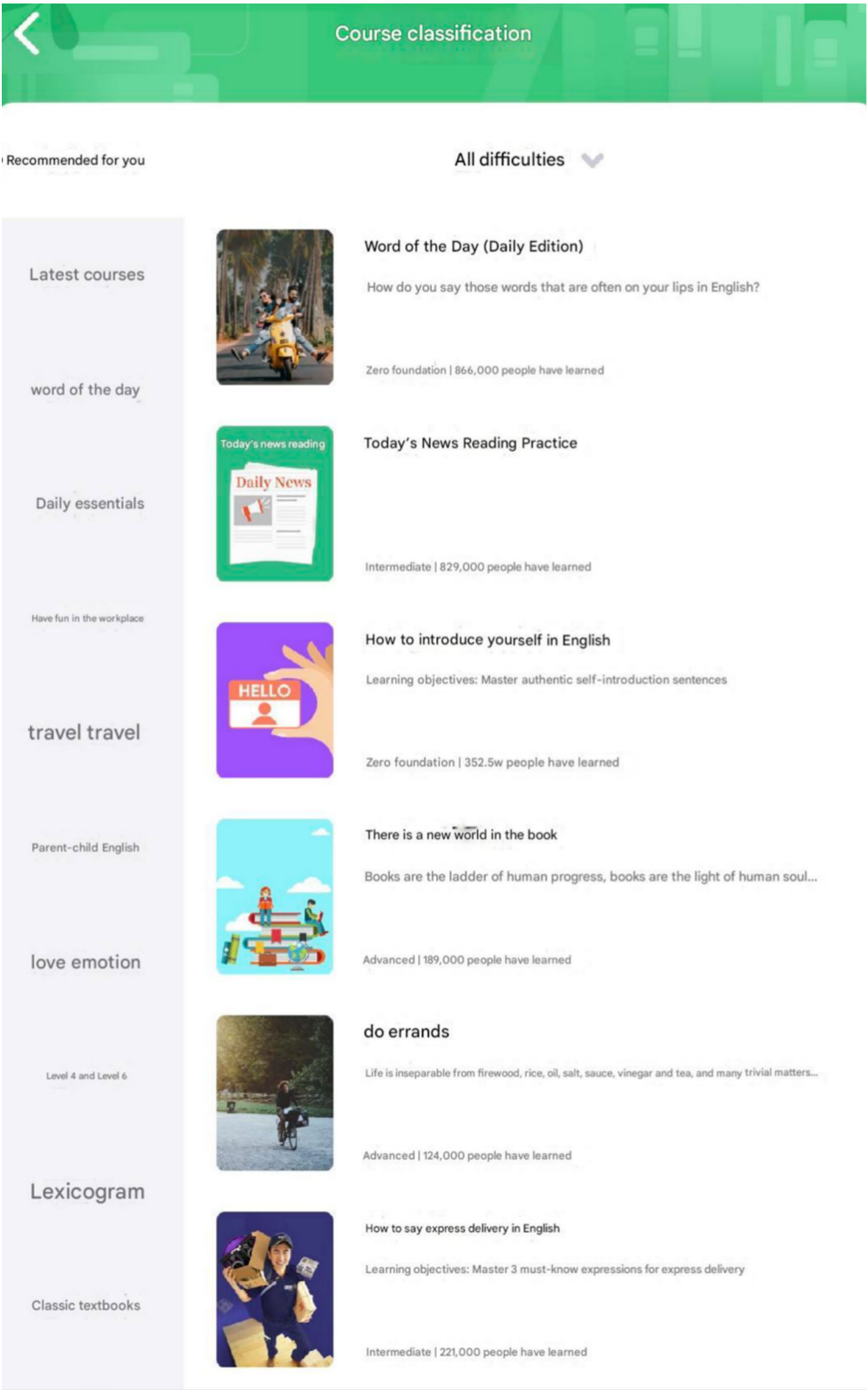


Fig. 3 Translation version of the interface for practice resources. Experimental group using the AI-powered mobile app to select the learning materials.

their speaking practice, students were encouraged to share their scores and learning path reports via the WeChat group, allowing the teacher to monitor their progress. Each subsequent intervention week, they followed the same structure for the control and experimental groups. The videos sent to the control group were

derived from instructional materials typically used in regular practice sessions. Meanwhile, the experimental group selected listening materials entirely based on their preferences, guided by their weekly English-speaking test results. No time constraints were imposed on either group; both groups had the flexibility to

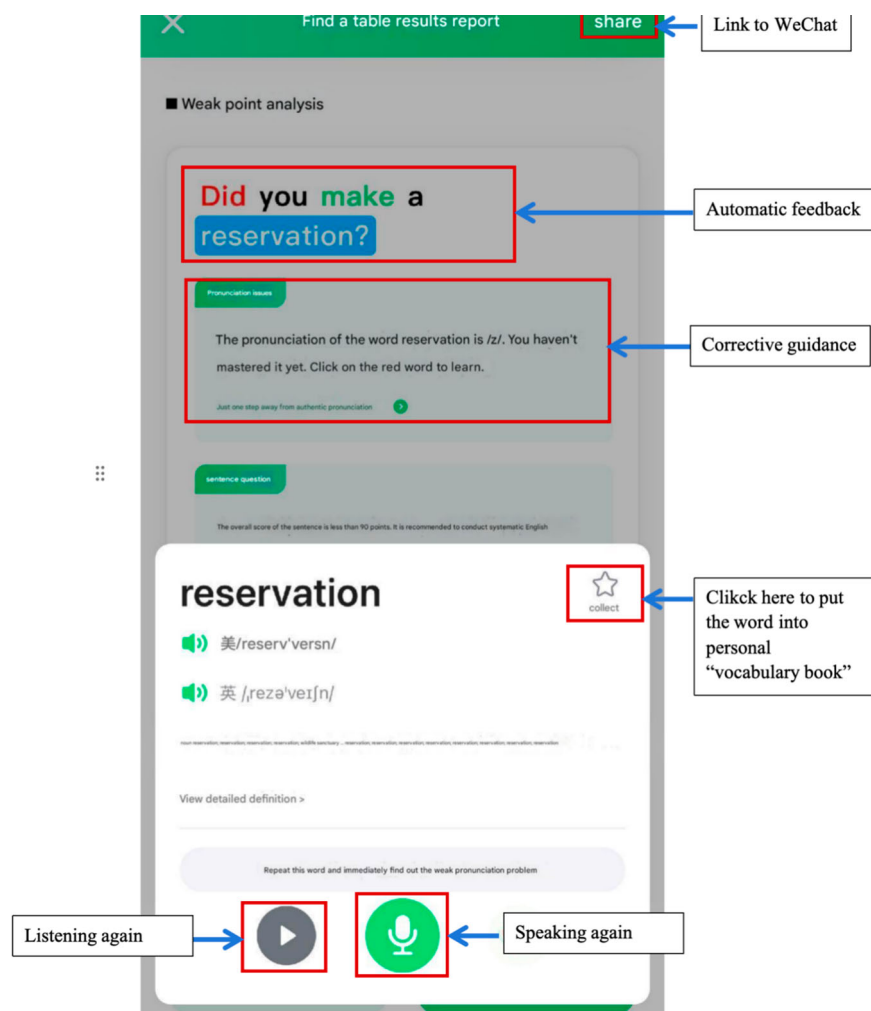


Fig. 4 Translation version of the interface for automatic feedback and corrective guidance. Experimental group using the AI-powered mobile app to practice English speaking.

Table 2 Independent samples *t*-test for overall speaking performance pretest scores.

	Group	Mean	SD	<i>t</i> (61)	<i>p</i>
Overall speaking pretest scores	Experimental	5.17	0.84	-1.58	0.12
	Control	4.84	0.83		

complete the assignments at a time and place. Following the intervention, the final week (Week 12) was designated for the posttest of English-speaking performance for both groups.

Data analysis. This study utilized SPSS 27 for all statistical analyses. Initially, an independent samples *t*-test was applied to the pretest scores of speaking performance between the experimental and control groups to verify no significant differences in speaking abilities at the start. To explore the first research question, an independent samples *t*-test was performed on the posttest scores to measure improvement in the experimental group. For the second research question, a multivariate analysis of variance (MANOVA) was employed to assess progress in the four speaking sub-skills between the groups after the intervention.

Results

Overall EFL speaking performance. An independent *t*-test was carried out on the pretest speaking performance scores between the two groups, as shown in Table 2. The analysis revealed no significant difference in overall speaking performance ($t(61) = -1.58, p > 0.05$) between the two groups. This result highlights the statistical equivalence regarding their overall speaking pretest scores between the groups before the intervention.

To answer the first research question, a paired sample *t*-test was constructed to compare the pretest and posttest overall speaking performance within both two groups. Relevant data from these tests can be found in Table 3. The results of the paired sample *t*-tests indicate significant differences between the pretest and posttest mean scores for experimental group ($t(31) = -11.96, p = 0.00$) with a mean difference of $-0.71, 95\% \text{ CI} [-0.82, -0.58]$ and control group ($t(30) = -5.36, p = 0.00$) with a mean difference of $-0.39, 95\% \text{ CI} [-0.53, -0.24]$. However, the experimental group (Cohen's $d = -2.11$) demonstrated a larger effect size compared to the control group (Cohen's $d = -0.96$). This suggests improvements in overall speaking performance after the intervention in both groups. Furthermore, to determine which app contributed more effectively to speaking performance, an independent samples *t*-test was conducted on the posttest scores of speaking performance between the control and experimental groups. The results, presented

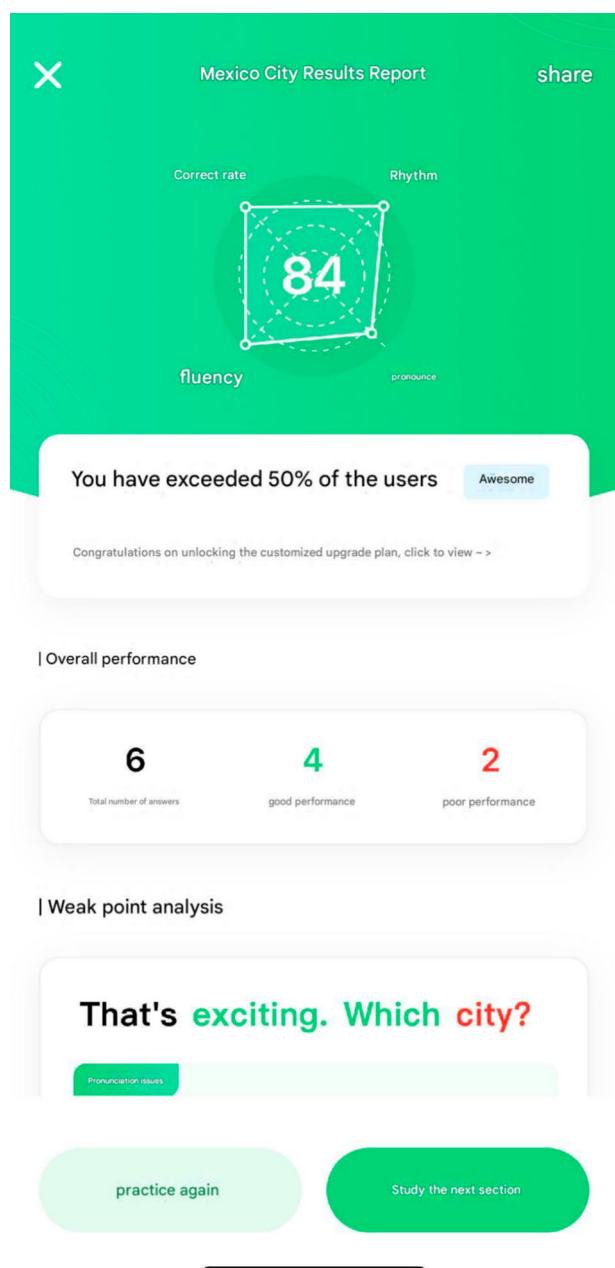
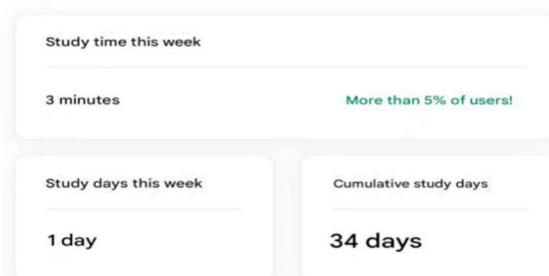


Fig. 5 Translation version of the interface for overall scores of speaking practice. Experimental group using the AI-powered mobile app to obtain the assessment scores.

in Table 4, revealed significant differences ($t(61) = 2.85, p = 0.006 < 0.05$), indicating that the experimental group (Mean = 5.88, Standard deviation = 0.91) got higher speaking performance scores than the control group (Mean = 5.23, Standard deviation = 0.90) after the intervention, with a 95% confidence interval ranging from 0.19 to 1.11. The effect size was large (Cohen's $d = 0.72$).

Four sub-skills of EFL speaking performance. To further investigate which sub-skills of the English-speaking performance test exhibited significant effects due to using the AI-powered mobile app, a series of independent sample t -tests and a multivariate analysis of covariance (MANOVA) were performed in this study. Initially, a series of independent samples t -tests were conducted with a Bonferroni adjustment, setting the alpha level at 0.01 (0.05 divided by 4 to account for multiple

my week



Data details

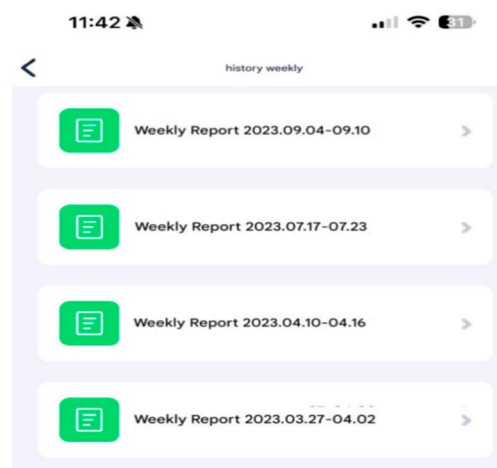
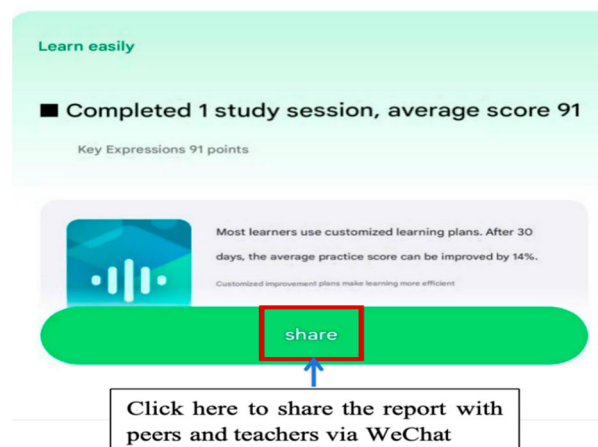


Fig. 6 Translation version of the interface for comprehensive analysis of the learning path report. Experimental group using the AI-powered mobile app to get the report.

comparisons). An independent t -test was conducted on the pretest scores of the four sub-skills between the two groups, as shown in Table 5. The analysis indicated that there were no significant differences in the pretest scores of the four sub-skills between the groups ($P_s > 0.01$). This outcome confirms that there were no significant differences in the pretest scores for the four sub-skills between the two groups before the intervention, establishing homogeneity. Subsequently, a series of independent t -tests were conducted on the posttest scores for the four sub-skills to evaluate changes in speaking performance between the two groups after the intervention. According to Table 6, the

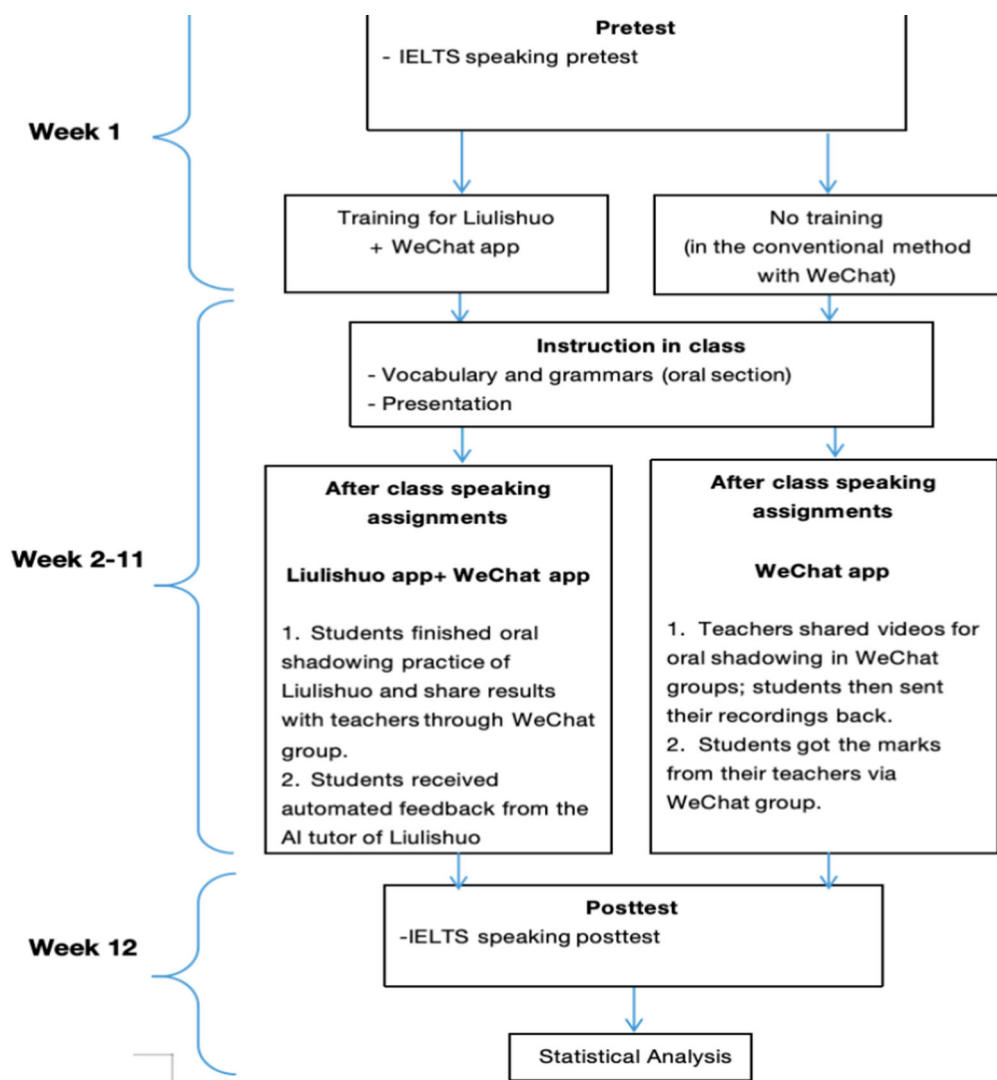


Fig. 7 Experimental procedure. Research methodology flowchart.

Table 3 Paired sample *t*-test comparing speaking performance pre and posttest.

Test	Control group				Experimental group			
	Mean	SD	<i>t</i> (30)	<i>p</i> (two-tail)	Mean	SD	<i>t</i> (31)	<i>p</i> (two-tail)
Pre	4.84	0.83	−5.36	0.00	5.17	0.84	−11.96	0.00
Post	5.23	0.90			5.88	0.91		

independent *t*-tests for posttest scores revealed significant differences in pronunciation ($F(1, 60)$, $p = 0.00 < 0.01$, Cohen's $d = 0.42$) and fluency ($F(1, 60)$, $p = 0.01 < 0.01$, Cohen's $d = 0.84$), using a Bonferroni-adjusted alpha level of $p = 0.01$. However, there were no significant differences in grammar ($F(1, 60)$, $p = 0.06 > 0.01$, Cohen's $d = 0.48$) and vocabulary ($F(1, 60)$, $p = 0.08 > 0.01$, Cohen's $d = 0.44$) between the experimental and control groups. Cohen's d values of 0.42 and 0.84 indicate a small to medium effect for pronunciation and a larger effect for fluency, respectively. This suggests that the experimental intervention had a stronger impact on fluency, while its effect on pronunciation was relatively weaker.

MANOVA was conducted to confirm whether four sub-skills differed significantly between control and experimental after the

intervention. Prior to conducting the MANOVA, its assumptions were rigorously checked, encompassing normality, outliers, linearity, multicollinearity, homogeneity of variance, and homogeneity of variance-covariance matrices. All these assumptions were met (see Appendix A), thereby qualifying the adjusted mean scores for further analysis through the subsequent MANOVA. The multivariate test of the posttest scores across four sub-scores (Wilk's Lambda) revealed a significant main effect distinguishing the experimental group from the control group across all four sub-scores ($F(4, 58) = 4.73$, $p = 0.000$, Wilks' $\lambda = 0.75$, $\eta^2 = 0.25$). Further analysis of between-subjects effects aimed to discern differences within these sub-scores. Employing a Bonferroni-adjusted alpha level of 0.0125 (Pallant, 2005), the outcomes detailed in Table 7 from this one-way MANOVA showed no significant impact of the AI-powered mobile

Table 4 Independent samples t-test for English speaking posttest scores between Experimental (E) and Control (C) groups.						
Group	Leneve's test for equality of variances			t-test for equality of means		
	F	Sig.	t(61)	Sig. (two-tail)	Mean difference	Std. error difference
E	0.001	0.97	2.85	0.006	0.65	0.23
C						

Table 5 Independent samples t-test for four sub-skills pretest scores between Experimental (E) and Control (C) groups.					
Sub-skills	Group	Mean	SD	t(61)	p (two-tail)
V	C	5.15	0.96	0.22	0.83
	E	5.09	0.93		
G	C	4.77	0.91	−1.18	0.24
	E	5.05	0.92		
P	C	4.84	0.93	−1.52	0.13
	E	5.19	0.89		
F	C	4.76	1.03	−1.78	0.08
	E	5.20	0.96		
V vocabulary, G grammar, P pronunciation, F fluency.					

Table 6 Independent samples t-test for four sub-skills posttest scores between Experimental (E) and Control (C) groups.					
Sub-skills	Group	Mean	SD	t(61)	p (two-tail)
V	C	5.61	0.82	1.79	0.08
	E	6.02	0.98		
G	C	5.03	1.01	1.91	0.06
	E	5.50	0.92		
P	C	5.24	1.02	3.73	0.00
	E	6.21	1.05		
F	C	4.97	1.05	3.34	0.01
	E	5.89	1.14		
V vocabulary, G grammar, P pronunciation, F fluency.					

Table 7 MANOVA results for the sub-skills of overall speaking performance.						
sub-skill	Type III sum of squares	df	Mean square	F	Sig.	Partial Eta squared (η^2)
V posttest	2.55	1	2.55	3.11	0.083	0.05
G posttest	3.45	1	3.45	3.66	0.061	0.06
P posttest	15.02	1	15.02	13.91	0.000	0.19
F posttest	13.41	1	13.41	11.16	0.001	0.16
V vocabulary, G grammar, P pronunciation, F fluency.						

application on Vocabulary ($F(1, 61) = 3.11, p = 0.083 > 0.0125, \eta^2 = 0.05$) and Grammar ($F(1, 61) = 3.66, p = 0.061 > 0.0125, \eta^2 = 0.057$) in comparison to the control group. Conversely, significant improvements were noted in Pronunciation ($F(1, 61) = 13.91, p = 0.000 < 0.0125, \eta^2 = 0.19$) and Fluency ($F(1, 61) = 11.16, p = 0.001 < 0.0125, \eta^2 = 0.16$) within the experimental group. The effect sizes for Vocabulary and Grammar are small to medium, whereas, for Pronunciation and Fluency, they are medium to large. These findings elucidate that the AI-powered mobile app distinctly bolstered pronunciation and fluency, albeit without markedly affecting vocabulary and grammar, relative to the control group.

Discussion

This study enriches the empirical evidence by examining the integration of AI-powered mobile applications in the context of mobile-assisted language learning (MALL), offering a detailed examination of how such AI tools impact EFL speaking performance in the context of China. For the first research question, the

primary finding of this study underscores a significant improvement in EFL speaking performance across both the control and experimental groups. However, a comparative analysis of posttest scores revealed that the experimental group significantly outperformed the control group, indicating that the AI-powered mobile app significantly enhances students' EFL speaking performance compared to the SNS mobile app. These results echo the findings from previous research (Stockwell, 2022; Ferguson et al., 2022; Fathi et al., 2024), reinforcing the efficacy of AI applications in enhancing language learning outcomes. Stockwell (2022) claimed that AI technology offers significant advantages over conventional communication mobile apps in enhancing language learning within the MALL context. Ferguson et al. (2022) found that AI-powered gaming apps offering simple tasks can enhance learning outcomes more effectively than non-AI apps by engaging students actively. Similarly, Fathi et al. (2024) reported that AI-mediated activities offer environments for more frequent and targeted practice, with instant feedback tailored to individual needs, thus boosting speaking performance. Correspondingly, this study confirms that AI-powered mobile apps significantly enhance speaking skills by consistently delivering personalized, ZPD-oriented instructions and feedback, improving interactions in the target language.

Additionally, for the second research question, in this study, the AI-powered mobile app demonstrated significant improvements in pronunciation and fluency sub-skills of speaking performance among Chinese EFL undergraduate students, as evidenced by the considerable enhancements in posttest scores for these areas. However, the impact on vocabulary and grammar sub-skills, while showing a trend toward improvement, did not reach statistical significance. Regarding this, this study is partially consistent with prior studies by Junaidi (2020) and Zou et al. (2023). These studies demonstrated that AI-powered mobile apps can significantly enhance all four speaking sub-skills.

The rationale for the partially consistent results of this study with previous research can be attributed to the MALL context. Many existing studies lack a control group or use non-MALL instruction as the control, which does not provide a fair comparison to the AI-powered mobile app, thus affecting the clarity of results regarding AI effectiveness across different linguistic aspects. Sharples et al. (2016) highlight how smartphones with various apps have transformed language learning inside and outside traditional settings (Kukulska-Hulme, 2009). Despite the prevalence of mobile-assisted language learning (MALL), many studies do not consider this context (Ekinici, 2020; Viberg, Kukulska-Hulme (2022); Mihaylova et al., 2022). This study, therefore, uniquely positions its research within the MALL context and designs a control group that also uses a common SNS mobile app, allowing for a more relevant comparison and better understanding of AI-powered tools in language learning. Moreover, the mixed results of this study can be understood through its theoretical framework. Vygotsky's sociocultural theory posits that the more ZPD-oriented assistance learners receive, the greater their learning achievements (Stockwell, 2022; Kukulska-Hulme, 2024). In this context, Liulishuo, an AI-powered language learning app, provides more personalized and instant feedback on learners' speaking performance compared to conventional SNS apps, with a particular focus on pronunciation and fluency. However, it lacks targeted assistance for grammar and vocabulary, a limitation rooted in its feature set. For instance, while Liulishuo offers instant feedback and guidance based on the analysis of students' recordings for fluency and pronunciation, it provides limited analysis and support for grammar and vocabulary. This limitation may have contributed to the pronounced improvement in pronunciation and fluency while leading to the non-significant effects on vocabulary and grammar development.

Although the effects on vocabulary and grammar sub-skills were not statistically significant, the AI-powered mobile app demonstrated notable improvements in overall speaking performance, fluency, and pronunciation, underscoring its potential benefits for EFL learning, particularly in enhancing speaking skills. The app provides students with ample opportunities to practice English speaking through tailored, individualized learning experiences. The findings suggest that integrating AI tools into after-class assignments can offer valuable guidance for educators on effectively incorporating technology into their teaching practices. For app developers, the results highlight areas for further optimization, encouraging the enhancement of features to support all four speaking sub-skills comprehensively.

Conclusion

This study aimed to enhance EFL speaking performance through the AI mobile app Liulishuo, which provides real-time feedback and personalized speaking tasks tailored to individual learner needs. The findings of this study show the significant effects of the Liulishuo on improving EFL overall speaking performance when compared to the conventional communication mobile app. Furthermore, it explores the effectiveness of the AI app across four speaking sub-skills: vocabulary, grammar, pronunciation, and fluency. The findings indicate that Liulishuo significantly improves pronunciation and fluency sub-skills, likely due to its instance feedback and personalized instruction mechanisms, but shows no marked effects on vocabulary and grammar. These insights extend the discourse on employing advanced technologies for language learning, particularly considering the non-significant results for vocabulary and grammar in this study. Future studies should place more emphasis on the comparison among the four sub-skills of speaking performance, specifically investigating how AI-powered mobile apps like Liulishuo can be further optimized to enhance vocabulary and grammar, as well as examining the underlying factors contributing to the varying levels of improvement across different speaking sub-skills.

Within the context of this study, it is essential to acknowledge certain limitations that might impact the findings. First, the generalizability of the results may be limited by the specific characteristics of the participants involved. Secondly, there is a technological limitation due to the use of only one AI-powered mobile application, Liulishuo, in the experimental group, primarily focusing on its free features because of budget constraints. This restriction may have limited the app's ability to comprehensively target all four sub-skills, particularly vocabulary and grammar, which often require more structured and diverse task types than those provided in the free version. Different AI apps with distinct features may lead to different results. Third, the ten-week intervention provides only a static perspective, which may not capture the progression of EFL speaking performance over a longer period. These limitations highlight areas for future research. To improve the generalizability of the findings, future studies should include a more diverse participant pool across different countries, universities, and academic levels. Additionally, exploring how various AI applications influence outcomes could deepen our understanding of the effectiveness of AI in enhancing specific language skills, offering valuable insights for the development of more precise and effective language learning tools. Moreover, future research should consider implementing a longitudinal time-series experimental design to validate the results over an extended period, thereby providing more robust and reliable conclusions.

Ultimately, the findings of this study highlight the potential of an AI-powered tool (Liulishuo) in enhancing EFL speaking performance and suggest directions for further research and practical application. Researchers are encouraged to investigate the impact

of Liulishuo and other AI-powered tools on all four sub-skills, particularly vocabulary and grammar, which showed non-significant results in this study. This could involve examining how additional features, such as more diverse and context-specific tasks, could be integrated to address these sub-skills effectively. For practitioners, the findings suggest that integrating AI tools like Liulishuo into after-class assignments can provide instance feedback and personalized instruction, thereby enhancing students' speaking proficiency. Educators are encouraged to integrate AI-powered apps like Liulishuo into traditional teaching methods, leveraging their features to address specific student needs.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding authors upon reasonable request.

Received: 16 June 2024; Accepted: 27 February 2025;

Published online: 15 March 2025

References

- Abdullah Z, Mohamad Said MNH (2022) Engaging and empowering Malaysian students through open and distance learning in the post-COVID era. In *Frontiers in Education* (vol. 7, p. 853796). Frontiers Media SA
- Ahmadi P, Samad AA, Noordin N (2013) Identity formation of TEFL graduate students through oral discourse socialization. *Theory Pract Lang Stud* 3(10):1764–1769
- Aljaafreh A, Lantolf JP (1994) Negative feedback as regulation and second language learning in the zone of proximal development. *Mod Lang J* 78(4):465–483
- Amoah S, Yeboah J (2021) The speaking difficulties of Chinese EFL learners and their motivation towards speaking the English language. *J Lang Linguistic Stud* 17(1):56–69
- Bortoluzzi M, Bertoldi E, Marenzi I (2021) (Re-)contextualising storytelling with children in English L2: mobile-assisted language teacher education. In *Mobile assisted language learning across educational contexts*. Routledge
- Brown HD, Lee H (2015) *Teaching by principles: an interactive approach to language pedagogy*, 4th edn. Boston: Pearson Education, Inc
- Burston J (2015) Twenty years of MALL project implementation: a meta-analysis of learning outcomes. *ReCALL* 27(1):4–20
- Butler YG, Lee J, Peng X (2022) Failed policy attempts for measuring English speaking abilities in college entrance exams: cases from China, Japan, and South Korea. *Engl Today* 38(4):271–277
- Cai Y, Zhang LJ (2023) Effects of mobile-supervised question-driven collaborative dialogues on EFL learners' communication strategy use and academic oral English performance. *Front Psychol Educ Psychol* 2023:1142651
- Chen X, Xie H, Zou D, Hwang GJ (2020) Application and theory gaps during the rise of artificial intelligence in education. *Comput Educ Artif Intell* 1(10002):1–20
- Chwo GSM, Marek MW, Wu WCV (2018) Meta-analysis of MALL research and design. *System* 74:62–72
- Dennis NK (2024). Using AI-powered speech recognition technology to improve english pronunciation and speaking skills. *Technol Edu* 12:107
- EI Shazly R (2021) Effects of artificial intelligence on English speaking anxiety and speaking performance: a case study. *Expert Syst* 38(3):1–15
- Ekinci E (2020) Mobile assisted language learning (MALL) As a foreign language aid: a literature review. *J Soc Sci Humanit* 5(9):152–166
- Evita D, Muniroh S, Suryati N (2021) The correlation between EFL students' use of ICT level and their autonomous learning level. *J Lang Lit Arts* 1(11):1479–1493
- Fathi J, Rahimi M, Derakhshan A (2024) Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System* 121:103254
- Ferguson C, van den Broek EL, van Oostendorp H (2022) AI-induced guidance: preserving the optimal zone of proximal development. *Comput Educ Artif Intell* 3:1–9
- Foroutan M, Noordin N (2012) Effect of dialogue journal writing through the use of conventional tools and e-mail on writing anxiety in the ESL context. *Engl Lang Teach* 5(1):10–19
- Green A, O'Sullivan B (2019) Language learning gains among users of English Liulishuo. LAIX
- Hafour MF (2022) Interactive digital media assignments: effects on EFL learners' overall and micro-level oral language skills. *Comput Assist Lang Learn*. <https://doi.org/3610.top/10.1080/09588221.2022.2067180>
- Han J, Lee D (2024) Research on the development of principles for designing elementary English speaking lessons using artificial intelligence chatbots. *Humanit Soc Sci Commun* 11(1):1–15
- Hwang GJ, Rahimi M, Fathi J (2024) Enhancing EFL learners' speaking skills, foreign language enjoyment, and language-specific grit utilising the affordances of a MALL app: a microgenetic perspective. *Comput Educ* 214(2024):1–28
- Hwang WY, Guo BC, Hoang A, Chang CC, Wu NT (2022) Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Comput Assist Lang Learn*. <https://doi.org/10.1080/09588221.2022.2095406>
- Jao CY, Yeh HC, Huang WR, Chen NS (2022) Using video dubbing to foster college students' English-speaking ability. *Comput Assist Lang Learn* 37(1–2):1–23
- Jing M (2018) China looks to school kids to win the global AI race. *South China Morning Post*. <https://www.scmp.com/tech/china-tech/article/2144396/china-looks-school-kids-win-global-ai-race>
- Junaidi J (2020) Artificial intelligence in EFL context: rising students' speaking performance with Lyra virtual assistance. *Int J Adv Sci Technol Rehabil* 29(5):6735–6741
- Kang H, Lin X (2019) Lifelong learning on the go: English language mobile learning in China. *New Dir Adult Contin Educ* 2019(62):49–60
- Karim SA, Hamzah AQS, Anjani NM, Prianti J, Sihole IG (2023) Promoting EFL students' speaking performance through ELSA speak: an artificial intelligence in English language learning. *J Lang Lang Teach* 11(4):655–668
- Khodabandeh F (2022) Exploring the applicability of virtual reality-enhanced education on extrovert and introvert EFL learners' paragraph writing. *Int J Educ Technol High Educ* 19(1):27
- Kukulska-Hulme A (2009) Will mobile learning change language learning? *ReCALL*, 21(2):157–165
- Kukulska-Hulme A (2016) Mobile assistance in language learning: a critical appraisal. In: Agnieszka P, Mohamed A (eds) *The international handbook of mobile-assisted language learning*. China Central Radio & TV University Press Co., Ltd, pp. 138–160
- Kukulska-Hulme A (2019) Intelligent assistants in language learning: friends or foes? In: *World conference on mobile and contextual learning*. The Open University, pp. 127–131
- Kukulska-Hulme A (2024) A hopeful future for mobile language learning. In: *the Bloomsbury handbook of language learning and technology*. Bloomsbury Academic
- Kukulska-Hulme A, Beirne E, Conole G, Costello E, Coughlan T, Ferguson R, Whitelock D (2020) *Innovating pedagogy 2020: Open university innovation report 8*. The Open University
- Kukulska-Hulme A, Chen Y (2018) Rethinking distance education in China: mobile-assisted language learning designs in a blended foreign language curriculum. In: Palalas A (ed) *Blended language learning: international perspectives on innovative practices*. Beijing: China Central Radio and Television University Press Co., Ltd, pp. 1–20
- Lantolf JP (2000) Second language learning as a mediated process. *Language Teaching* 33(2):79–96
- Lantolf JP, Poehner ME, Thorne SL (2020). Sociocultural theory and L2 development. In: *theories in second language acquisition*. Routledge, pp. 223–247
- Lantolf JP, Thorne SL, Poehner ME (2014) Sociocultural theory and second language development. In: *Theories in second language acquisition*. Routledge, pp. 221–240
- Leong LM, Ahmadi SM (2017) An analysis of factors influencing learners' english speaking skill. *Int J Res English Educ* 2:34–41
- Li F, Fan S, Wang Y (2022) Mobile-assisted language learning in Chinese higher education context: a systematic review from the perspective of the situated learning theory. *Educ Inf Technol* 27(7):9665–9688
- Luckin R, Holmes W, Griffiths M, Forcier LB (2016) *Intelligence unleashed: an argument for AI in education*. Pearson, London
- Meyasa KY, Santosa MH (2023) Telegram as a social networking service (SNS) for students' speaking activity in the English as a Foreign Language (EFL) context. In: Made Hery S (ed) *Pedagogy-driven technology integration in English language teaching*. Nilacakra Publisher, pp. 105–119
- Mihaylova M, Gorin S, Reber TP, Rothen N (2022) A meta-analysis on mobile-assisted language learning applications: benefits and risks. *Psychol Belgica* 62(1):252
- Morgana V, Kukulska-Hulme A (eds) (2021) *Mobile assisted language learning across educational contexts*, 1st edn. Routledge
- Nam B, Jiang X (2023) Aspiration for cosmopolitan capital and ambiguous loss: Chinese exchange students' experiences in US higher education institutions during the COVID-19 pandemic. *Asia Pac J Educ* 43(4):1256–1271

- Nami F (2020) Educational smartphone apps for language learning in higher education: students' choices and perceptions. *Australas J Educ Technol* 36(4):82–95
- Okumura S (2022) Toward successful telecollaboration using SNSs in EFL instruction: what elements should be incorporated and considered? *Informat Commun Stud* 65:1–10
- Pallant J (2005) SPSS survival manual: a step by step guide to data analysis using SPSS for Windows. Australia: Allen & Unwin
- Peeters W, Pretorius M (2020) Facebook or fail-book: exploring “community” in a virtual community of practice. *ReCALL* 32(3):291–306
- Pegrum M (2019) *Mobile lenses on learning*. Springer, Singapore
- Pegrum M, Hockly N, Dudeney G (2022) *Digital Literacies* (2nd ed.), Routledge. <https://doi.org/10.4324/9781003262541>
- Pu P, Chang DYS (2023) Effects of different input modes on blended EFL speaking instruction: a quasi-experimental study. *Comput Assist Lang Learn*. <https://doi.org/10.1080/09588221.2023.2273853>
- Rajendran T, Yunus MM (2021) A systematic literature review on the use of mobile-assisted language learning (MALL) for enhancing speaking skills among ESL and EFL learners. *Int J Acad Res Progress Educ Dev* 10(1):586–609
- Reinders H, Pegrum M (2017) Supporting language learning on the move. In: *SLA research and materials development for language learning*. Routledge, pp. 219–232
- Reinders H, Stockwell G (2017) Computer-assisted SLA. In: *the Routledge handbook of instructed second language acquisition*. Routledge, pp. 361–375
- Sabili AS, Nazri MA, Surayya SA, Wati L (2024) Smart learning: the impact of ai technologies on english language proficiency in the digital era. *J Rev Pendidikan dan Pengajaran* 7(4):14040–14049
- Sharples M, de Roock R, Ferguson R, Gaved M, Herodotou C, Koh E, Kukulska-Hulme A, Looi C-K, McAndrew P, Rienties B, Weller M, Wong LH (2016) *Innovating pedagogy 2016: exploring new forms of teaching, learning and assessment, to guide educators and policy makers* (Open University Innovation Report 5). The Open University
- Shortt M, Tilak S, Kuznetcova I, Martens B, Akinkuolie B (2023) Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020. *Comput Assist Lang Learn* 36(3):517–554
- Stockwell G (2022) *Mobile assisted language learning: concepts, contexts and challenges*. Cambridge University Press
- Sun Z, Lin CH, You J, Shen HJ, Qi S, Luo L (2017) Improving the English-speaking skills of young learners through mobile social networking. *Comput Assist Lang Learn* 30(3–4):304–324
- Tai TY, Chen HHJ, Todd G (2022) The impact of a virtual reality app on adolescent EFL learners' vocabulary learning. *Comput Assist Lang Learn* 35(4):892–917
- Thorne SL (2003) Artifacts and cultures-of-use in intercultural communication. *Lang Learn Technol* 7(2):38–67
- Ur P (2002) The English teacher as professional. In: *Methodology in language teaching: an anthology of current practice*. Cambridge Univ. Press, pp. 388–392
- Viberg O, Kukulska-Hulme A, Peeters W (2023) Affective support for self-regulation in mobile-assisted language learning. *Int J Mob Blend Learn (IJMBL)* 15(2):1–15
- Viberg O, Kukulska-Hulme A (2022) Fostering learners' self-regulation and collaboration skills and strategies for mobile language learning beyond the classroom. In: *the Routledge handbook of language learning and teaching beyond the classroom*. Routledge, pp. 142–154
- Vygotsky L (1978) Interaction between learning and development. *Read Dev Child* 23(3):34–41
- Vygotsky LS (1987) *The collected works of LS Vygotsky: the fundamentals of defectology*, vol. 2. Springer Science & Business Media
- Wang H, Smyth R, Cheng Z (2017) The economic returns to proficiency in English in China. *China Econ Rev* 43(12):91–104
- Wei J, Yang H, Duan J (2022) Investigating the effects of online English film dubbing activities on the intelligibility and comprehensibility of Chinese students' English pronunciation. *Theory Pract Lang Stud* 12(9):1911–1920
- Wei R, Su J (2012) The statistics of English in China: an analysis of the best available data from government sources. *Engl Today* 28(3):10–14
- Yang H, Kyun S (2022) The current research trend of artificial intelligence in language learning: a systematic empirical literature review from an activity theory perspective. *Australas J Educ Technol* 38(5):180–210
- Yang X, Hu J (2023) Chinese college students' preferences for mobile-assisted language learning classes and their effects on student engagement in English language classrooms. *Innov Lang Learn Teach* 17(5):953–966
- Zhai C, Wibowo S (2023) A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Comput Educ Artif Intell* 4:100134
- Zhou Z (2021) A systematic literature review on the use of mobile-assisted language learning (MALL) for enhancing speaking skills in Chinese EFL context. *Int J Front Sociol* 3(15):12–24
- Zhu D (2022) English performance of Chinese University students The influence of different factors on English learning. In: *2021 international conference on education, language and art (ICELA)*, Atlantis Press, pp. 843–848
- Zou B, Guan X, Shao Y, Chen P (2023) Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability* 15(4):2872
- Zou B, Du Y, Wang Z, Chen J, Zhang W (2023) An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *Sage Open* 13(3):1–9

Acknowledgements

The authors thank all the participants for their cooperation in data collection. Also we are grateful to all the reviewers and editors for their precious suggestions and comments on the manuscript.

Author contributions

All the authors contributed equally to this study.

Competing interests

The authors declare no competing interests.

Ethical approval

This study was conducted as part of the author's PhD research under the supervision of Universiti Putra Malaysia (UPM), and it complies with the ethical standards set by the Ethics Committee for Research Involving Human Subjects at UPM (JKEUPM). The ethical approval for this study was granted under reference number JKEUPM-2022-910 on 15 December 2022, with approval valid until 15 December 2023. The study involves university students from a Chinese university and investigates the effects of the artificial intelligence mobile application Liulishuo on speaking performance. Data collection activities, such as questionnaires and speaking tests, were included in the approved research plan. Prior to the study's initiation at Jilin Technology and Business University, written permission was obtained from the university, confirming institutional support for the research and its application to the doctoral thesis at Universiti Putra Malaysia. Formal ethical approval was granted by Universiti Putra Malaysia, and the study adhered to the Declaration of Helsinki's guidelines, ensuring compliance with both international ethical standards and local regulations. By securing institutional approval from Jilin Technology and Business University, the study ensured it met the ethical and regulatory requirements of both the home institution (UPM) and the host institution (Jilin Technology and Business University).

Informed consent

Written informed consent was obtained from all adult participants prior to their involvement in the study. The consent process was conducted in April 2023 by the principal researcher. Participants were provided with an information sheet detailing the study's objectives, procedures, potential risks, data usage, and confidentiality measures. Participants explicitly consented to participation, data collection, data usage, and publication of research findings. All participants were assured that their anonymity would be maintained throughout the study and that their personal data would only be used for research purposes. This study employed a quasi-experimental research design, where participants were assigned to either an experimental or control group. The experimental group used an AI-powered mobile application (Liulishuo) as part of their English-learning process, while the control group followed traditional methods. This study did not involve any medical, psychological, or high-risk interventions. No vulnerable individuals, such as minors, patients, or refugees, were included in this study. No financial or material incentives were provided to participants. Participants were informed that their participation was voluntary, and they had the right to withdraw at any stage without any consequences. Participants were informed that there were no foreseeable risks associated with their participation in this study.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1057/s41599-025-04688-0>.

Correspondence and requests for materials should be addressed to Ma Mingyan, Nooreen Noordin or Abu Bakar Razali.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025