Impact of the Application of Mobile Integrated Speech Recognition and Automated Writing Evaluation Software on the University Learners' EFL Speaking Competence

Wenyi Li[1,*], Maslawati Mohamad[1], and Huay Woon You[2]

[1]Faculty of Education, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[2]Pusat PERMATA@Pintar Negara, Universiti Kebangsaan Malaysia, Selangor, Malaysia
Email: p121641@siswa.ukm.edu.my (W.L.); maslawati@ukm.edu.my (M.M.); hwyou@ukm.edu.my (H.W.Y.)
*Corresponding author
Manuscript received April 17, 2025; revised May 12, 2025; accepted June 5, 2025; published September 18, 2025

***Abstract*—Developing speaking competence remains a challenge for university learners in English as a Foreign Language contexts due to limited opportunities for oral practice and individualized feedback. This study examined the impact of a mobile-based approach that integrates automatic speech recognition and automated writing evaluation on the speaking competence of Chinese university learners. A quasi-experimental design was implemented with 104 first-year students enrolled in the compulsory English course. Participants were randomly assigned to an experimental group, which engaged in weekly speaking tasks supported by speech recognition and writing evaluation technologies, or a control group that followed traditional instruction. Both groups received the same syllabus and classroom teaching to control for instructional differences. Speaking performance was assessed through pre- and post-tests rated across four analytic dimensions: grammar and vocabulary, discourse management, pronunciation, and interactive communication. Results showed that the experimental group made significantly greater gains than the control group in all four dimensions, with the largest improvements observed in pronunciation and interactive communication, followed by grammatical accuracy and discourse organization. In addition, qualitative data from reflective journals, focus group interviews, and annotated screenshots indicated that learners perceived the integrated use of speech recognition and writing evaluation tools as effective for identifying language problems, promoting self-regulated practice, and increasing speaking confidence. These findings suggest that mobile-based automatic speech recognition and automated writing evaluation can provide effective, accessible, and scalable support for multidimensional oral development in university-level English language education.**

***Keywords*—speaking competence, automatic speech recognition, automated writing evaluation, mobile learning, oral English development**

I. Introduction

In English as a Foreign Language (EFL) contexts, speaking competence is widely recognized as one of the most challenging skills to develop, particularly among university-level learners in China [1]. Although recent curriculum reforms have placed greater emphasis on communicative ability, persistent issues such as limited fluency, grammatical inaccuracy, and low confidence in oral expression remain prevalent [2]. These difficulties are often attributed to a lack of speaking opportunities, minimal corrective feedback, and insufficient individualized support in traditional classroom settings [3]. While spoken language is increasingly valued in policy and assessment, it remains underemphasized in practice. This persistent disconnect between institutional goals and classroom realities underscores the need for pedagogical solutions that can support more autonomous and feedback-rich speaking practice.

The increasing availability of Mobile-Assisted Language Learning (MALL) tools has introduced new opportunities for enhancing oral language skills [4]. Among these, Automatic Speech Recognition (ASR) provides learners with real-time transcriptions of spoken language, facilitating pronunciation monitoring and fluency development [3]. Automated Writing Evaluation (AWE), originally designed for written feedback, offers grammar- and syntax-based suggestions that can be applied to transcribed speech [5]. Few empirical studies have examined the effects of integrating ASR and AWE into a structured learning cycle to support multidimensional speaking competence, encompassing knowledge of language and discourse, core speaking skills, and communication strategies [6].

Research on ASR in second language learning has primarily focused on its role in improving segmental and suprasegmental pronunciation features [7, 8]. Most ASR-related studies are conducted in fixed-content learning applications, where learners practice predefined dialogues or vocabulary, limiting opportunities for spontaneous and flexible speech production. AWE, in turn, has largely been examined in writing instruction, with limited attention to its applicability in oral language development. Few studies have explored how AWE feedback supports learners' grammatical revision of transcribed speech. Moreover, existing work tends to treat ASR and AWE as separate tools, overlooking their pedagogical potential when integrated into a single, feedback-rich practice cycle. Also lacking are investigations into how learners engage with and perceive this integration in mobile-supported, self-regulated learning environments.

To address these issues, the current study draws on the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2). This framework offers a multidimensional perspective for examining learners' perceptions of technology-enhanced speaking practice. Constructs such as performance expectancy, effort expectancy, and habit are particularly relevant in mobile learning environments and have been associated with learners' perceived gains in fluency, accuracy, and sustained oral engagement. However, few studies—especially within the Computer-Assisted Pronunciation Training (CAPT) literature—have connected these constructs to broader aspects of oral skill development. CAPT research has traditionally emphasized pronunciation

improvement through isolated and repetitive drills, often overlooking learners' perceptions and self-regulated engagement in real-world mobile contexts. These theoretical and practical gaps call for a more integrated, learner-centered approach.

In response to this gap, the present study implements a mobile-based instructional design that integrates ASR and AWE to support EFL learners' speaking development. The study evaluates the impact of this integrated approach and learners' perceptions of its impact on multiple dimensions of speaking competence.

The study is guided by the following research questions:

1) To what extent does the use of mobile-based ASR and AWE affect the speaking competence of Chinese university EFL learners?
2) How do learners perceive the impact of ASR- and AWE-supported speaking practice on different dimensions of their speaking competence?

Based on Research Question 1, the following hypotheses were formulated for statistical testing:

H1a: The Experimental Group (EG) will demonstrate a statistically significant improvement in speaking competence after the intervention.

H1b: The post-test speaking competence scores of the EG will be significantly higher than those of the Control Group (CG).

II. Literature Review

A. Speaking Competence in EFL Contexts

Speaking competence is widely recognized as a multidimensional construct that encompasses grammatical accuracy, lexical resource, pronunciation, fluency, and pragmatic appropriateness in real-time communication [9]. According to Goh and Burns, second language speaking involves three core components: knowledge of language and discourse, core speaking skills, and communication strategies [6]. This framework continues to inform recent research that seeks to identify pedagogical approaches capable of supporting integrated oral development [10].

In EFL university settings, especially in China, speaking remains a persistent challenge. Despite curricular reforms promoting communicative competence, students frequently exhibit hesitation, disfluency, and syntactic inaccuracy during spontaneous speech [11]. These limitations are commonly linked to teacher-fronted classroom practices that prioritize reading and grammar, while offering minimal opportunities for real-time spoken interaction. Furthermore, learners often lack access to individualized feedback on their oral output, which restricts their ability to notice errors and refine their speech production [12]. In this context, fostering speaking competence requires pedagogical designs that address both linguistic precision and interactive spontaneity, with sustained learner engagement and timely feedback as essential components.

B. MALL and Speaking Development

MALL has gained prominence as a flexible and learner-centered approach to second language acquisition, particularly in speaking instruction [13]. MALL enables learners to engage in speaking practice beyond classroom hours, offering personalized access to tasks, models, and feedback at their own pace [14]. In EFL contexts, where limited exposure to spoken English constrains fluency development, mobile technologies have been shown to increase speaking opportunities and enhance learners' motivation and autonomy [15].

Recent empirical studies have provided evidence of the positive impact of MALL on speaking performance. For example, a study by Al-Abri *et al.* found that mobile-based speaking activities significantly improved learners' fluency and lexical diversity [16]. Similarly, Wu *et al.* reported that sustained use of speaking apps helped students internalize sentence patterns and develop greater confidence in oral production. Meta-analytic reviews also indicate that MALL-supported instruction, particularly with structured speaking tasks and feedback, yields moderate to large effect sizes on various aspects of speaking competence [17].

However, much of the research in this area tends to prioritize learner motivation, usability, or general attitudes, with limited attention to how mobile tools support development across specific speaking subskills such as grammatical accuracy, discourse organization, or strategic interaction [18]. As such, further research is needed to clarify the pedagogical value of MALL for developing integrated speaking competence.

C. ASR in Language Learning

ASR has attracted increasing attention in second language acquisition due to its ability to convert speech into text in real-time, thereby offering learners immediate feedback on their spoken output [19]. Numerous studies have confirmed the benefits of ASR in supporting pronunciation development, especially in terms of segmental features (e.g., individual sounds) and suprasegmental features (e.g., intonation and stress). For instance, Ngo *et al.* found that ASR-assisted pronunciation training significantly improved learners' vowel articulation [7]. Similarly, Amrate and Tsai reported enhanced rhythm and intonation among EFL learners after using ASR-based practice tools that provided phonetic visualizations [8].

Beyond pronunciation, ASR-supported practice has also been associated with increased speech intelligibility and learner motivation. Jiang *et al.* demonstrated that the availability of real-time speech-to-text feedback helped learners detect mispronunciations and build confidence through self-paced repetition [20]. Such low-stakes practice environments reduce speaking anxiety and foster learner autonomy, particularly in EFL settings where access to corrective feedback is often limited [20].

Despite these promising outcomes, the use of ASR in language learning remains largely constrained to pronunciation-focused applications. Most mobile-based ASR tools are embedded within structured learning apps that offer pre-set vocabulary or sentence patterns (e.g., language learning games or scripted drills) [21]. These applications provide limited flexibility for learners to generate personalized spoken content or engage in open-ended tasks. As a result, the pedagogical value of ASR has often been confined to the repetition of fixed phrases, with less emphasis on higher-level speaking development such as syntactic complexity, grammatical accuracy, or discourse

organization [20].

Moreover, the potential of dictation-style ASR applications—where learners use the technology to generate and revise spontaneous spoken texts—has received limited scholarly attention [20]. Unlike controlled input systems, dictation-based ASR can support freer speaking practice by enabling learners to monitor, transcribe, and reflect on their language output [20]. However, this functionality is rarely explored in current EFL speaking instruction research, leaving a gap in understanding how ASR can be utilized beyond phonological training.

The present study addresses this gap by integrating ASR into a mobile learning cycle that supports personalized speaking practice. Rather than relying on fixed content, learners in this study use ASR to generate open-ended spoken responses, which are then revised with the help of automated grammar feedback. This approach enables the design to utilize the transcription function of ASR not only for pronunciation support, but also for the development of broader speaking competence.

D. AWE and Its Potential for Speaking

AWE systems—such as Grammarly, Criterion, and Pigai—have been widely adopted in language education to provide real-time, individualized feedback on grammar, vocabulary usage, and sentence structure. These systems are primarily designed for writing instruction, offering learners opportunities to identify linguistic errors and make targeted revisions based on system-generated feedback [22].

In EFL contexts, AWE systems have demonstrated considerable effectiveness in improving learners' writing performance. Recent studies have shown that AWE tools not only enhance grammatical accuracy but also support the development of lexical range and syntactic complexity by enabling learners to revise their texts based on instant, individualized feedback [23]. These tools have been particularly effective in promoting self-directed learning and metalinguistic awareness, as learners engage with feedback iteratively and develop the ability to identify and correct recurring errors [24]. Moreover, AWE has been recognized as a scalable solution in large-class settings, where teacher feedback may be limited, offering timely and consistent linguistic support to learners across proficiency levels [25].

Although AWE has been extensively applied in writing instruction and shown to improve writing accuracy, revision practices, and learner autonomy [26], its application in oral language learning contexts, particularly mobile-supported speaking tasks, has not yet been systematically examined in the literature. This study addresses that gap by examining how the combined use of ASR and AWE within a mobile-based environment supports the development of multiple dimensions of EFL learners' speaking competence.

E. UTAUT2 and Speaking Development

UTAUT2 provides a multidimensional framework for understanding learners' behavioral engagement with educational technologies. Although originally developed in consumer behavior research, UTAUT2 has been increasingly adopted to investigate MALL, including contexts involving oral skill development. While relatively few studies directly apply UTAUT2 to second language speaking, emerging research suggests that its core constructs are relevant to learner behavior and performance in technology-mediated speaking practice.

Performance Expectancy (PE) refers to the perceived effectiveness of a technology in improving learning outcomes. In the context of speaking development, Guskaroska [27] found that EFL learners' perceived gains in pronunciation accuracy and oral fluency through ASR tools contributed strongly to their intention to continue using such applications. Similarly, Nguyen [28] emphasized that learners' belief in the benefits of mobile tools for language production positively influenced their engagement in self-directed speaking tasks.

Effort Expectancy (EE) denotes the perceived ease of using the system. Hirschi *et al.* [29] reported that adult ESOL learners were more likely to engage in mobile-assisted pronunciation training when the ASR interface was simple, responsive, and required minimal instructional support. Tools that were intuitive and required little cognitive overhead were associated with higher rates of spontaneous speaking practice.

Social Influence (SI) involves the extent to which learners perceive that significant others, such as teachers or peers, support their use of a given technology. In Nguyen's [28] study, peer and teacher encouragement played a critical role in shaping learners' attitudes toward mobile speaking tools, especially in collaborative or feedback-oriented settings.

Facilitating Conditions (FC) refer to the availability of institutional, technical, or instructional support that enables effective tool usage. Nguyen [28] also highlighted that stable access to devices, application reliability, and scaffolding from instructors were crucial for maintaining sustained learner engagement, especially during autonomous speaking activities.

Hedonic Motivation (HM) represents the degree of enjoyment derived from using the technology. Shu, Huang, and Xing [30] found that multimodal human–computer interaction, such as visual and voice-based prompts, created a more emotionally engaging environment, particularly in oral communication activities. Learners reported reduced anxiety and increased motivation when the tools offered interactive or entertaining elements.

Price Value (PV), though less frequently studied in educational contexts, has been linked to learners' preference for freely accessible tools. Khan *et al.* [31] observed that undergraduate learners were more likely to adopt mobile speaking platforms when no additional costs were involved, even if premium versions offered enhanced features. Cost-related concerns influenced both initial adoption and long-term retention.

Habit (HT) reflects the extent to which learners use a technology automatically as a result of repeated prior experience. In speaking development, habit formation is essential for building fluency and confidence. Khan *et al.* [31] reported that students who incorporated mobile tools into their daily language routines demonstrated greater autonomy and more consistent oral practice over time.

Overall, UTAUT2 offers a useful lens for examining not only learner acceptance of mobile technologies but also the behavioral and affective pathways through which these tools support speaking competence. In this study, UTAUT2 is employed to interpret learners' perceptions of ASR–AWE

integration and its relevance to oral skill development in a mobile-supported learning environment.

III. Materials and Methods

A. Research Design

This study employed a mixed-methods quasi-experimental design to evaluate the impact of integrating ASR and AWE on Chinese university EFL learners' speaking competence. The rationale for this design lies in its ability to capture both quantitative gains in speaking performance and qualitative insights into learners' experiences with mobile-supported speaking practice.

Two groups were involved: an EG, which received speaking instruction supported by ASR- and AWE-based mobile practice, and a CG, which followed regular speaking activities without technological assistance. Pretest and posttest speaking assessments were conducted for both groups to measure development in speaking competence. Meanwhile, three qualitative sources—reflective journals, focus group interviews, and annotated screenshots—were used to triangulate findings and enrich the interpretation of learners' engagement and responses throughout the intervention.

This research design aligns with Creswell's recommendations for mixed-methods experimental studies, particularly in language education contexts where learner experience and performance outcomes must be interpreted together to inform practice [32].

B. Participants

A total of 104 first-year Business English major students participated in this study. All were enrolled in the compulsory College English course at a Chinese university during the semester when the intervention took place. The course is part of the national English curriculum for undergraduate students, designed to improve general language proficiency in listening, speaking, reading, and writing. The original course did not include any mobile-supported speaking component. The mobile-based ASR and AWE intervention was introduced exclusively for EG.

Participants were selected through a pre-study background survey. Inclusion criteria included: (1) prior exposure to MALL, (2) self-reported moderate to intermediate speaking proficiency based on the Prescription of English Speaking Performance Levels in China's Standards of English Language Ability [33]. This classification corresponds to Level 5 of the national standard, which describes learners who can communicate in routine academic and social situations, express opinions on familiar topics, and maintain basic coherence and grammatical control in speaking. This level is approximately equivalent to B1 on the CEFR scale and reflects an intermediate level of English-speaking proficiency [34]. This benchmark ensured baseline comparability between groups and supports the interpretation of observed learning gains. Based on these criteria, participants were randomly assigned to two groups: 52 in EG and 52 in CG.

All participants provided written informed consent and were assured that participation was voluntary, anonymous, and would not affect their academic evaluation in the course. Ethical approval for the study was obtained from the university's research ethics committee.

C. Intervention Procedure

The intervention was conducted over 14 weeks, aligned with the academic calendar of the College English course. Both EG and CG followed the same syllabus, classroom instruction, and weekly speaking assignments, with the only difference being the integration of ASR and AWE technologies for the EG.

The College English course allocated approximately one hour per week for speaking activities, including classroom-based discussions and oral assignments. To avoid confounding effects due to differences in time on task, EG was assigned one additional hour per week for ASR–AWE-supported practice. Speaking assignments were matched in topic, complexity, and timing for both groups, with the only difference being the mode of completion. To standardize engagement and control for excessive or uneven practice beyond assigned tasks, participants in EG were required to submit weekly annotated screenshots and reflective journals as verification of task completion. These screenshots were used to verify adherence to time requirements and ensure that only the scheduled one-hour ASR–AWE activity per week was considered in the analysis.

To implement the intervention, NetEase Youdao Dictionary was selected for EG. This app was chosen based on the following considerations:

- It is freely accessible across major mobile platforms, enabling learners to practice anytime and anywhere without financial or technical barriers.
- It integrates both ASR and AWE functionalities, allowing users to receive real-time transcriptions of their speech alongside automated feedback on grammar, sentence structure, and word choice.
- It is one of China's most widely used language learning apps, with a large user base and high download volume, indicating its functional stability and user acceptance.

Students in the EG engaged in a structured, iterative speaking practice cycle that combined ASR and AWE functionalities. The procedure is illustrated in Fig. 1, which presents a decision-based workflow guiding learners through repeated cycles of recording, feedback, revision, and re-recording.

The cycle began with students recording their spoken responses using the app's ASR feature. The system instantly produced a transcript of the spoken content. Students then evaluated whether the ASR had accurately captured their intended utterances. If any errors or omissions were identified, they re-recorded the task until accurate recognition was achieved, thereby engaging in targeted pronunciation adjustment.

Once a satisfactory transcript was generated, students copied the ASR-converted text into the AWE interface to receive feedback on grammar, vocabulary, and sentence structure. Based on this feedback, they revised the spoken content to meet grammatical and syntactic norms. If revisions were made, students returned to the ASR step to re-record the improved version. This recursive, self-regulated loop supported both pronunciation improvement and language form control.

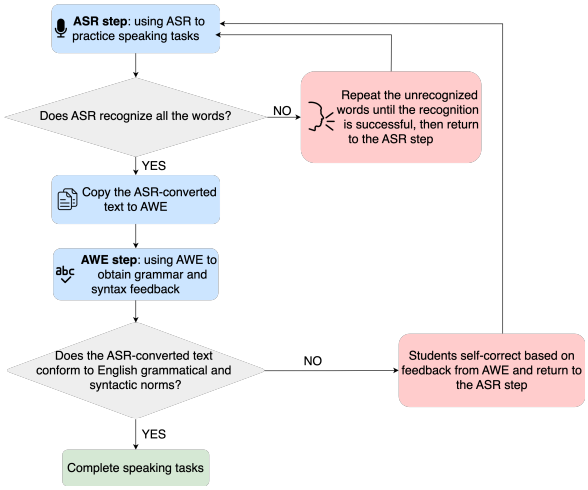

Fig. 1. Workflow of the ASR–AWE integrated speaking practice cycle.

The cycle concluded only when the spoken output met both phonological and grammatical criteria. Students then submitted their final recordings along with annotated screenshots showing the ASR and AWE feedback, forming their weekly speaking portfolios. In addition, reflective journals were collected for qualitative analysis.

By contrast, students in the CG completed the same speaking tasks through conventional methods such as in-class presentations and pair discussions without any use of ASR or AWE. To ensure instructional consistency, both groups were taught by the same instructor. While general classroom feedback was provided to all students, the instructor did not intervene in the technology-based practice process for the EG. The overall design and timeline of the intervention are shown in Fig. 2.

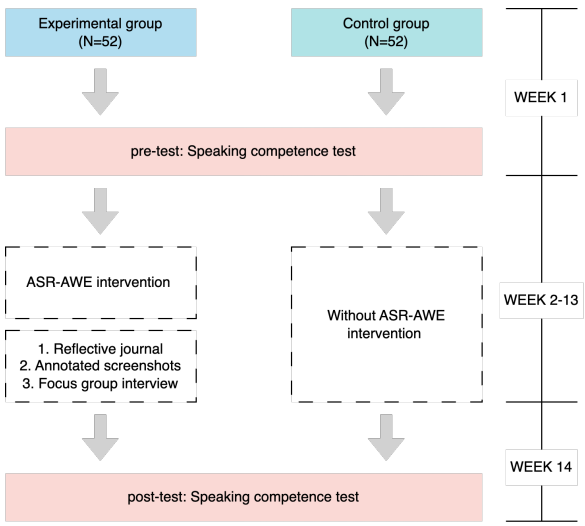

Fig. 2. Experimental procedure and data collection timeline.

D. Instruments

1) Speaking test

To evaluate students' speaking competence before and after the intervention, a speaking test was administered to both groups. The test tasks were adapted from the official Cambridge English: Business Vantage (BEC Vantage) Speaking Test, which is designed to assess communicative performance in business-related English contexts at the B2 level of the Common European Framework of Reference for Languages.

The speaking test followed the standard format of the BEC Vantage Speaking component, which includes three parts: a personal interview (Part 1), a collaborative task (Part 2), and a discussion based on visual input (Part 3). Each test session lasted approximately 14 minutes and was conducted in pairs to simulate authentic workplace interaction.

Students' performances were audio-recorded and independently rated by two certified BEC Vantage Speaking Examiners, both of whom had over 15 years of experience in English oral assessment. The evaluation employed the B2 Business Vantage Speaking Assessment Scale, which consists of four analytic categories:

1) Grammar and Vocabulary: accuracy and range of grammatical forms and vocabulary.
2) Discourse Management: ability to produce extended and organized speech using appropriate cohesive devices.
3) Pronunciation: intelligibility, control of intonation, sentence stress, and individual sounds.
4) Interactive Communication: ability to initiate, respond, link contributions, and maintain the flow of interaction.

Each dimension was scored on a scale from 0 to 5, with band descriptors outlining specific performance features. The official assessment scale is provided in Table A1. To ensure inter-rater reliability, intraclass correlation coefficients were calculated to assess the level of agreement between the two examiners.

2) Reflective journals

To explore learners' perceptions and experiences with mobile-supported speaking practice, EG was asked to submit weekly reflective journals throughout the 14-week intervention. The design of the journal prompts was based on UTAUT2 proposed by Venkatesh *et al.*, which includes the following seven dimensions [35]:

1) PE: learners' beliefs about how ASR and AWE help improve their speaking competence.
2) EE: learners' perceptions of ease or difficulty in using ASR and AWE.
3) SI: the extent to which learners felt encouraged or discouraged by peers, family, or instructors.
4) FC: the availability of resources, time, and technical support during using ASR and AWE.
5) HM: learners' enjoyment, interest, or emotional responses to the use of ASR and AWE.
6) PV: learners' evaluation of whether the benefits of using ASR and AWE outweigh the monetary cost (not applicable in this study, as the ASR and AWE features in NetEase Youdao Dictionary are freely accessible).
7) HT: whether and how learners began to develop routines or habitual use patterns related to ASR and AWE.

Each reflective journal included open-ended guiding questions corresponding to UTAUT2 dimensions, allowing learners to reflect on specific aspects of their technology use (see Table A2). Reflective journals were written in Chinese to ensure depth and authenticity of expression and were later translated and coded using thematic analysis.

3) Focus group interviews

At the end of the intervention, semi-structured focus group interviews were conducted with EG. Each group consisted of 6–7 participants.

The interview protocol included open-ended questions

based on the UTAUT2 dimensions and learners' experiences with ASR and AWE (see Table A3). Interviews were conducted in Chinese to ensure comfort and depth of response and were audio-recorded and transcribed for analysis.

4) Annotated screenshots

To complement self-reported data, students in EG were also asked to submit annotated screenshots documenting their weekly use of ASR and AWE. These screenshots captured moments of system feedback, personal revision, and reflection.

Students used color-coding and brief comments to indicate system-generated corrections and their own responses. These screenshots served as visual artifacts that triangulated with reflective journals and interview data during qualitative coding.

E. Data Collection and Analysis

1) Quantitative analysis

All statistical analyses were conducted using SPSS 29. Descriptive statistics were calculated to summarize the central tendency and variability of speaking scores across conditions. The Shapiro–Wilk test was first used to examine the normality assumption, which determined the choice of subsequent inferential methods. If normality was met, independent and paired samples t-tests were conducted to assess between-group and within-group differences, respectively. In cases where normality was violated, non-parametric alternatives (i.e., Mann–Whitney U and Wilcoxon signed-rank tests) were employed. Effect sizes were reported using Cohen's d for t-tests and r for non-parametric tests.

2) Qualitative analysis

Qualitative data were collected from three sources: (1) reflective journals, (2) focus group interviews, and (3) annotated screenshots from students in EG.

All qualitative materials were transcribed, translated (where necessary), and analyzed using thematic analysis [36]. The coding process followed a three-level procedure:

1) First-order codes: meaningful phrases or sentences from participants' responses
2) Second-order categories: conceptual themes grouped under UTAUT2 dimensions
3) Aggregate dimensions: aligned with the six constructs of UTAUT2 (PE, EE, SI, FC, HM, HT)

To ensure trustworthiness, investigator triangulation was applied, where two researchers independently coded a subset of data and reached consensus through discussion. Representative quotes were selected to illustrate key findings.

F. Ethical Considerations

This study was conducted by ethical standards for research involving human participants. Before data collection, ethical approval was obtained from the Research Ethics Committee of a university in China. All procedures, including recruitment, consent, and data handling, were designed to protect participants' rights, privacy, and well-being.

All participants were informed of the research purpose, procedures, and their right to withdraw at any time without penalty. Written informed consent was obtained from each participant before the study began. Students were assured that their participation was voluntary and would not affect their course grades, attendance, or standing within the class.

To ensure confidentiality, all personal identifiers were removed during data processing. Each participant was assigned a unique code in transcripts, analyses, and reporting. All digital and physical data were securely stored in encrypted folders accessible only to the research team and will be retained for a limited period before being permanently deleted.

IV. Result and Discussion

A. Overview of the Section

This section presents the results of the study, organized according to the two research questions. Quantitative data are analyzed to evaluate the impact of the ASR–AWE intervention on learners' speaking competence (RQ1), with hypotheses H1a and H1b tested through descriptive and inferential statistics. Qualitative findings are then presented to explore learners' perceptions of the intervention (RQ2), thematically categorized under the UTAUT2 framework.

B. Results for Research Question 1: Impact of ASR-AWE on Speaking Competence

1) Descriptive analysis of speaking competence

a) Overall score summary

Table 1 presents the descriptive statistics of speaking test overall scores for EG and CG at the pre-test and post-test stages. The reported indicators include mean, median, mode, standard deviation, and variance.

Table 1. Descriptive statistics of speaking test overall Scores

Group	Time	Mean	Median	Mode	SD	Variance
EG	Pre-test	12.28	12.33	13.00	0.91	0.82
	Post-test	14.01	14.00	14.00	0.92	0.85
CG	Pre-test	12.19	12.33	13.00	1.02	1.03
	Post-test	12.98	13.00	13.00	0.92	0.85

As shown in the Table 1, both groups demonstrated improvement in their post-test scores compared to the pre-test; however, the degree of improvement differed markedly. EG's mean score increased by 1.73 points (from 12.28 to 14.01), while CG showed a smaller increase of 0.79 points (from 12.19 to 12.98). This notable difference suggests that the integrated use of mobile ASR and AWE may have contributed meaningfully to the development of students' speaking competence.

In addition to higher mean scores, EG's median and mode reached 14.00 in the post-test, indicating a strong and consistent upward shift across participants. The relatively stable standard deviation and variance values imply that the observed improvement was not limited to a few outliers but rather reflected a more general enhancement within the group.

These initial descriptive results provide preliminary support for the effectiveness of the ASR–AWE integrated intervention in enhancing EFL university students' speaking competence. Further inferential statistical analyses will be conducted to examine the significance of these observed differences.

b) Dimension score overview

To provide a more detailed view of learners' performance across different speaking dimensions, Table 2 reports the descriptive statistics of four core dimensions—grammar and vocabulary, discourse management, pronunciation, and interactive communication—for both groups at pre- and post-test stages. Figs. 3 and 4 visually illustrate the changes in dimension scores over time, allowing for easier comparison of developmental trends.

Table 2. Descriptive statistics of speaking test dimension scores

Group	Time	Dimension	Mean	SD	Variance
EG	Pre-test	D1	3.24	0.04	0.09
		D2	3.23	0.04	0.10
		D3	2.83	0.05	0.15
		D4	2.97	0.05	0.14
	Post-test	D1	3.49	0.32	0.10
		D2	3.41	0.37	0.13
		D3	3.60	0.37	0.14
		D4	3.51	0.42	0.17
CG	Pre-test	D1	3.22	0.48	0.23
		D2	3.10	0.17	0.03
		D3	2.76	0.42	0.18
		D4	3.11	0.25	0.06
	Post-test	D1	3.37	0.38	0.14
		D2	3.25	0.41	0.17
		D3	3.15	0.39	0.16
		D4	3.20	0.38	0.14

Notes: D1: grammar and vocabulary; D2: discourse management; D3: pronunciation; D4: interactive communication

As shown in Fig. 3, EG experienced consistent and marked improvement across all four dimensions after the ASR–AWE intervention. The most notable gains were pronunciation, with a mean increase from 2.83 to 3.60, followed by interactive communication (2.97 to 3.51). These improvements suggest that the ASR-AWE's dual feedback mechanism facilitated linguistic precision and communicative fluency. Grammar and discourse-related scores also rose steadily, indicating balanced growth across knowledge-based and performance-based speaking components.

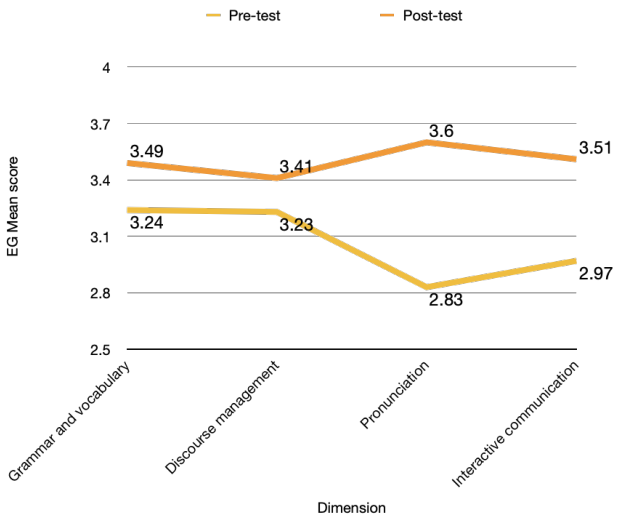

Fig. 3. Mean score trends across speaking test dimension scores in EG.

In contrast, Fig. 4. shows that CG also made some gains across dimensions, but the changes were smaller and more variable. For instance, pronunciation only increased from 2.76 to 3.15, and interactive communication from 3.11 to 3.20. The CG's dimension scores remained consistently lower than those of the EG at the post-test stage.

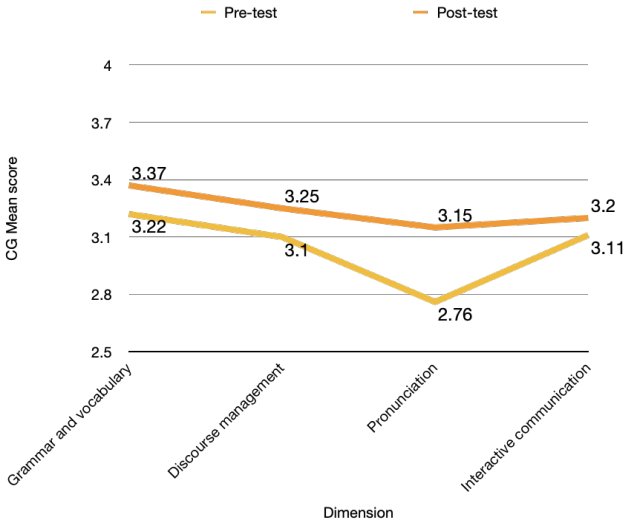

Fig. 4. Mean score trends across speaking test dimension scores in CG.

Beyond mean scores, the Standard Deviation (SD) and variance figures in Table 2 offer further insight into the consistency of learner performance. In the EG, post-test SDs and variances across dimensions remained relatively low and stable (e.g., SD = 0.37 for pronunciation, variance = 0.14), suggesting that the observed improvements were not driven by a few high-performing individuals, but reflected broad-based gains across the group. This is particularly evident in discourse management and pronunciation, where increased means were accompanied by moderate and consistent variability.

In contrast, the CG's variability values were either unchanged or slightly higher, particularly in the pronunciation and discourse management dimensions. For example, the variance for CG's pronunciation dimension at post-test was 0.16—comparable to EG—but with a much smaller gain in mean score. This may indicate more scattered individual outcomes within CG, possibly due to lack of structured feedback mechanisms.

Overall, these findings reinforce the effectiveness of the ASR–AWE not only in improving speaking competence across multiple dimensions, but also in fostering coherent and stable progress among learners. The convergence of improved mean scores with steady variability measures provides stronger support for the robustness of the intervention's impact.

2) Hypothesis testing: within-group changes (H1a)

a) Normality test

Before performing inferential statistical analyses, normality tests were conducted to examine whether the speaking test scores met the assumptions required for parametric tests. The Shapiro–Wilk test was employed due to its higher power for small to moderate sample sizes. The results of the normality tests for both the overall scores and the four dimension scores are summarized in Tables 3 and 4, respectively.

As shown in Table 3, the pre-test scores of EG ($p = 0.006$) and CG ($p = 0.025$) significantly deviated from normality. In contrast, the post-test scores for both groups met the normality assumption ($p > 0.05$), suggesting acceptable distribution at the later testing stage. Given that two out of

four datasets violated the assumption, non-parametric tests—specifically, the Wilcoxon Signed-Rank Test (for within-group comparisons) and the Mann–Whitney U Test (for between-group comparisons)—were selected for subsequent analyses of total speaking scores.

Table 3. Shapiro–wilk test for speaking test overall scores

Group	Time	W Statistic	Sig. (*p*)
EG	Pre-test	0.933	0.006
	Post-test	0.960	0.078
CG	Pre-test	0.948	0.025
	Post-test	0.965	0.134

Table 4. Shapiro–wilk test for speaking test dimension scores

Group	Time	Dimension	W Statistic	Sig. (*p*)
EG	Pre-test	D1	0.76	< 0.001
	Post-test		0.88	< 0.001
	Pre-test	D2	0.88	< 0.001
	Post-test		0.85	< 0.001
	Pre-test	D3	0.85	< 0.001
	Post-test		0.82	< 0.001
	Pre-test	D4	0.87	< 0.001
	Post-test		0.88	< 0.001
CG	Pre-test	D1	0.93	0.005
	Post-test		0.81	< 0.001
	Pre-test	D2	0.67	< 0.001
	Post-test		0.68	< 0.001
	Pre-test	D3	0.87	< 0.001
	Post-test		0.83	< 0.001
	Pre-test	D4	0.73	< 0.001
	Post-test		0.80	< 0.001

Notes: D1: grammar and vocabulary; D2: discourse management; D3: pronunciation; D4: interactive communication

To determine whether parametric tests were suitable for analyzing the four speaking dimensions, the Shapiro–Wilk test was also applied to each dimension across both groups and time points. The results in Table 4 indicate that all 16 datasets violated the normality assumption ($p < 0.001$ or $p < 0.01$), with W-statistics ranging from 0.67 to 0.93. This pattern was consistent across all dimensions, including grammar and vocabulary, discourse management, pronunciation, and interactive communication. Notably, the discourse management scores exhibited particularly low W-values (e.g., .67 for CG pre-test), indicating substantial deviation from a normal distribution.

Given these findings, non-parametric tests were deemed the most appropriate analytical choice for all inferential comparisons involving the four dimensions. This decision ensured methodological consistency and protected against potential violations of statistical assumptions that could compromise the validity of the results.

b) Wilcoxon signed-rank test

To examine the effectiveness of the ASR–AWE intervention within each group, Wilcoxon signed-rank tests were conducted to compare pre-test and post-test scores. As shown in Table 5, EG and CG demonstrated statistically significant improvements in overall speaking performance.

Table 5. Wilcoxon signed-rank test for speaking test overall scores

Group	Ranks summary	Z	p	r
EG	+52/-0/=0	−6.285	< .001	0.87
CG	+39/-5/=8	−5.503	< .001	0.76

The EG exhibited a complete shift toward higher scores, with all 52 participants showing positive ranks ($Z = -6.285$, $p < 0.001$), resulting in a large effect size ($r = 0.87$). This indicates a strong and consistent improvement across EG. The CG also showed a significant increase ($Z = -5.503$, $p < 0.001$), though five participants experienced lower post-test scores and eight had tied scores. The effect size for the CG ($r = 0.76$) was slightly lower, suggesting that while the CG improved, the magnitude of change was less substantial than in the EG.

To explore which specific aspects of speaking competence improved over time, Wilcoxon signed-rank tests were also applied to each of the four assessed dimensions: grammar and vocabulary, discourse management, pronunciation, and interactive communication. The results are summarized in Table 6. To better visualize the relative impact of the intervention across speaking dimensions in EG, a radar chart was used to depict the effect sizes (r) for each assessed component (see Figs. 5 and 6).

Table 6. Wilcoxon signed-rank test for speaking test dimension scores

Group	Dimension	Rank summary	Z	*p*	r
EG	D1	+30/−8/=14	−3.558	< 0.001	0.49
	D2	+25/−10/=17	−3.039	0.002	0.42
	D3	+48/−0/=4	−6.083	< 0.001	0.83
	D4	+40/−2/=10	−5.507	< 0.001	0.77
CG	D1	+29/−14/=9	−2.038	0.042	0.28
	D2	+15/−10/=27	−2.541	0.011	0.35
	D3	+31/−4/+17	−4.793	< 0.001	0.67
	D4	+13/−6/=33	−2.001	0.045	0.28

Notes: D1: grammar and vocabulary; D2: discourse management; D3: pronunciation; D4: interactive communication

In EG, all four dimensions showed statistically significant improvements from pre- to post-test: Pronunciation ($Z = -6.083$, $p < 0.001$, $r = 0.83$) and interactive communication ($Z = -5.507$, $p < 0.001$, $r = 0.77$) exhibited the largest effect sizes, suggesting that these were the most substantially impacted areas. Grammar and vocabulary ($Z = -3.558$, $p < 0.001$, $r = 0.49$) and discourse management ($Z = -3.039$, $p = 0.002$, $r = 0.42$) also showed moderate to large effects, indicating well-rounded growth across linguistic and communicative domains.

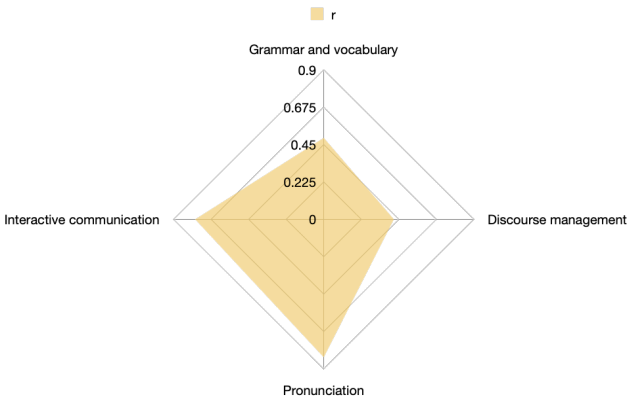

Fig. 5. Effect sizes of speaking test dimension scores in EG.

In CG, all four dimensions also reached statistical significance, but with smaller effect sizes overall: The largest improvement was observed in pronunciation ($Z = -4.793$, $p < 0.001$, $r = 0.67$), followed by discourse management ($Z = -2.541$, $p = 0.011$, $r = 0.35$). Grammar and vocabulary ($Z = -2.038$, $p = 0.042$, $r = 0.28$) and interactive communication ($Z = -2.001$, $p = 0.045$, $r = 0.28$) demonstrated smaller effect sizes, suggesting more modest changes.

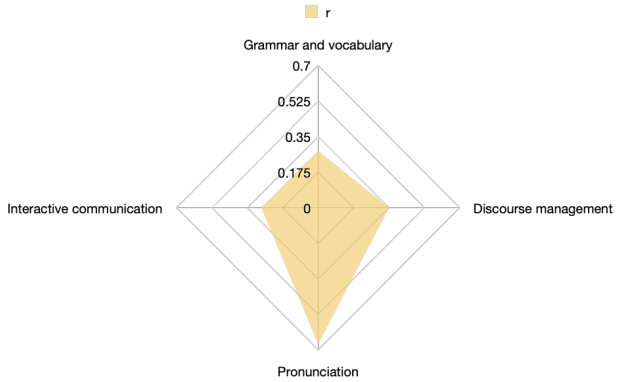

Fig. 6. Effect sizes of speaking test dimension scores in CG.

Taken together, the within-group comparisons confirm that while both groups benefitted over time, EG experienced more substantial and consistent gains, particularly in pronunciation and interactional fluency. These results provide strong evidence of the positive impact of the ASR–AWE intervention on multiple dimensions of speaking competence.

3) Hypothesis testing: between-group changes (H1b)

a) Mann-whitney U test

To further evaluate the differential impact of the ASR–AWE intervention, Mann–Whitney U tests were conducted to compare the performance of EG and CG at both the pre-test and post-test stages. The results for the overall speaking scores are presented in Table 7.

Table 7. Mann–whitney U test for speaking test overall scores

Time	U	Z	p	r
Pre-test	1307.000	−0.297	0.767	0.03
Post-test	589.000	−4.998	< 0.001	0.49

As shown in Table 7, the pre-test comparison between the EG and CG revealed no statistically significant difference (U = 1307.000, Z = –0.297, p = 0.767), indicating that the two groups were comparable in speaking proficiency at baseline. However, at the post-test stage, the difference became statistically significant (U = 589.000, Z = –4.998, p < 0.001), with a moderate effect size (r = 0.49). This suggests that EG outperformed CG following the intervention, supporting the efficacy of the integrated ASR–AWE approach.

Mann–Whitney U tests were also performed for each of the four speaking dimensions. Table 8 shows the results for both the pre- and post-test stages.

Table 8. Mann–whitney U test for speaking test dimension scores

Dimension	Time	U	Z	*p*	r
D1	Pre-test	1276.5	−0.512	0.609	0.05
	Post-test	1089.0	−1.774	0.076	0.17
D2	Pre-test	1047.0	−2.205	0.027	0.22
	Post-test	995.0	−2.491	0.013	0.24
D3	Pre-test	1237.0	−0.788	0.431	0.08
	Post-test	579.5	−5.225	< 0.001	0.51
D4	Pre-test	1080.0	−1.987	0.047	0.20
	Post-test	774.5	−3.900	< 0.001	0.38

Notes: D1: grammar and vocabulary; D2: discourse management; D3: pronunciation; D4: interactive communication.

At the pre-test stage, none of the dimensions showed significant differences between the groups, except for discourse management (p = 0.027, r = 0.22) and interactive communication (p = 0.047, r = 0.20), which showed small effect sizes. This suggests slight baseline variation, though the differences were relatively minor.

At the post-test stage, statistically significant differences emerged in three out of the four dimensions:

Pronunciation: Z = –5.225, p < 0.001, r = 0.51 (large effect). Interactive communication: Z = –3.900, p < 0.001, r = 0.38 (moderate effect). Discourse management: Z = –2.491, p = 0.013, r = 0.24 (small to moderate effect).

While grammar and vocabulary did not reach statistical significance (p = 0.076), the effect size (r = 0.17) indicates a small trend in favor of EG.

These findings suggest that EG made significantly greater improvements than CG in several core dimensions of speaking competence, particularly in pronunciation and interactive communication.

C. Results for Research Question 2: Learners' Perceptions of ASR-AWE-supported Speaking Practice

1) Overview of thematic analysis approach

To explore students' perceptions of the intervention, a thematic analysis was conducted on qualitative data collected from reflective diaries, focus group interviews, and open-ended questionnaire responses. The analysis followed Braun and Clarke's six-phase framework, involving familiarization with the data, initial coding, theme development, theme refinement, definition and naming, and final reporting [36].

A theory-driven coding approach was adopted, using the six dimensions of UTAUT2 as the analytical framework: PE, EE, SI, FC, HM and HT. Based on this structure, the analysis aimed to examine how learners perceived the use of the ASR–AWE integrated approach in enhancing their speaking competence. Table 9 presents the multiple sub-themes and sub-sub-themes emerging under each theme. A detailed breakdown of sub-themes and additional illustrative quotes across all UTAUT2 themes is provided in Table A4.

2) Theme 1: Performance expectancy

Performance expectancy emerges as the most prominently reported theme in participants' reflections. Participants widely perceived that the integration of mobile ASR and AWE contributed to meaningful improvements in their speaking competence. These perceived gains extended across linguistic accuracy, phonological awareness, discourse organization, and strategic self-monitoring during oral production.

Three key sub-themes were identified under this theme. The first sub-theme, enhancing the language and discourse knowledge, encompassed learners' development in grammar, vocabulary, pronunciation, and discourse organization. Many reported that the feedback loop helped them identify persistent language problems, such as tense misuse, pronunciation errors, and unclear expression. Some students also reported learning new lexical items and improving the structural clarity of their extended responses. As one participant explained, *"The system showed me how I kept using incorrect tenses when speaking."* (P07).

The second sub-theme, enhancing the core speaking skills, included improvements in articulation, speech function, and coherence. Participants described how repeating sentences until accurate ASR transcription was achieved led to clearer

pronunciation. They also noted increased fluency and pragmatic appropriateness when expressing requests or opinions. In several cases, students reported using feedback to restructure responses for more effective openings and closings.

Table 9. Themes and sub-themes based on UTAUT2 framework

Themes	Sub-themes	Sub sub-themes
1. Performance expectancy	Enhancing the language and discourse knowledge	Grammatical knowledge Phonological knowledge Lexical knowledge Discourse knowledge
	Enhancing the core speaking skills	Pronunciation skills Speech function Discourse organization
	Enhancing the communication strategies	Cognitive strategies Metacognitive strategies Discourse planning and monitoring
2. Effort expectancy	Perceived ease	Easy to use Convenience and flexibility
	User interface	Comfortable with ASR's pronunciation Comfortable with the user interface
3. Social Influence	Peer influence Academic influence	Support from peer Support from instructor
4. Facilitating conditions	Resource support	Authoritative dictionaries Exchange learning strategy
	Operating conditions	Lack of offline grammar correction
5. Hedonic motivation	Enjoyment	Vocabulary learning through song lyrics Vocabulary learning through short video
6. Habit	Dependency	Dependency on ASR-AWE for improving speaking competence Frequent use of ASR-AWE
	Habitual usage	Routine use of ASR-AWE in language practice

The third sub-theme, enhancing the communication strategies, reflected participants' increased use of cognitive, metacognitive, and discourse-monitoring strategies. Students described how they began mentally planning their speech, monitoring recurring issues through transcripts, and consciously revising content to improve logical flow and connectedness. These reflections suggest that learners were not merely receiving feedback passively but were actively shaping their oral output through self-regulated cycles of production and revision.

3) Theme 2: Effort expectancy

Effort expectancy captures learners' perceptions of how easy and convenient it was to use the mobile ASR and AWE during speaking practice. Participants generally described the practice as accessible, intuitive, and suitable for self-directed learning.

Two key sub-themes emerged: perceived ease and interface experience. Learners consistently found the process of recording, reviewing, and revising to require minimal technical effort, which encouraged spontaneous and frequent use. Many also emphasized the flexibility of practice, noting that it could easily fit into short breaks or fragmented schedules. In addition, participants appreciated the user interface, highlighting the clarity of pronunciation models and the smooth navigation layout, which contributed to a more focused and confident learning experience.

As one participant explained, *"I just opened the app, clicked record, and started speaking. It was so simple."* (P11) This comment exemplifies how the low operational demands of the integrated tools enabled learners to direct their attention toward meaningful language production rather than technical challenges.

4) Theme 3: Social influence

Social influence, while not as frequently mentioned as other themes, plays a meaningful role in shaping learners' engagement with mobile ASR and AWE. Participants' reflections revealed that encouragement and endorsement from others often reinforced their motivation to try and persist with the tools.

Two sub-themes emerged under this category: peer influence and academic influence.

Under peer influence, learners commonly cited the impact of classmates or roommates on their decision to adopt certain learning strategies. Observing others using the feedback functions encouraged students to emulate these behaviors, particularly when it came to refining spoken grammar and structure. As one participant shared, *"My roommate told me she always used the grammar suggestions to polish her answers."* (P29) This kind of informal peer modeling created a sense of shared learning norms and increased learner initiative.

Academic influence stemmed from teacher recommendations and validation. Participants noted that when instructors explicitly endorsed the use of the tools for independent error correction, it boosted their confidence and normalized regular usage. As P35 explained, *"Our teacher said this tool can really help us correct ourselves, so I used it weekly."* Instructor support helped align students' autonomous use of the technology with broader instructional goals, reinforcing both acceptance and sustained engagement.

Together, these reflections indicate that social influence, particularly through peer modeling and teacher validation, served as both a motivational catalyst and a behavioral anchor in learners' use of ASR and AWE.

5) Theme 4: Facilitating condition

Facilitating condition refers to the external resources and environmental factors that influence learners' engagement with mobile ASR and AWE. Participants reported both

enabling supports and practical constraints that shaped their experience with the technology.

Two sub-themes were identified: resource support and operating conditions.

Under resource support, learners valued features that enhanced their individual or collaborative learning. The built-in dictionary function, for instance, was frequently used to clarify word meanings and collocations during the revision process. As one participant noted, *"I often checked word meanings and collocations using the dictionary inside the app."* (P13) Others highlighted the benefits of peer interaction, describing how they exchanged annotated transcripts and provided each other with feedback to improve spoken output.

In contrast, operating conditions presented certain barriers to seamless use. The most commonly cited issue was the requirement for stable internet connectivity to access real-time grammar feedback. As P41 explained, *"When the Wi-Fi was weak, the grammar feedback didn't show up, so I had to wait."* While this did not prevent continued use of the tools, it did interrupt the immediacy of the revision cycle and occasionally delayed completion of the tasks.

Overall, learners' reflections suggest that facilitating conditions involved a dynamic interplay of affordances and limitations. While built-in resources and social collaboration enhanced the learning environment, technological dependencies such as internet access occasionally disrupted the flow of practice.

6) Theme 5: Hedonic motivation

Hedonic motivation refers to the pleasure and enjoyment learners experience when using mobile ASR and AWE. Although it was not the most dominant theme, participants consistently described how emotionally engaging features contributed to greater interest and sustained interaction with the tools.

The central sub-theme was enjoyment, particularly through vocabulary learning supported by multimodal content. Students reported that song lyrics helped reduce the stress of memorization and created a more relaxed learning environment. Some also highlighted the role of short videos, which made target words more vivid and memorable through visual and auditory cues. As one participant recalled, *"Some words came with a mini video. I clicked it out of curiosity, and it stuck in my head."* (P20)

These findings suggest that enjoyment was not merely incidental but an active driver of engagement. When learning felt personally meaningful and entertaining, students were more likely to explore language repeatedly and integrate the practice into their routines. Hedonic features thus supported both motivation and retention, especially in self-paced mobile learning contexts.

7) Theme 6: Habit

Habit refers to the extent to which learners internalize mobile ASR and AWE as part of their regular learning routines. Over time, participants reported that repeated exposure and perceived benefits led to increasingly consistent and autonomous usage of the tools.

Two sub-themes were identified: dependency and habitual usage. Under the first, several students expressed a growing reliance on the feedback process to feel confident and prepared before speaking tasks. Some stated that they performed noticeably better only after practicing with ASR and AWE, reflecting an emerging sense of technological dependence for skill enhancement. As one participant shared, *"I feel I speak better only when I've practiced using ASR–AWE before class."* (P06)

The second sub-theme involved the routinization of tool use. Participants described how practices such as reviewing drafts with ASR or verifying recordings with AWE became embedded in their weekly study patterns. This shift from occasional use to regular, self-initiated practice suggested a high level of acceptance and behavioral integration. Learners no longer treated the tools as supplementary but rather as essential components of their speaking development.

These findings highlight the potential of mobile-assisted technologies to support sustainable language learning behaviors. When learners recognize tangible benefits, consistent use can evolve into enduring habits that reinforce motivation, self-monitoring, and long-term competence.

D. Discussion

1) Summary of key findings

This study investigated the impact of mobile ASR and AWE on the speaking competence of Chinese university EFL learners. Two research questions guided the investigation: (1) To what extent does mobile-based ASR and AWE affect learners' speaking competence? (2) How do learners perceive the impact of ASR- and AWE-supported speaking practice on various dimensions of their oral performance?

The quantitative findings demonstrated a clear and substantial impact of the ASR–AWE intervention. While both EG and CG showed statistically significant gains from pre-test to post-test, the improvement in EG was notably greater. Within-group comparisons revealed more consistent and marked progress among EG participants, with all students achieving higher scores following the intervention. Between-group analysis confirmed that the groups were equivalent at baseline, but diverged significantly at the post-test stage, with EG outperforming CG by a meaningful margin. These results indicate that the integrated ASR–AWE approach was effective in supporting measurable improvements in speaking competence. Similar findings have been reported in studies where mobile ASR and AWE significantly enhanced learners' pronunciation, fluency, and grammatical accuracy in oral tasks [4, 7, 10].

The qualitative findings complemented and enriched the quantitative results. Drawing on the UTAUT2 framework, six key themes emerged from students' reflective journals, focus group interviews, and annotated screenshots. Performance expectancy was the most frequently cited theme, with students reporting gains in grammar, pronunciation, vocabulary, discourse organization, and communication strategies. This dimension has also been emphasized in prior studies on mobile-assisted speaking instruction [27, 37]. Effort expectancy was also prominent, as learners described the tools as intuitive, time-efficient, and easy to integrate into their study routines- a finding consistent with UTAUT2-based MALL studies highlighting ease of use and technology acceptance [28, 38]. In addition, students highlighted the motivational support received from peers and instructors (social influence), the availability of useful

resources and system features (facilitating conditions), the enjoyment of multimedia-enhanced vocabulary learning (hedonic motivation), and the development of routine or habitual usage (habit). These aspects have also been identified in recent research exploring learner perceptions of ASR- and AWE-based tools [29, 39]. Collectively, these perceptions suggest that students found the ASR–AWE tools not only beneficial but also accessible, engaging, and sustainable over time.

In summary, the findings from both quantitative and qualitative strands converged to affirm the effectiveness of the ASR–AWE integrated approach. The intervention enhanced students' speaking competence in measurable ways and was perceived positively across multiple dimensions of user experience.

2) Interpretation of findings

The findings of this study offer important insights into how the integration of ASR and AWE technologies within a mobile-supported environment can enhance EFL learners' speaking competence.

Quantitative results showed that EG significantly outperformed CG in post-test speaking scores, indicating that the integrated use of ASR and AWE was effective in promoting measurable oral proficiency gains. These improvements were not limited to pronunciation or grammar alone; rather, the cycle of transcription, automated feedback, and revision encouraged learners to reflect meta-linguistically and refine their performance across multiple layers of speaking competence. This aligns with recent studies showing that structured engagement with ASR and feedback tools facilitates improvements beyond phonological skills, extending into syntactic accuracy and discourse fluency [2, 4, 5]. Unlike prior studies that employed ASR primarily for pronunciation training [7, 8], this study demonstrated that when learners engage with both ASR and AWE in a structured, recursive process, they also benefit in discourse organization and syntactic control.

The qualitative findings provided further explanatory depth. Performance expectancy emerged as the most salient UTAUT2 dimension, with students reporting perceived improvements in grammar, pronunciation, and coherence. These learner perceptions aligned with the observed post-test gains and further underscored the pedagogical value of the dual-technology approach. Similar results were observed in recent research where students cited increased perceived usefulness and motivation after using integrated mobile feedback tools [40, 41]. Students' reflections also highlighted the importance of effort expectancy: ASR and AWE were considered easy to use, flexible, and convenient, which helped sustain learner motivation and consistent use over time- findings echoed in studies using the UTAUT2 model to explore technology adoption in language learning [28, 38].

Notably, many students described how the ASR and AWE routine became part of their regular learning habits, supported by facilitating conditions such as mobile access and clear system interfaces. This habitual use suggests that the technologies not only enhanced linguistic outcomes but also contributed to behavioral change. Such patterns mirror recent findings showing that mobile ASR and AWE encourage routinized and autonomous language practice [31]. Although hedonic motivation was less frequently mentioned, several learners noted that interactive and multimodal features helped reduce anxiety and foster a more engaging learning environment, an important consideration in speaking-focused EFL settings [30, 38].

The convergence between learner perceptions and observed gains suggests a close alignment between subjective experience and objective improvement. For instance, the quantitative increase in discourse management scores is mirrored by learners' reflections on improved organization and coherence under performance expectancy. Similarly, gains in phonological control and grammatical range and accuracy correspond with feedback-driven revisions described under both performance and effort expectancy, where students emphasized ease of repeated practice. Other constructs, such as social influence and facilitating conditions, while not directly linked to scoring dimensions, were associated with increased willingness and capacity to sustain speaking routines—factors likely contributing to broader gains in fluency and strategic planning. These parallels affirm that learners' engagement with mobile ASR and AWE not only shaped their attitudes but also translated into measurable linguistic development across multiple dimensions of speaking competence.

These findings contribute to the growing literature on mobile-assisted language learning by demonstrating the value of combining ASR and AWE into a cohesive instructional design. While earlier studies have shown that mobile tools can support lexical and fluency development [16, 17] or promote self-confidence [20], the present study extends these insights by highlighting how the strategic integration of feedback technologies supports the holistic development of speaking competence, including grammatical accuracy, discourse organization, and communicative strategies. It is also worth noting that EG engaged in reflective tasks such as journal writing and screenshot annotation, which may have heightened their engagement and metalinguistic awareness. While these tasks were not designed to deliver instructional feedback, their presence could have indirectly supported the learning process.

In sum, the intervention was effective because it engaged learners in an ongoing process of production, feedback, and revision, supporting both accuracy and fluency development. This recursive, self-regulated process allowed them to monitor, evaluate, and refine their output, making the learning experience more interactive, individualized, and effective. In doing so, the study addresses persistent challenges in EFL contexts, such as limited speaking time and lack of personalized feedback, by offering a scalable model that supports comprehensive oral development.

3) Contributions of the study

This study provides several contributions to the understanding of how mobile-based technologies can support the development of speaking competence in EFL contexts.

First, it addresses a clear research gap by investigating the integrated use of ASR and AWE in speaking instruction. Although previous studies have explored ASR and AWE separately [26, 27, 29], limited research has examined their combined use to support both pronunciation and grammar development in oral tasks. For instance, Guskaroska [27] and Hirschi *et al.* [29] focused on learners' acceptance and performance in ASR-supported pronunciation learning, while

Barrot [26] reviewed trends in automated writing evaluation for instruction and assessment. However, these studies did not consider the pedagogical value of integrating both technologies in a single feedback cycle. The present study addresses this gap by combining ASR and AWE to enhance multiple dimensions of speaking competence through a recursive, feedback-driven approach.

Second, the study offers practical insights for learners in contexts with limited access to individualized oral feedback. The ASR–AWE practice cycle allowed learners to engage in repeated speaking tasks, receive system-generated feedback, and revise their output independently. This process supported learners' self-monitoring and encouraged consistent oral practice beyond the classroom. The structured yet flexible workflow may serve as a useful reference for those seeking to support speaking improvement through mobile tools.

This study establishes a foundational basis for future research on learner interaction with integrated educational technologies. At the same time, it offers new empirical understanding of how mobile ASR and AWE can be effectively combined to support speaking practice and learner autonomy in EFL university settings.

4) Limitations

While the study provides meaningful findings regarding the integration of ASR and AWE in speaking instruction, several limitations should be acknowledged.

The study relied on specific ASR and AWE tools whose feedback capabilities may not fully cover all aspects of speaking competence. For instance, the AWE system primarily focused on grammatical accuracy and did not offer discourse-level or pragmatic feedback. Similarly, the ASR component provided transcription and pronunciation models but lacked interactive features. These tool-specific limitations may have shaped the scope of learner engagement and feedback interpretation.

A further limitation concerns the difference in engagement formats between the two groups. While CG followed regular instructions, EG submitted reflective journals and annotated screenshots throughout the intervention. These tasks were intended for data collection and did not involve any pedagogical guidance or feedback; however, they may have encouraged greater cognitive involvement and motivation among experimental participants. Therefore, this engagement-related bias should be considered when assessing the extent to which the observed gains were due to the ASR-AWE technology itself, rather than to increased learner investment.

Despite these constraints, the study offers a valuable starting point for understanding how integrated mobile technologies can support EFL speaking development. Future research can build on this work by expanding its scope and refining the technological components involved.

5) Implications and suggestions

The findings of this study yield several implications for EFL learners, instructors, and researchers interested in the use of mobile technologies to support speaking development.

For learners, the ASR–AWE integrated approach provides a practical method for improving spoken English outside the classroom. The ability to receive immediate feedback on both pronunciation and grammar encourages self-monitoring and promotes repeated, focused practice. Learners are encouraged to develop personalized routines using these tools—such as reviewing system feedback, revising spoken content, and recording improved versions—to strengthen their oral accuracy and fluency over time.

For instructors and course designers, the results suggest that integrating ASR and AWE into speaking activities can enhance the effectiveness of oral training, particularly when classroom time for speaking is limited. Teachers may incorporate ASR–AWE cycles into homework assignments, pronunciation tasks, or speaking portfolios to support learner autonomy. Instructors should also provide initial guidance on how to interpret and apply system feedback, as some learners may need scaffolding to maximize the benefits of technology-assisted revision.

For researchers, this study highlights the value of investigating how different types of automated feedback can be combined to support multidimensional language development. Future studies may explore the long-term effects of such integrated practice, compare different types of ASR or AWE tools, or examine how learners at different proficiency levels engage with this approach. In addition, further research could explore how peer collaboration or teacher input might be blended with ASR–AWE systems to support social and interactive dimensions of speaking competence.

Finally, future research and implementation efforts must also consider potential challenges related to scalability and institutional readiness. Variations in device availability, internet connectivity, and digital literacy across learner populations may affect the feasibility of widespread adoption. Although the integrated use of mobile ASR and AWE was designed for self-regulated learning, its successful integration into instructional contexts may still depend on initial teacher guidance, ongoing technical support, and adequate infrastructure. In addition to addressing these practical issues, future studies should also consider design refinements to reduce potential motivational bias. For example, comparing two equally engaging technology-supported conditions, such as ASR versus AWE, or two ASR platforms, while ensuring that both groups complete comparable reflective or metacognitive activities, would allow more valid attribution of observed gains to specific technological features. Addressing both logistical and methodological concerns will be essential for ensuring that mobile-based speaking interventions can be equitably, sustainably, and rigorously implemented across diverse educational settings.

V. Conclusion

This study aimed to evaluate the impact of ASR and AWE on the speaking competence of Chinese university EFL learners. Guided by two research questions, it employed a mixed-methods quasi-experimental design to evaluate both the performance outcomes and learner perceptions associated with ASR–AWE-supported speaking practice.

The findings showed that learners who engaged in the ASR-AWE integrated practice demonstrated significantly greater improvements in their speaking performance compared to those who followed traditional instruction. These gains were evident across key dimensions of speaking competence, including pronunciation, grammatical accuracy,

and discourse organization. Thematic analysis further revealed that students viewed the system as effective, easy to use, and motivating, with many incorporating it into their regular study habits. The convergence of statistical gains and perceived usefulness supports the pedagogical value of combining ASR and AWE to promote multidimensional oral development.

In terms of practical implications, the results suggest that ASR- and AWE-supported mobile learning can offer a scalable and repeatable complement to classroom-based instruction. However, cost-related constraints must be considered. While some learners benefited from freely available features of applications such as Youdao, full access to advanced functions may require paid subscriptions, which can limit uptake for certain users. To address access disparities, institutions may consider supporting technology adoption through mechanisms such as bulk licensing, negotiated institutional access, or the use of functionally equivalent open-source alternatives. At the same time, learners can be guided to critically evaluate and select mobile tools that align with their learning needs and digital literacy levels. Furthermore, targeted teacher professional development is essential to ensure that instructors can effectively guide learners in interpreting system-generated feedback and integrating it into iterative speaking practice.

Several limitations of the present study should be acknowledged. The sample was limited to first-year Business English majors at a single university, which may constrain the generalizability of findings. Future studies should examine more diverse learner populations and conduct cross-platform comparisons to explore whether similar effects are observed across different ASR and AWE tools. In addition, a delayed post-test was not administered in this study, which restricts conclusions about the long-term retention of speaking improvements. Follow-up research incorporating longitudinal tracking would provide stronger evidence regarding the sustainability of learning gains.

In sum, the integration of mobile ASR and AWE represents a promising approach to supporting EFL learners' speaking development through individualized, feedback-driven, and self-regulated practice. As mobile learning technologies continue to evolve, their pedagogical potential will increasingly depend on thoughtful implementation strategies, inclusive access policies, and support for teacher adaptation. This study contributes to an emerging understanding of how integrated feedback tools can be effectively leveraged to enhance speaking competence in technology-mediated language learning environments.

APPENDIX

Table A1. BEC Business vantage speaking assessment scale

B2	D1	D2	D3	D4
5	Shows a good degree of control of a range of simple and some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on a wide range of familiar topics.	Produces extended stretches of language with very little hesitation. Contributions are relevant and there is a clear organization of ideas. Uses a range of cohesive devices and discourse markers.	Is intelligible. Intonation is appropriate. Sentence and word stress are accurately placed. Individual sounds are articulated clearly.	Initiates and responds appropriately, linking contributions to those of other speakers. Maintains and develops the interaction and negotiates towards an outcome.
4	Performance shares features of Bands 3 and 5.			
3	Shows a good degree of control of simple grammatical forms and attempts some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on a range of familiar topics.	Produces extended stretches of language despite some hesitation. Contributions are relevant and there is very little repetition. Uses a range of cohesive devices.	Is intelligible. Intonation is generally appropriate. Sentence and word stress is generally accurately placed. Individual sounds are generally articulated clearly.	Initiates and responds appropriately. Maintains and develops the interaction and negotiates towards an outcome with very little support.
2	Performance shares features of Bands 1 and 3.			
1	Shows a good degree of control of simple grammatical forms. Uses a range of appropriate vocabulary when talking about everyday situations	Produces responses which are extended beyond short phrases, despite hesitation. Contributions are mostly relevant, despite some repetition. Uses basic cohesive devices.	Is mostly intelligible and has some control of phonological features at both utterance and word levels.	Initiates and responds appropriately. Keeps the interaction going with very little prompting and support.
0	Performance below Band 1.			

Notes: D1: grammar and vocabulary; D2: discourse management; D3: pronunciation; D4: interactive communication

Table A2. Reflective journal

Construct	Questions
Performance Expectancy	1.How do you find ASR and AWE useful for improving speaking competence?
Effort Expectancy	2. How easy does using ASR and AWE to improve speaking competence seem to be?
Social influence	3. How would your family, friends, and teacher perceive you use ASR and AWE to improve speaking competence?
Facilitating Conditions	4. Do you have the necessary resources and knowledge that would allow you to use ASR and AWE to improve speaking competence?
Hedonic Motivation	5. For you, what is interesting, fun, and enjoyable when you use ASR and AWE to improve speaking competence?
Habit	6. Has using ASR and AWE to improve your speaking competence become a habit for you?

Table A3. Focus group interview

Construct	Questions
Performance Expectancy	1.How do you find ASR and AWE useful for improving speaking competence?
Effort Expectancy	2. How easy does using ASR and AWE to improve speaking competence seem to be?

Social influence	3. How would your family, friends, and teacher perceive you use ASR and AWE to improve speaking competence?
Facilitating Conditions	4. Do you have the necessary resources and knowledge that would allow you to use ASR and AWE to improve speaking competence?
Hedonic Motivation	5. For you, what is interesting, fun, and enjoyable when you use ASR and AWE to improve speaking competence?
Habit	6. Has using ASR and AWE to improve your speaking competence become a habit for you?

Table A4. Thematic quotes from learner reflections organized by UTAUT2

Theme	Sub-themes	Sub sub-themes	Representative Quote
Performance Expectancy	Enhancing the language and discourse knowledge	Grammatical knowledge	"The system showed me how I kept using incorrect tenses when speaking." (P07)
		Phonological knowledge	"I realized my pronunciation was unclear when ASR failed to recognize my words." (P22)
		Lexical knowledge	"I learned how to use some advanced words by checking suggestions in AWE." (P13)
		Discourse knowledge	"Seeing the transcript helped me organize long answers more logically." (P34)
	Enhancing the core speaking skills	Pronunciation skills	"I repeated the same sentence until ASR recognized every word. That improved clarity." (P26)
		Speech function	"It helped me adjust how I made requests and gave opinions more naturally." (P48)
		Discourse organization	"The feedback made me restructure my answer to have a clearer beginning and ending." (P09)
	Enhancing the communication strategies	Cognitive strategies	"I began planning what to say in my head while using the tool, which helped fluency." (P30)
		Metacognitive strategies	"After checking the transcript, I always reflected on how I could improve next time." (P24)
		Discourse planning and monitoring	"Now I try to make my speech more connected and logical when I revise it." (P46)
Effort Expectancy	Perceived ease	Easy to use	"I just opened the app, clicked record, and started speaking. It was so simple." (P11)
		Convenience and flexibility	"Even if I only had 10 minutes between classes, I could still do a full practice." (P08)
	User interface	Comfortable with ASR's pronunciation	"The pronunciation examples were clear and easy to imitate." (P31)
		Comfortable with the user interface	"The layout was clean, and I could easily find what I needed." (P19)
Social Influence	Peer influence	Support from peer	"My roommate told me she always used the grammar suggestions to polish her answers." (P29)
	Academic influence	Support from instructor	"Our teacher said this tool can really help us correct ourselves, so I used it weekly." (P35)
Facilitating Condition	Resource support	Authoritative dictionaries	"I often checked word meanings and collocations using the dictionary inside the app." (P13)
		Exchange learning strategy	"We sometimes shared transcripts and gave each other feedback to revise sentences." (P22)
	Operating conditions	Lack of offline grammar correction	"When the Wi-Fi was weak, the grammar feedback didn't show up, so I had to wait." (P41)
Hedonic Motivation	Enjoyment	Vocabulary learning through song lyrics	"I liked checking words from songs. It felt more relaxing than just memorizing." (P12)
		Vocabulary learning through short video	"Some words came with a mini video. I clicked it out of curiosity, and it stuck in my head." (P20)
Habit	Dependency	Dependency on ASR-AWE for improving speaking competence	"I feel I speak better only when I've practiced using ASR–AWE before class." (P06)
		Frequent use of ASR-AWE	"I use it at least three times a week. It's become part of how I study now." (P17)
	Habitual usage	Routine use of ASR-AWE in language practice	"I always check my answer with ASR–AWE before I record the final version." (P21)

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

W.Y.L., M.B.M., and H.W.Y conducted the research, analyzed the data, and wrote the paper. All authors had approved the final version.

References

[1] C. Shi, A. Kassim, and N. R. M. Radzuan, "Improving EFL learners' English public speaking performance through project-based learning strategy at tertiary level," *Language Teaching Research Quarterly,* vol. 42, pp. 126–144, 2024.

[2] G. Zhang and H. Lu, "Impact of English-speaking environments and Chinese language pronunciation on the speaking proficiency of English learners in China: A comprehensive study," *Journal of Psycholinguistic Research,* vol. 53, no. 3, p. 45, 2024.

[3] B. Zou, S. Liviero, Q. Ma, W. Zhang, Y. Du, and P. Xing, "Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice," *System,* vol. 126, 103497, 2024.

[4] G.-J. Hwang, M. Rahimi, and J. Fathi, "Enhancing EFL learners' speaking skills, foreign language enjoyment, and language-specific grit utilising the affordances of a MALL app: A microgenetic perspective," *Computers & Education,* vol. 214, 105015, 2024.

[5] Y. Xue, "Towards automated writing evaluation: A comprehensive review with bibliometric, scientometric, and meta-analytic approaches," *Education and Information Technologies,* vol. 29, no. 15, pp. 19553–19594, 2024.

[6] C. C. Goh and A. Burns, *Teaching Speaking: A Holistic Approach,* Cambridge University Press, 2012.

[7] T. T.-N. Ngo, H. H.-J. Chen, and K. K.-W. Lai, "The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A

meta-analysis," *ReCALL,* vol. 36, no. 1, pp. 4–21, 2024.
[8] M. Amrate and P.-H. Tsai, "Computer-assisted pronunciation training: A systematic review," *ReCALL,* pp. 1–21, 2024.
[9] M. H. Al-Khresheh, "The role of presentation-based activities in enhancing speaking proficiency among Saudi EFL students: A quasi-experimental study," *Acta Psychologica,* vol. 243, 104159, 2024.
[10] M. S. Uztosun, "Cultivating oral communication skills in academic contexts," *Teaching English for Academic Purposes: Theory into Practice*, Springer, 2024, pp. 143–167.
[11] G. Yang, Y. Wang, Y. Zhang, M. Yang, Q. Zeng, and Z. Song, "An empirical study of AI-supported interleaved training strategy to improve EFL students' English impromptu speaking performance, learning engagement, technology acceptance and epistemic network structure," *The Asia-Pacific Education Researcher,* pp. 1–22, 2025.
[12] R. Shadiev, Y. Feng, R. Zhussupova, and F. Altinay, "Effects of speech-enabled corrective feedback technology on EFL speaking skills, anxiety and confidence," *Computer Assisted Language Learning,* pp. 1–37, 2024.
[13] K. Karakaya and A. Bozkurt, "Mobile-Assisted Language Learning (MALL) research trends and patterns through bibliometric analysis: Empowering language learners through ubiquitous educational technologies," *System,* vol. 110, 102925, 2022.
[14] H. B. S. G. Al Shihri, O. H. A. Mahfoodh, and A. B. B. M. A. Khan, "Examining the effect of the integration of multiple MALL applications on EFL students' academic vocabulary acquisition: A mixed-methods study," *Cogent Education,* vol. 12, no. 1, 2473229, 2025.
[15] J. Fathi, M. Rahimi, S. Nourzadeh, and R. Shadiev, "Examining the role of voice blogging in developing speaking complexity, accuracy, and fluency in EFL learners: A social-constructivist perspective," *Interactive Learning Environments,* pp. 1–28, 2025.
[16] A. Al-Abri, F. R. Madiseh, and M. M. Moghaddam, "Exploring learning-oriented assessment in enhancing students' lexical fluency through MALL," *The Asia-Pacific Education Researcher,* vol. 34, no. 1, pp. 1–13, 2025.
[17] T.-T. Wu, I. P. Hapsari, and Y.-M. Huang, "Effects of incorporating AI chatbots into think–pair–share activities on EFL speaking anxiety, language enjoyment, and speaking performance," *Computer Assisted Language Learning,* pp. 1–39, 2025.
[18] G.-J. Hwang, M. Rahimi, and J. Fathi, "Enhancing EFL learners' speaking skills, foreign language enjoyment, and language-specific grit utilising the affordances of a MALL app: A microgenetic perspective," *Computers & Education,* vol. 214, 105015, 2024.
[19] Y. Liu, F. binti Ab Rahman, and F. binti M. Zain, "A systematic literature review of research on automatic speech recognition in EFL pronunciation," *Cogent Education,* vol. 12, no. 1, 2466288, 2025.
[20] M. Y. C. Jiang, M. S. Y. Jong, W. W. F. Lau, C. S. Chai, and N. Wu, "Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom," *Journal of Computer Assisted Learning,* vol. 39, no. 1, pp. 125–140, 2023.
[21] S. McCrocklin, C. Fettig, and S. Markus, "Salukispeech: integrating a new ASR tool into students' English pronunciation practice," *Pronunciation in Second Language Learning and Teaching Proceedings,* vol. 12, no. 1, 2022.
[22] L. Ding and D. Zou, "Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence," *Education and Information Technologies,* vol. 29, no. 11, pp. 14151–14203, 2024.
[23] M. Chen and Y. Cui, "The effects of AWE and peer feedback on cohesion and coherence in continuation writing," *Journal of Second Language Writing,* vol. 57, 100915, 2022.
[24] J. Wilson and C. MacArthur, "Exploring the role of automated writing evaluation as a formative assessment tool supporting self-regulated learning in writing," *The Routledge International Handbook of Automated Essay Evaluation*, Routledge, 2024, pp. 197–220.
[25] Q.-K. Fu, D. Zou, H. Xie, and G. Cheng, "A review of AWE feedback: Types, learning outcomes, and implications," *Computer Assisted Language Learning,* vol. 37, no. 1–2, pp. 179–221, 2024.
[26] J. S. Barrot, "Trends in automated writing evaluation systems research for teaching, learning, and assessment: A bibliometric analysis," *Education and Information Technologies,* vol. 29, no. 6, pp. 7155–7179, 2024.
[27] A. Guskaroska, "Exploring technology acceptance of ASR for pronunciation learning," *Iowa State University*, 2024.
[28] T. T. Nguyen, "Exploring the interaction of Meta-UTAUT in the mobile-assisted language learning environment for EFL learning," *Journal of Educators Online,* vol. 21, no. 4, 2024.
[29] K. Hirschi, O. Kang, J. Hansen, C. Cucchiarini, and H. Strik, "Mobile-assisted pronunciation training with adult ESOL learners: Background, acceptance, effort, and accuracy," *Pronunciation in Second Language Learning and Teaching Proceedings,* vol. 12, no. 1, 2022.
[30] D. Shu, C. Huang, and Y. Xing, "Analysis on the promotion of classroom atmosphere in multimodal English teaching based on human-computer interaction," *International Journal of Human–Computer Interaction,* vol. 40, no. 13, pp. 3516–3527, 2024.
[31] O. Khan, Z. Ullah, I. Bakht, and S. Bibi, "Learners' autonomy and Mobile Assisted Language Learning (MALL): Perceptions of undergraduate students," *Panacea Journal of Linguistics & Literature,* vol. 3, no. 1, pp. 331–343, 2024.
[32] W. Creswell John and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Los Angeles London New Dehli Singapore: Sage Publications, 2022.
[33] Ministry of Education of the People's Republic of China, *China's Standards of English Language Ability*, Beijing: Higher Education Press, 2018. (in Chinese)
[34] C. Peng, J. Liu, and H. Cai, "Aligning China's standards of English language ability with the common European framework of reference for languages," *The Asia-Pacific Education Researcher,* vol. 31, no. 6, pp. 667–677, 2022.
[35] V. Venkatesh, J. Y. Thong, and X. Xu, "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology," *MIS Quarterly,* pp. 157–178, 2012.
[36] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology,* vol. 3, no. 2, pp. 77–101, 2006.
[37] J. Xie and J. Jiao, "Enhancing language education in developing countries through intelligent transformation: A comprehensive study," *Educational Technology Research and Development,* vol. 73, no. 2, pp. 1263–1291, 2025.
[38] Y. Zhang and W. Pan, "A UTAUT2 model expansion: Investigating the effect of interactive learning environment and gamification on duolingo user base in China," *International Journal of Human–Computer Interaction,* pp. 1–15, 2024.
[39] A. Bessadok and M. Hersi, "A structural equation model analysis of English for specific purposes students' attitudes regarding computer-assisted language learning: UTAUT2 model," *Library Hi Tech,* vol. 43, no. 1, pp. 36–55, 2025.
[40] X. Wang *et al.*, "Perceived usefulness predicts second language learners' continuance intention toward language learning applications: A serial multiple mediation model of integrative motivation and flow," *Education and Information Technologies,* pp. 1–17, 2022.
[41] H. B. S. G. Al Shihri, O. H. A. Mahfoodh, and A. B. B. M. A. Khan, "Examining the effect of the integration of multiple MALL applications on EFL students' academic vocabulary acquisition: A mixed-methods study," *Cogent Education,* vol. 12, no. 1, 2473229, 2025.

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).