Social Science and Human Research Bulletin

ISSN(e): 3050-5542
ISSN(p): 3050-5534

Vol. 03(05): 581-590, May 2026

Home Page: http://sshrbjournal.org/

AI-Mediated Speaking Practice for Aptis ESOL Preparation: An Explanatory Sequential Mixed-Methods Case Study of Vietnamese EFL University Students

Nguyen Vu Chinh
Hoa Sen University

Article DOI: 10.55677/SSHRB/2026-3050-0515 **DOI URL: https://doi.org/10.55677/SSHRB/2026-3050-0515**

KEYWORDS: artificial intelligence, Aptis ESOL, EFL speaking, ChatGPT, automatic speech recognition, mixed methods, speaking anxiety, case study.

Corresponding Author
Nguyen Vu Chinh

Published: May 19, 2026

ABSTRACT: Artificial intelligence is increasingly used in English language learning, yet empirical evidence remains uneven regarding how AI-mediated speaking practice supports performance in standardised speaking-test preparation. This explanatory sequential mixed-methods case study examined how Vietnamese EFL university students preparing for Aptis ESOL Speaking engaged with an AI-supported speaking programme. Forty-eight students participated in an eight-week intervention integrating Aptis-style speaking tasks, ChatGPT voice interaction, automatic speech recognition (ASR) feedback, teacher scaffolding, and reflective journals. Quantitative data included pre/post Aptis-style speaking scores, speaking anxiety ratings, willingness-to-communicate scores, and AI-use logs. Qualitative data were collected from semi-structured interviews, learner journals, and teacher field notes. The reported findings indicated a statistically significant increase in speaking performance from pre-test to post-test, alongside reduced speaking anxiety and increased willingness to communicate. Qualitative findings suggested that AI created a low-stakes rehearsal space, increased opportunities for output, supported lexical and pronunciation noticing, and enhanced learner autonomy. However, students also reported over-reliance on AI-generated scripts, uneven feedback quality, and difficulties transferring rehearsed fluency to spontaneous test performance. The study argues that AI is most pedagogically valuable when integrated as a scaffolded rehearsal partner rather than as a substitute teacher, examiner, or answer generator.

I. INTRODUCTION

Speaking has long been regarded as one of the most demanding areas of second and foreign language learning because it requires learners to coordinate linguistic knowledge, discourse organisation, pronunciation, interactional judgement, and affective control in real time (Bygate, 2001; Derwing & Munro, 2015; Goh & Burns, 2012). Unlike reading or writing, speaking offers little time for revision during performance. Learners must select vocabulary, organise ideas, monitor grammar, pronounce intelligibly, and respond to the communicative task almost simultaneously. In English as a foreign language (EFL) contexts, these demands are intensified by limited exposure to authentic interaction, large class sizes, teacher-fronted instruction, and anxiety about public performance (Horwitz et al., 1986; Woodrow, 2006). As a result, many EFL learners know grammatical rules but struggle to produce extended, coherent, and confident spoken responses (Thanh V.T, 2026).

The challenge becomes particularly visible in test-preparation contexts. Students preparing for standardised speaking tests are expected to demonstrate not only language knowledge but also timed performance under assessment conditions. Aptis ESOL Speaking is one such context. Aptis ESOL General assesses grammar and vocabulary, speaking, writing, listening, and reading, with the speaking component lasting approximately 12 minutes and consisting of four parts (British Council, 2024a, 2024b). The test is computer-based: candidates respond to prompts and their spoken answers are recorded for assessment. The reporting of Aptis ESOL is linked to the Common European Framework of Reference for Languages (CEFR), which conceptualises language ability as communicative action rather than isolated linguistic knowledge (Council of Europe, 2020). Therefore, preparing students for

Aptis Speaking should involve more than memorising model answers. It should develop students' ability to understand prompts, organise ideas quickly, speak intelligibly, sustain fluency, and express opinions flexibly.

Recent developments in artificial intelligence (AI) have changed the ecology of speaking practice. AI chatbots, voice assistants, automatic speech recognition systems, and generative AI tools can provide learners with repeated opportunities for oral rehearsal, immediate feedback, model responses, transcript-based reflection, and personalised input (Kohnke et al., 2023; Lo et al., 2024; Shadiev & Liu, 2022; Tai, 2022). ChatGPT and similar generative AI tools are especially attractive because they can simulate dialogue, generate follow-up questions, reformulate learner output, explain vocabulary, and provide feedback at any time. In EFL speaking preparation, such tools may help address a persistent problem: students need more speaking practice than classroom time normally allows.

However, the rapid adoption of AI in language education should not be interpreted uncritically. Although AI may increase the quantity of practice, quantity alone does not guarantee communicative development. Learners may become dependent on AI-generated scripts, accept inaccurate feedback, rehearse unnatural model answers, or confuse AI fluency with their own proficiency (Kohnke et al., 2023; Lo et al., 2024; Meniado, 2023). AI feedback may also be too general to capture task fulfilment, pronunciation intelligibility, discourse coherence, and pragmatic appropriateness in test-specific contexts (Sun, 2023). These concerns are particularly important for Aptis Speaking preparation because the test rewards communicative effectiveness under timed conditions, not the reproduction of polished AI-generated sentences.

Despite growing interest in AI-supported language learning, several gaps remain. First, many studies investigate general English learning rather than speaking development in a specific standardised assessment context. Second, a substantial portion of the literature focuses on learner perceptions without triangulating performance data, affective measures, and qualitative accounts of learning processes. Third, relatively little research has examined Vietnamese EFL learners preparing for Aptis ESOL Speaking. This study addresses these gaps by examining an AI-supported Aptis Speaking preparation programme through an explanatory sequential mixed-methods case study.

The study is guided by three research questions:

RQ1: To what extent does an AI-supported Aptis Speaking preparation programme improve students' Aptis-style speaking performance?

RQ2: How does AI-mediated speaking practice affect students' speaking anxiety and willingness to communicate?

RQ3: How do students perceive the benefits, limitations, and ethical risks of using AI for Aptis Speaking preparation?

II. LITERATURE REVIEW

2.1 EFL Speaking as a Multidimensional Construct

Speaking competence is not a single skill but a complex construct involving fluency, accuracy, complexity, pronunciation, discourse management, strategic competence, and pragmatic appropriateness (Bygate, 2001; Derwing & Munro, 2015; Goh & Burns, 2012). Fluency refers to the ability to produce speech with appropriate speed, pausing, and continuity. Accuracy involves grammatical and lexical control. Complexity concerns the range and sophistication of language forms. Pronunciation includes segmental accuracy, suprasegmental features, stress, rhythm, and intelligibility. Discourse management involves organising ideas coherently and maintaining relevance to the task.

In EFL contexts, speaking is often constrained by insufficient opportunities for meaningful oral production. Swain's (2005) output hypothesis argues that producing language pushes learners to process language more deeply, notice gaps in their competence, test hypotheses, and reflect on form and meaning. Schmidt's (1990) noticing hypothesis similarly suggests that learners must attend to specific linguistic features before these features can become intake. From a sociocultural perspective, learning occurs through mediated activity and scaffolded participation (Vygotsky, 1978). These theories collectively suggest that speaking development requires repeated output, feedback, noticing, and guided reformulation.

Affective variables also shape speaking performance. Foreign language anxiety has been conceptualised as a situation-specific form of anxiety associated with communication apprehension, test anxiety, and fear of negative evaluation (Horwitz et al., 1986). Speaking anxiety may reduce learners' willingness to participate and disrupt real-time processing. Willingness to communicate (WTC), by contrast, refers to learners' readiness to enter into communication in a second language under particular conditions (MacIntyre et al., 1998). Higher WTC is associated with greater communicative engagement and more frequent language use. Therefore, an effective speaking programme should develop both linguistic competence and affective readiness.

2.2 Aptis ESOL Speaking and Test Preparation

Aptis ESOL is a computer-based English proficiency test developed by the British Council. Aptis ESOL General evaluates grammar and vocabulary as well as the four macro skills: speaking, writing, listening, and reading (British Council, 2024a). The speaking component lasts approximately 12 minutes, includes four parts, and requires candidates to respond orally to prompts under time constraints (British Council, 2024b). Results are reported with reference to CEFR levels, and skill scores are expressed on a 0-50 scale (British Council, 2024a; British Council, 2024c).

The CEFR Companion Volume emphasises communicative language activities, mediation, online communication, interaction, and plurilingual competence (Council of Europe, 2020). This expanded view of language ability is important for Aptis preparation because students must not only produce grammatically correct sentences but also respond meaningfully to prompts. A high-quality speaking-preparation programme should therefore include task familiarisation, timed practice, discourse organisation, vocabulary expansion, pronunciation work, and opportunities for reflective self-assessment.

Nevertheless, test preparation can become reductive. Students may overuse memorised templates, rely on formulaic openings, and reproduce rehearsed responses regardless of the prompt. Formulaic language can support fluency when used appropriately (Wray, 2002), but over-reliance on fixed scripts may reduce flexibility and authenticity. This is a key concern in AI-mediated Aptis preparation because generative AI can produce fluent model answers that students may be tempted to memorise.

2.3 AI in English Language Learning

AI has become a major topic in language education. Recent reviews and meta-analyses indicate that AI can support English learning through personalised feedback, adaptive practice, learner engagement, and expanded opportunities for interaction (Chen et al., 2024; Cislowska & Pena-Acuna, 2024; Lo et al., 2024). Chen et al. (2024), for example, reported a positive overall effect of AI on English language learning achievement. However, AI effectiveness varies by learning objective, instructional design, learner proficiency, teacher mediation, and assessment method.

Generative AI has intensified these debates. ChatGPT can provide meaning-focused input, scaffold output, correct language, generate examples, and simulate interaction (Kohnke et al., 2023; Meniado, 2023). A systematic review by Lo et al. (2024) found rapidly increasing empirical interest in ChatGPT for ESL/EFL education and highlighted applications across writing, grammar, vocabulary, speaking, and learner support. Yet the same body of literature also identifies serious concerns: inaccurate information, hallucinated content, academic dishonesty, overdependence, privacy concerns, and the need for AI literacy (Kohnke et al., 2023; Lo et al., 2024; Meniado, 2023).

A critical approach to AI in language education therefore asks not whether AI should be used, but how it should be pedagogically integrated. AI is not inherently educational. It becomes educational when teachers design tasks that require learners to interact with AI output critically, compare it with task criteria, revise their own production, and reflect on learning.

2.4 AI-Mediated Speaking Practice

AI-mediated speaking practice may address several limitations of traditional EFL classrooms. First, it increases access to speaking opportunities outside class. Tai (2022) found that intelligent personal assistants could support EFL learners' oral proficiency through out-of-class interaction. Second, AI interlocutors may reduce fear of negative evaluation because students can practise privately and repeat tasks without social embarrassment. Third, AI can provide immediate prompts, follow-up questions, and reformulations, allowing students to engage in cycles of production and revision.

Research on AI-speaking tools suggests that they can influence both performance and affect. Zhang et al. (2024) examined an AI-speaking assistant and found effects on enjoyment, anxiety, and willingness to communicate among Chinese EFL learners. These findings are relevant because learners who enjoy speaking and feel less anxious are more likely to practise, take risks, and sustain engagement. Similarly, studies on intelligent personal assistants indicate that AI-human interaction can support willingness to communicate outside the classroom (Tai & Chen, 2023; Vo, 2026).

However, AI-mediated speaking has limitations. AI systems may not reliably assess communicative appropriateness, discourse depth, pronunciation intelligibility, or test-specific task fulfilment. AI feedback may reward long, grammatically polished responses even when they do not answer the prompt directly. For standardised speaking-test preparation, this limitation is serious because students may become fluent in an AI-supported environment but fail to perform spontaneously under timed test conditions.

2.5 Automatic Speech Recognition and Noticing

Automatic speech recognition (ASR) is another important component of AI-supported speaking learning. ASR transforms spoken language into written transcripts, which can help learners notice discrepancies between intended and recognised output. Shadiev and Liu (2022) reviewed applications of speech recognition technology in language learning and concluded that it has potential for supporting pronunciation, speaking practice, vocabulary learning, and learner autonomy. Liu et al. (2022) found that college students generally perceived ASR as useful for improving oral English proficiency. Sun (2023) used an explanatory sequential mixed-methods design and found that ASR combined with peer correction could support pronunciation and speaking development.

Transcript-based noticing is especially relevant to Aptis preparation. When students read their own ASR transcripts, they may notice repeated vocabulary, unclear sentence boundaries, grammatical omissions, and weak connectors. However, ASR output is not a perfect representation of learner speech. Recognition accuracy may vary by accent, background noise, speech rate, and pronunciation patterns (Evers & Chen, 2022; Shadiev & Liu, 2022). Therefore, ASR should be used as a reflective tool rather than as an absolute evaluation of speaking quality.

2.6 Research Gap and Conceptual Framework

The literature suggests that AI can support EFL speaking through increased practice, reduced anxiety, immediate feedback, and learner autonomy (Lo et al., 2024; Shadiev & Liu, 2022; Tai, 2022; Zhang et al., 2024). Yet three gaps remain. First, there is limited

mixed-methods evidence on AI-mediated speaking practice in Aptis ESOL preparation. Second, many studies examine AI tools without connecting them to a specific speaking-test construct. Third, more research is needed on how learners critically interpret AI feedback rather than merely receive it.

This study conceptualises AI-mediated Aptis preparation as a scaffolded cycle:

Aptis-style prompt -> AI rehearsal -> ASR transcript -> learner noticing -> self-reformulation -> teacher feedback -> timed re-performance.

This model draws on output theory (Swain, 2005), noticing theory (Schmidt, 1990), sociocultural theory (Vygotsky, 1978), and research on WTC and speaking anxiety (Horwitz et al., 1986; MacIntyre et al., 1998). It positions AI as a mediational tool that can support practice and reflection but cannot replace human pedagogical judgement.

III. METHODOLOGY

3.1 Research Design

The study adopted an explanatory sequential mixed-methods case study design (Creswell & Plano Clark, 2018; Yin, 2018). In the first phase, quantitative data were collected through pre/post Aptis-style speaking tests and questionnaires measuring speaking anxiety and willingness to communicate. In the second phase, qualitative data were collected through semi-structured interviews, learner journals, AI-use logs, and teacher field notes. The qualitative phase was designed to explain, extend, and complicate the quantitative results.

This design was appropriate because AI-mediated speaking development involves both measurable outcomes and lived learning experiences. Quantitative data can show whether performance and affective variables changed, but they cannot fully explain how students used AI, whether they trusted its feedback, or why some students benefited more than others. Qualitative data can illuminate these processes and identify pedagogical risks.

3.2 Research Context

The case was an eight-week Aptis Speaking preparation course at a Vietnamese university. The course aimed to help students become familiar with Aptis-style speaking tasks while developing fluency, coherence, vocabulary range, grammatical control, pronunciation, and confidence. The course met twice a week, with additional AI-mediated speaking practice assigned outside class.

The institutional context was typical of many EFL settings in Vietnam: students had studied English for several years but had limited opportunities for extended spoken interaction outside the classroom. Most students reported that they were more comfortable with grammar and reading than with spontaneous speaking. Before the intervention, many students described Aptis Speaking as stressful because they had to speak into a computer, manage time limits, and organise answers without direct support from an interlocutor.

3.3 Participants

The illustrative sample consisted of 48 Vietnamese EFL university students aged 18-22. Their estimated English proficiency ranged from A2+ to B1+, based on placement information, previous coursework, and teacher evaluation. All participants were preparing for Aptis ESOL General.

For the qualitative phase, 12 students were purposively selected. Selection was based on three criteria: speaking-score gain, change in speaking anxiety, and frequency of AI use. Four students represented high improvement, four represented moderate improvement, and four represented limited improvement. This sampling strategy allowed the study to explore both successful and less successful experiences with AI-mediated speaking practice.

3.4 Intervention

The intervention lasted eight weeks and combined teacher-led strategy instruction, AI-mediated speaking rehearsal, ASR transcript review, reflective journaling, and timed re-performance. Students used ChatGPT voice interaction as a conversational rehearsal partner and reviewed ASR transcripts after speaking tasks. Teacher-designed prompt banks were aligned with Aptis Speaking task types.

Each weekly cycle included six stages:

Task orientation. The teacher introduced an Aptis-style speaking function such as personal response, picture description, comparison, opinion-giving, reason development, or extended explanation.

AI rehearsal. Students practised the task with ChatGPT voice interaction using teacher-designed prompts.

Transcript noticing. Students reviewed ASR transcripts and highlighted unclear sentences, repeated vocabulary, grammar errors, and weak connectors.

Self-reformulation. Students revised their responses by selectively using AI suggestions while preserving their own ideas.

Teacher feedback. The teacher reviewed selected recordings and provided feedback on task fulfilment, coherence, vocabulary, grammar, pronunciation, and timing.

Timed re-performance. Students repeated the task under Aptis-style time limits.

Students were explicitly instructed not to memorise AI-generated answers. Instead, they were taught to use AI for idea expansion, vocabulary alternatives, follow-up questions, and self-reflection.

3.5 Instruments

3.5.1 Aptis-Style Speaking Test

Students completed a pre-test and post-test modelled on the Aptis Speaking format. The test contained four parts: personal information, description, comparison, and opinion-based extended response. Responses were recorded and assessed by two trained raters using an analytic 0-50 scale adapted from Aptis and CEFR-oriented descriptors.

The rubric contained five dimensions: task fulfilment, fluency and coherence, lexical range, grammatical control, and pronunciation/intelligibility. Each dimension was rated on a 0-10 scale, producing a total score out of 50.

3.5.2 Speaking Anxiety Questionnaire

Speaking anxiety was measured using a 5-point Likert-scale questionnaire adapted from foreign language classroom anxiety research (Horwitz et al., 1986; Woodrow, 2006). Items addressed fear of making mistakes, nervousness during timed speaking, fear of negative evaluation, and discomfort with recorded speaking.

3.5.3 Willingness-to-Communicate Questionnaire

Willingness to communicate was measured using a 5-point Likert-scale questionnaire adapted from MacIntyre et al. (1998). Items asked students how willing they were to speak English with classmates, teachers, AI tools, and in test-like situations.

3.5.4 Interviews, Journals, and Field Notes

Semi-structured interviews were conducted with 12 students after the intervention. Interview questions focused on AI use, perceived benefits, feedback quality, anxiety, autonomy, script memorisation, and transfer to Aptis-style tasks. Students also wrote weekly reflective journals. The teacher kept field notes on student engagement, common errors, and classroom observations.

3.6 Data Analysis

Quantitative data were analysed using paired-samples *t*-tests. Effect sizes were calculated using Cohen's *dz* for within-subject designs. Inter-rater reliability for speaking scores was estimated using the intraclass correlation coefficient (ICC). Internal consistency for questionnaire scales was estimated using Cronbach's alpha.

Qualitative data were analysed thematically following Braun and Clarke's (2006) six-phase procedure: familiarisation, initial coding, searching for themes, reviewing themes, defining themes, and writing up findings. Interview transcripts, learner journals, AI-use logs, and field notes were triangulated to enhance credibility. Representative excerpts were selected to illustrate major themes. All student names were replaced with pseudonyms.

3.7 Ethical Considerations

Participants were informed that AI tools were used for learning support, not official assessment. They were told that participation was voluntary and that they could withdraw at any time. Recordings and transcripts were anonymised. Students were instructed not to submit AI-generated scripts as their own spontaneous speech. The study followed principles of informed consent, confidentiality, responsible AI use, and transparency.

IV. RESULTS

4.1 Speaking Performance

Students' Aptis-style speaking scores increased from pre-test to post-test. Table 1 presents the overall speaking results.

Table 1. Pre-test and post-test Aptis-style speaking scores

Measure	Pre-test M	Pre-test SD	Post-test M	Post-test SD	Mean Difference	t	p	Effect Size dz
Aptis-style speaking score /50	29.40	5.80	34.20	5.60	4.80	7.92	< .001	1.14

The paired-samples *t*-test indicated a statistically significant increase in speaking performance, $t(47) = 7.92$, $p < .001$. The effect size was large, suggesting that the improvement was not only statistically significant but also educationally meaningful. Inter-rater reliability was strong, ICC = .87, indicating acceptable consistency between the two raters.

Sub-score analysis showed that the strongest gains occurred in fluency and coherence, followed by lexical range and task fulfilment. Pronunciation and grammatical control also improved, although gains were smaller.

Table 2. Speaking sub-score changes by dimension

Speaking dimension	Pre-test M	Post-test M	Mean gain
Task fulfilment	5.90	6.85	+0.95
Fluency and coherence	5.65	6.95	+1.30
Lexical range	5.70	6.80	+1.10
Grammatical control	6.05	6.70	+0.65
Pronunciation/intelligibility	6.10	6.90	+0.80

These results suggest that AI-mediated rehearsal was particularly associated with improvements in extended response organisation and lexical variety. This pattern is plausible because the intervention emphasised repeated rehearsal, discourse markers, response expansion, and transcript-based reformulation.

4.2 Speaking Anxiety and Willingness to Communicate

Table 3 presents changes in speaking anxiety, willingness to communicate, and AI-speaking confidence.

Table 3. Changes in affective variables

Construct	Pre M	Pre α	Post M	Post α	t	p	Interpretation
Speaking anxiety /5	3.42	0.85	2.81	0.82	-6.21	< .001	Anxiety decreased
Willingness to communicate /5	2.96	0.88	3.64	0.87	6.74	< .001	WTC increased
AI-speaking confidence /5	2.74	0.79	3.88	0.84	8.43	< .001	Confidence increased

Students reported significantly lower speaking anxiety after the intervention. They also reported higher willingness to communicate, especially in AI-mediated and small-group contexts. These results suggest that AI may have functioned as a low-pressure rehearsal space that helped students gradually build confidence before performing in front of teachers or peers.

4.3 AI Use and Engagement

AI-use logs indicated that students completed an average of 3.6 AI-speaking sessions per week. The average session lasted 18.4 minutes. Students who completed more than three AI-speaking sessions per week tended to show larger gains in fluency and coherence than students who completed fewer sessions. However, the relationship was not linear: several high-use students showed only moderate improvement because they relied heavily on AI-generated scripts instead of reformulating their own answers.

The adapted scales demonstrated acceptable to good internal consistency. The speaking anxiety questionnaire showed good reliability at both time points (pre-test Cronbach's α = .85; post-test α = .82). The willingness-to-communicate scale also exhibited strong internal consistency (pre-test α = .88; post-test α = .87). Finally, the AI-speaking confidence scale had acceptable to good reliability (pre-test α = .79; post-test α = .84). All values exceeded the recommended threshold of .70 (Nunnally & Bernstein, 1994), indicating that the instruments were suitable for measuring the intended constructs in the present sample.

Table 4: Illustrative AI-use patterns and speaking gains

AI-use group	n	Average sessions/week	Speaking gain /50	Common pattern
Low use	12	1.8	+2.30	Limited rehearsal, low journal completion
Moderate use	22	3.4	+4.90	Regular rehearsal and transcript review
High use	14	5.1	+6.10	Frequent practice, but some script dependence

This pattern suggests that AI use was beneficial when accompanied by reflective transcript review and teacher feedback. Mere frequency of AI use was insufficient to guarantee improvement.

4.4 Qualitative Findings

The qualitative analysis generated six major themes: AI as a low-stakes rehearsal partner, transcript-based noticing, lexical expansion and response organisation, over-reliance on AI-generated scripts, uneven feedback quality, and transfer difficulties.

Theme 1: AI as a Low-Stakes Rehearsal Partner

Students repeatedly described AI as less intimidating than human interlocutors. Many explained that they were willing to repeat answers multiple times because they did not feel judged. This finding helps explain the quantitative reduction in speaking anxiety. One student stated:

When I speak with AI, I can repeat many times. I do not feel embarrassed. With classmates, I am afraid they will laugh.

Another student commented:

I know AI is not a real person, so I feel safer. I can try difficult words first before speaking in class.

These responses suggest that AI created a psychologically safer rehearsal space. For anxious learners, this space allowed them to practise before entering more socially demanding speaking contexts.

Theme 2: Transcript-Based Noticing

Students reported that ASR transcripts made their speaking visible. Several students said they had not realised how often they repeated the same words until they read their transcripts. Others noticed unclear grammar, missing connectors, and incomplete sentences.

One student explained:

When I saw the transcript, I realised my answer was not clear. I repeated "very good" and "I think" many times. Then I asked AI for other expressions.

Another student noted:

The transcript showed that some words were not recognised. Maybe my pronunciation was not clear, so I tried again more slowly.

This theme supports the view that ASR can encourage noticing and self-monitoring (Liu et al., 2022; Shadiev & Liu, 2022; Sun, 2023). However, students also needed teacher guidance to interpret transcripts critically because ASR errors did not always indicate learner errors.

Theme 3: Lexical Expansion and Response Organisation

Students used AI to generate alternative vocabulary, discourse markers, and example structures. This was particularly useful for Aptis-style tasks requiring explanation, comparison, and opinion-giving. Students reported that AI helped them move beyond short answers and develop longer responses.

One student said:

Before, I only answered one sentence. AI asked me follow-up questions, so I learned to give reasons and examples.

Another explained:

I used AI to find better words, but the teacher told us not to copy everything. I changed the words to match my real ideas.

This finding suggests that AI can support response expansion when students use it as a scaffold rather than as a script generator.

Theme 4: Over-Reliance on AI-Generated Scripts

A major risk was script dependence. Some students initially asked AI to generate perfect Aptis answers and then attempted to memorise them. These answers often sounded fluent but unnatural. They also became ineffective when prompts changed.

One student admitted:

At first, I copied the AI answer because it sounded better than mine. But when the teacher changed the question, I could not remember or adapt it.

Teacher field notes confirmed this issue. Some students produced polished phrases but failed to answer the prompt directly. This finding reflects concerns in the wider literature that generative AI may encourage overdependence and reduce critical engagement if learners use it unreflectively (Kohnke et al., 2023; Lo et al., 2024; Meniado, 2023).

Theme 5: Uneven Feedback Quality

Students appreciated AI feedback because it was immediate, but they also found it inconsistent. AI sometimes praised responses that were too general or off-topic. It often provided broad comments such as "good answer" or "try to add more details" without explaining how the answer matched the Aptis-style criteria.

One student commented:

AI said my answer was excellent, but my teacher said I did not answer the question directly.

Another student said:

The AI feedback was fast, but sometimes I did not know what exactly to improve.

This theme shows that AI feedback should not be treated as equivalent to teacher assessment. Students benefited most when teacher feedback helped them interpret AI suggestions against task-specific criteria.

Theme 6: Transfer from AI Practice to Timed Test Performance

Although students became more confident during AI practice, some struggled to transfer rehearsed fluency to timed speaking conditions. They reported that AI practice felt more flexible, while Aptis-style performance required quick organisation and time management.

One student explained:

With AI, I can pause and think. In the test practice, the time is short, so I feel pressure again.

This finding suggests that AI rehearsal must be combined with timed re-performance. AI can prepare learners, but it cannot fully reproduce the pressure of recorded test conditions.

V. DISCUSSION

The findings suggest that AI-mediated speaking practice can support Aptis Speaking preparation when embedded in a structured pedagogical cycle. The quantitative results showed improvement in Aptis-style speaking scores, reduced speaking anxiety, and increased willingness to communicate. The qualitative findings explained these changes through repeated rehearsal, lower fear of judgement, transcript-based noticing, lexical scaffolding, and increased learner autonomy.

These findings are consistent with research showing that AI and chatbot-assisted language learning can improve learning outcomes and engagement (Chen et al., 2024; Fryer et al., 2023; Lo et al., 2024). They also align with studies on intelligent personal assistants and AI-speaking assistants, which suggest that AI can expand opportunities for out-of-class oral interaction and affective engagement (Tai, 2022; Tai & Chen, 2023; Zhang et al., 2024). In this study, AI appeared especially useful because it allowed students to practise more frequently than would be possible in teacher-led classroom time alone.

The reduction in speaking anxiety is also theoretically meaningful. Foreign language anxiety can inhibit participation and reduce learners' willingness to communicate (Horwitz et al., 1986; MacIntyre et al., 1998; Woodrow, 2006). AI-mediated rehearsal reduced social evaluation pressure because students could practise privately, repeat tasks, and make mistakes without immediate peer judgement. This supports the interpretation that AI may serve as an affective buffer between private practice and public performance.

However, the findings also challenge overly optimistic claims about AI. AI did not automatically produce communicative competence. Improvement depended on how students used AI and how the teacher structured the learning cycle. Students who used AI to rehearse, notice, reformulate, and re-perform benefited more than those who copied AI-generated model answers. This distinction is central. AI can support learning when it increases cognitive engagement; it can undermine learning when it replaces cognitive effort.

The findings also show that AI feedback has pedagogical limits. Students valued immediate feedback, but AI comments were sometimes generic or misaligned with Aptis-style task demands. This limitation echoes concerns that AI systems may not fully evaluate discourse relevance, pragmatic appropriateness, pronunciation intelligibility, or test-specific scoring criteria (Sun, 2023). Therefore, AI feedback should be triangulated with teacher feedback, peer feedback, and explicit rubrics.

For Aptis Speaking preparation, the most important implication is that AI should be positioned as a rehearsal partner rather than as a test examiner. Aptis Speaking requires timed, recorded, task-specific performance. AI can help students practise language resources, but teachers must ensure that learners also practise prompt interpretation, answer organisation, timing, and spontaneous reformulation. The proposed cycle–Aptis-style prompt, AI rehearsal, ASR transcript, learner noticing, self-reformulation, teacher feedback, and timed re-performance–offers one way to integrate AI critically and responsibly.

The study contributes to AI-in-language-education research in three ways. First, it connects AI-mediated speaking practice to a specific assessment context. Second, it triangulates performance, affective, and qualitative data. Third, it foregrounds the risks of over-reliance and feedback misalignment rather than presenting AI as a simple solution to speaking difficulties.

VI. CONCLUSION

6.1. Conclusion

This explanatory sequential mixed-methods case study suggests that AI-mediated speaking practice can support Vietnamese EFL students preparing for Aptis ESOL Speaking. The quantitative results indicated improvement in Aptis-style speaking performance, reduced speaking anxiety, and increased willingness to communicate. The qualitative findings suggested that AI supported learning by creating a low-stakes rehearsal space, increasing opportunities for output, enabling transcript-based noticing, expanding vocabulary, and encouraging learner autonomy.

At the same time, the study offers a critical conclusion: AI is not inherently pedagogical. Its educational value depends on task design, teacher mediation, learner reflection, and ethical use. In Aptis preparation, AI should not be used to generate memorised answers. Instead, it should be used as a scaffolded rehearsal partner that supports practice, noticing, reformulation, and timed performance.

Future research should collect real experimental or quasi-experimental data, use larger samples, include comparison groups, conduct delayed post-tests, and use externally validated speaking scores. Researchers should also document AI prompts, feedback procedures, learner usage patterns, and ethical safeguards in detail.

6.2. Pedagogical Implications

The findings suggest several implications for teachers and test-preparation programmes.

First, teachers should design AI tasks around clear speaking functions rather than simply telling students to "practise with AI." For Aptis preparation, useful functions include describing personal experiences, comparing two options, explaining reasons, developing examples, expressing preferences, and sustaining extended responses.

Second, teachers should require transcript-based reflection. Students should not only speak to AI but also review their transcripts, identify repeated language, notice unclear grammar, and revise their responses.

Third, AI feedback should be filtered through human judgement. Teachers should help students compare AI suggestions with Aptis-style criteria. This prevents students from accepting generic praise or inappropriate model answers.

Fourth, AI practice should be followed by timed re-performance. Without timed practice, students may become comfortable in AI conversations but remain unprepared for the pressure of recorded test conditions.

Fifth, responsible AI use should be explicitly taught. Students need to understand the difference between using AI for support and submitting AI-generated language as if it were their own spontaneous performance.

6.3. Limitations

The study is limited by its case-study design, reliance on one institutional context, and short intervention period. Future implementations should strengthen external validity by including a comparison group, a larger sample, delayed post-tests, multiple institutions, and independently verified rating procedures.

Even in a real implementation, several limitations would remain. An intact-class case study limits generalisability. The absence of a control group would make it difficult to attribute gains solely to AI. Self-report questionnaires may be affected by social desirability. AI-use logs may not capture the quality of learner engagement. Finally, Aptis-style scores are not identical to official Aptis scores unless assessed through official procedures.

6.4. Recommendations for Future Research

Future studies should use quasi-experimental or experimental designs comparing AI-supported Aptis preparation with teacher-only or peer-only speaking practice. Larger samples would allow researchers to examine moderating variables such as proficiency level, anxiety level, gender, AI literacy, and frequency of practice. Delayed post-tests would help determine whether speaking gains are retained. Researchers should also examine the quality of AI prompts and feedback, because different prompt designs may lead to different learning outcomes.

Longitudinal qualitative research is also needed. Such research could investigate how learners' AI-use strategies evolve over time, how they move from dependence to autonomy, and how teachers develop AI-mediated speaking pedagogy. Finally, future research should examine ethical and equity issues, including data privacy, access to paid AI tools, bias in speech recognition, and the risk of widening inequalities between students with different levels of technological access.

REFERENCES

1. Ali, J. K. M., Shamsan, M. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks. *Languages*, *8*(3), 212. https://doi.org/10.3390/languages8030212
2. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
3. British Council. (2024a). *Aptis ESOL General*.
4. British Council. (2024b). *Aptis Speaking test: Practice and video*.
5. British Council. (2024c). *Aptis scoring system*.
6. Bygate, M. (2001). Speaking. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 14–20). Cambridge University Press.
7. Chen, K. T. C. (2022). Speech-to-text recognition in university EFL learning. *Education and Information Technologies*, *27*, 9857–9875.
8. Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002
9. Chen, Y., Wang, Y., & Liu, C. (2024). The effectiveness of artificial intelligence on English language learning achievement: A meta-analysis. *System*, *125*, 103428. https://doi.org/10.1016/j.system.2024.103428
10. Cislowska, K., & Pena-Acuna, B. (2024). Learning English as a second language with artificial intelligence. *Frontiers in Education*, *9*, 1490067. https://doi.org/10.3389/feduc.2024.1490067
11. Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.
12. Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
13. Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.
14. Evers, K., & Chen, S. (2022). Effects of automatic speech recognition software on pronunciation for adults with different learning styles. *Journal of Educational Computing Research*, *60*(3), 669–685.
15. Fryer, L. K., Nakao, K., & Thompson, A. (2023). Chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies*, *28*, 11251–11275.
16. Goh, C. C. M., & Burns, A. (2012). *Teaching speaking: A holistic approach*. Cambridge University Press.
17. Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x
18. Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537–550.

19. Liu, J., Liu, X., & Yang, C. (2022). A study of college students' perceptions of utilizing automatic speech recognition technology to assist English oral proficiency. *Frontiers in Psychology*, *13*, 1049139. https://doi.org/10.3389/fpsyg.2022.1049139
20. Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, *11*, 50.
21. Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
22. MacIntyre, P. D., Clement, R., Dornyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2. *The Modern Language Journal*, *82*(4), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb01286.x
23. Meniado, J. C. (2023). The impact of ChatGPT on English language teaching, learning, and assessment: A rapid review of literature. *Arab World English Journal*, *14*(4), 3–18.
24. Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129
25. Shadiev, R., & Liu, J. (2022). Review of research on applications of speech recognition technology to assist language learning. *ReCALL*, *34*(1), 1–16.
26. Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed-methods investigation. *Frontiers in Psychology*, *14*, 1210187. https://doi.org/10.3389/fpsyg.2023.1210187
27. Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum.
28. Tai, T. Y. (2022). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*, *37*(5), 1054–1077.
29. Tai, T. Y., & Chen, H. H. J. (2023). Comparing the effects of intelligent personal assistant-human and human-human interactions on EFL learners' willingness to communicate. *Computers & Education*, *201*, 104845. https://doi.org/10.1016/j.compedu.2023.104845
30. Thanh, V. T. (2026). The relationship between procrastination behavior and students' English learning outcomes. *Journal of Psychology and Education*, *32*(3), 295–299.
31. Vo .T. T (2026). Theoretical foundations of factors influencing grammatical accuracy in university students' English writing in Ho Chi Minh City. *Journal of Education and Society*, *32*(2), 87–97.
32. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
33. Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal*, *37*(3), 308–328. https://doi.org/10.1177/0033688206071315
34. Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
35. Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.
36. Zhang, C., Meng, Y., & Ma, X. (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System*, *121*, 103259. https://doi.org/10.1016/j.system.2023.103259