



AI-Scaffolded English Oral Performance and Learner Perceptions: A Comparative Case Study of Thai and Chinese University Students

Monthon Kanokpermpoon^a, Thayakorn Klomkaew^{b,*}, Yuwei Liang^c

^a monthon.k@litu.tu.ac.th, Language Institute, Thammasat University, Thailand
^b thayakorn.klom@gmail.com, English Program, Faculty of Humanities and Social Sciences, Kamphaeng Phet Rajabhat University, Thailand
^c liangyuwei@bjut.edu.cn, Beijing University of Technology, P. R. China
* Corresponding Author, thayakorn.klom@gmail.com

APA Citation: Kanokpermpoon, M., Klomkaew, T., & Liang, Y. (2026). AI-scaffolded English oral performance and learner perceptions: A comparative case study of Thai and Chinese university students. <i>LEARN Journal: Language Education and Acquisition Research Network</i> , 19(2), 1040-1072. https://doi.org/10.70730/DDOF5466	
Received 19/01/2026 Received in revised form 24/06/2026 Accepted 30/06/2026	ABSTRACT This study explored how AI scaffolding helped support English oral communication among Thai and Chinese EFL university students and investigated how these students perceived the usefulness and ease of use of AI tools. Using a comparative mixed-methods analysis, the study purposively selected ten first-year undergraduate students (five Thai and five Chinese participants). The students completed three phases of scaffolded speaking tasks using ChatGPT (for Thai participants) or Doubao (for Chinese participants), along with an extended Technology Acceptance Model (TAM) questionnaire. Speaking performances were analyzed using the Complexity, Accuracy, and Fluency Framework and an Analytic Scoring Rubric, while questionnaire data were examined through descriptive statistics and thematic analysis. The findings revealed that students from both groups improved their vocabulary and idea organization, particularly

	during structured question-and-answer interactions. By contrast, grammatical accuracy showed little improvement across the three phases. Overall, both Thai and Chinese students viewed AI tools positively and considered them useful and easy to use, although cultural differences influenced what they valued most in the learning experience. The study suggests that generative AI works best as a supportive speaking partner that complements, rather than replaces, human language teachers. Keywords: AI-scaffolding, oral performance, TAM perceptions, Thai students, Chinese students
--	---

Introduction

AI-powered chatbots such as ChatGPT and Doubao have become tools for language learning in addition to everyday work support. These chatbots allow learners to practice language skills in low-anxiety, on-demand environments. Unlike traditional computer-assisted language learning programs, chatbots can sustain interactive exchanges, provide instant feedback, and adjust to learner input (Fryer & Carpenter, 2006; Huang et al., 2022). These chatbot tools are particularly relevant to oral language learning due to their sustained interaction and timely feedback, which make them suitable for large university classrooms. However, the successful implementation of AI tools depends not only on technological capability but also on user acceptance. The Technology Acceptance Model (TAM) posits that perceived usefulness and perceived ease of use determine individuals' intentions to adopt new technologies (Davis et al., 1989). Subsequent research has extended TAM to include factors such as enjoyment, social influence, and technological familiarity (Chesney, 2006; Granić, 2023). In the contexts of language education, learners' perceptions significantly influence whether digital tools are meaningfully adopted and sustained (Huang et al., 2022). Therefore, investigating students' perceptions toward generative AI is crucial in comprehending whether these tools can be effectively integrated into oral language teaching in EFL contexts.

AI adoption in university classes has expanded rapidly across Asia, particularly in China and Thailand. In China, with national digital education strategies, AI in university teaching and grading has been implemented on a large scale (Zou et al., 2022). Similarly, Thailand has promoted the digitalization of higher education and employed AI-supported platforms to enhance EFL learners' English language proficiency (Taylor et al., 2023). Although research on AI-assisted language learning has gained significant

attention, studies specifically focusing on AI-enhanced speaking performance in Asian higher education remain scant (Shaengchart et al., 2023; Taylor et al., 2023). To better understand the dimensions of oral performances using AI scaffolding, this study examines four performance dimensions (fluency, grammatical complexity, vocabulary, and idea organization) alongside constructs from the extended Technology Acceptance Model (TAM) (perceived ease of use, perceived usefulness, attitude & enjoyment, output quality & result demonstrability, and social influence & facilitating conditions). Grounded in AI scaffolding for speaking and technology acceptance frameworks, the present study investigates Thai and Chinese university students' speaking interactions and their perceptions of AI's usefulness and ease of use in higher education. By comparing two higher education contexts in Thailand and China, this research seeks to answer the following research questions:

RQ1: To what extent does the use of generative AI as a speaking partner improve Thai and Chinese students' oral communication proficiency?

RQ2: How do Thai and Chinese students perceive the usefulness and ease of use of generative AI for English speaking practice?

Review of Literature

In this section, speaking interaction and AI use in scaffolding are discussed. This is followed by the expanded model of TAM for AI perceptions as well as empirical past studies.

Speaking Interaction and Artificial Intelligence (AI) through Scaffolding Technique

Oral communication continues to be a major challenge for EFL learners in higher education, particularly in contexts where instruction prioritizes grammar and examination rather than interactive language use (Charoenchotphokin, 2024). Many students experience speaking anxiety because they have limited exposure to real interaction and low confidence when participating in classroom discussions (Boonkit, 2010; Ork et al., 2024). These factors create a need for alternatives that provide supportive, structured opportunities for oral practice. AI, especially conversational tools such as ChatGPT and Doubao, can provide alternative platforms for collaborative speaking practice (Charoenchotphokin, 2024). By generating immediate dialogues and instant feedback, AI systems allow learners to practice on their own. However, the effectiveness of such tools depends on how they are pedagogically designed by teachers. This aligns with the interaction of the two proposed factors of TAM by Davis et al. (1989), which

are perceived usefulness and perceived ease of use, regarding individuals' intentions to adopt new technologies. In addition, the success of these two factors depends on the interaction between students and AI tools.

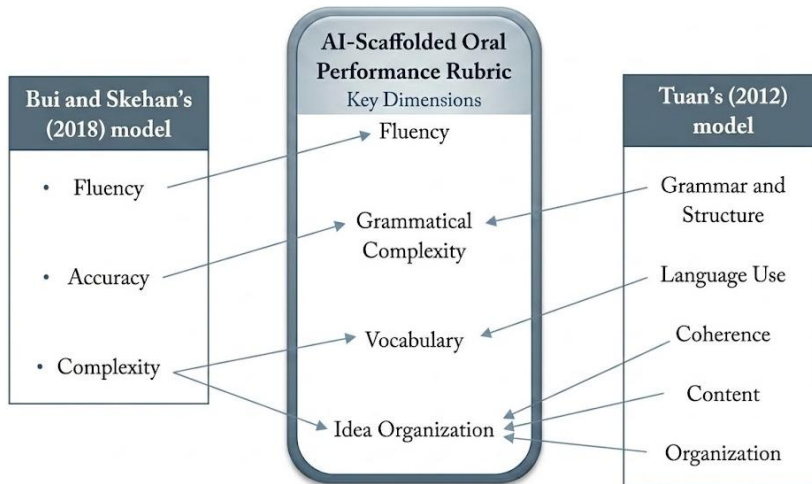
In this study, AI-supported oral performance was informed by the scaffolding approach (Wood et al., 1976), which considers learning as a gradual process supported by temporary guidance that is reduced as learners gain independence. By providing sentence models and vocabulary lists and gradually reducing prompts, ChatGPT operated as a digital scaffold within learners' Zone of Proximal Development (Van de Pol et al., 2010).

Charoenchotphokin's (2024) findings reveal that AI-scaffolded interaction supported learners' extended oral responses. Learners moved from depending heavily on high support for sentence models to producing longer and more comprehensible sentences with greater autonomy. This shows the concept of graduated support (Hammond & Gibbons, 2005) and is in line with Communicative Language Teaching practices, which support meaningful and task-based communication (Richards & Rodgers, 2001). Apart from linguistic improvement, learners showed a high degree of confidence and reduced anxiety. This suggests that AI-scaffolded learning can support both the cognitive and affective dimensions of speaking development. Despite some technical issues, AI played an effective role as a supplementary speaking partner rather than a replacement for teachers' classroom instruction. As such, AI-mediated scaffolding can increase opportunities for oral practice and create a constructive environment that bolsters learners to take communicative risks in structured tasks.

In examining EFL oral performances, this study employs Bui and Skehan's (2018) Complexity, Accuracy, and Fluency (CAF) approach as well as Tuan's (2012) Analytic Scoring Rubric for Speaking as the main analytical frameworks for assessing Thai and Chinese learners' oral performances. These two selected frameworks propose that language performance can be improved through three interrelated dimensions: complexity, accuracy, and fluency. Moreover, these dimensions help develop learners' interlanguage system and equip them with a better understanding of language production. Tuan's (2012) rubric, in particular, was developed in response to the limitations of holistic scoring methods. One of its major strengths is its diagnostic transparency, as each construct is scored independently (Bulut, 2022). This allows teachers and researchers to identify specific strengths and weaknesses more effectively rather than relying solely on a single overall impression. However, in speaking, the rubric only measures coherence, content, grammar and structure, language use, and organization, but it lacks fluency as a key speaking feature (Skehan, 2011). In the present study, the combined frameworks of Bui and Skehan (2018) and Tuan (2012) were used to measure the following key features of AI-scaffolded oral performances.

Figure 1

*Combined framework for AI-scaffolded oral performance rubric (Adapted from Bui & Skehan, 2018; Tuan, 2012) **



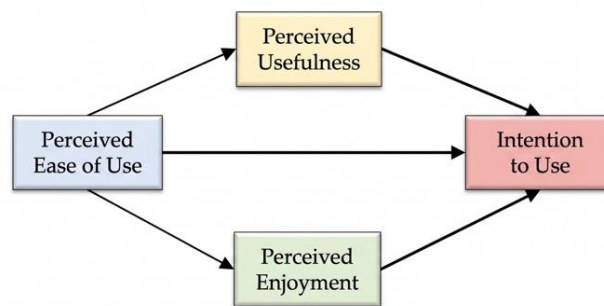
Technology Acceptance Model in Oral Communication

The Technology Acceptance Model (TAM) explains technology adoption through two key concepts: perceived usefulness (PU) and perceived ease of use (PEOU). Davis (1989) defined the former as the extent to which a system enhances performance and the latter as the degree to which it requires minimal effort. In technology-enhanced learning, usefulness refers to improvements in speaking fluency, pronunciation, confidence, and communicative effectiveness. Learners are unlikely to continue using speaking technologies such as AI conversational tools unless they perceive clear gains in their oral competence (Davis et al., 1989; Venkatesh & Davis, 2000). It can be argued that perceived usefulness drives a sustained adoption. Perceived ease of use is also important, even though often indirectly (Davis, 1989). Ease of use expands beyond technical simplicity since it includes smooth interaction, understandable feedback, and manageable cognitive load in speaking-focused technologies. For instance, if learners find the system complicated or unnatural, they may struggle to focus on communication itself. This results in a decline in both ease and usefulness perceptions (Davis, 1989; Davis et al., 1989). Ease of use, in contexts where speaking anxiety is common, may also reduce psychological barriers to participation (Venkatesh & Davis, 2000). In higher education settings, particularly in English language teaching (ELT), past studies (e.g., Chesney, 2006; Taylor et al., 2023) have extended TAM with a wider range of factors that contribute to technology

acceptance. Apart from perceived usefulness and ease of use, other factors such as self-efficacy, subjective norms, perceived enjoyment, social influence, system quality, accessibility, anxiety, and facilitating conditions have been identified (Taylor et al., 2023). These factors could affect users' decisions to adopt technology (Chesney, 2006; Granić, 2023). Among these newly proposed factors, perceived enjoyment has gained considerable attention in learning contexts (Chesney, 2006). In this factor, intrinsic motivation plays a crucial role in maintaining engagement (Davis et al., 1992; Van der Heijden, 2004). Unlike workplace technologies that are primarily task-oriented and productivity-driven, educational technologies often involve emotional and experiential elements that enhance learners' willingness to continue using them. Hence, Figure 2 below illustrates the extended TAM model.

Figure 2

*The Extended Technology Acceptance Model (TAM) (Adapted from Chesney, 2006)**



Previous studies

Past studies on scaffolding and AI-supported oral performance provide useful observations for the current study. According to Kim and Hwang (2026), Korean EFL university students used more complex vocabulary and sentence structures when completing expository speaking tasks than narrative tasks. Nonetheless, participants showed weak grammatical accuracy, poor pronunciation, and lower fluency. These findings support the Limited Attentional Capacity Model, which asserts that learners have difficulty handling several language demands at the same time during speaking tasks. Their findings also highlight an important issue for AI-mediated speaking tasks: the immediate pressure of oral communication can increase cognitive load.

With reference to Van de Pol et al. (2010), effective scaffolding comprises three important components: contingency, fading, and transfer of

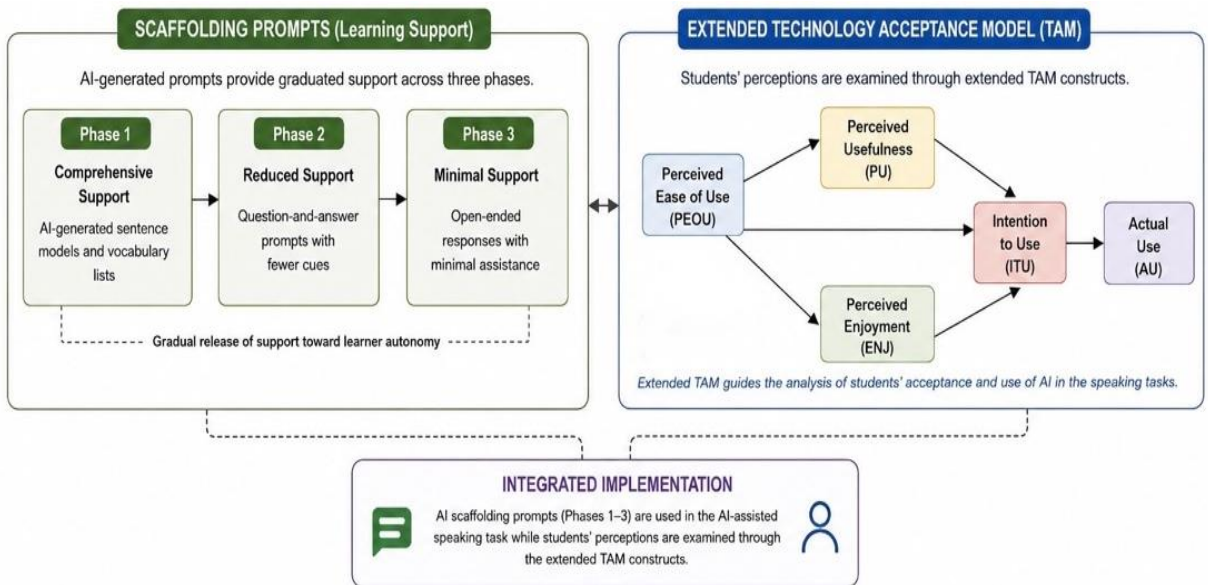
responsibility. That is, support should match learners' needs, gradually reduce over time, and ultimately help learners become more independent. This framework is in line with AI-assisted learning since AI tools can supply temporary help that guides learners within their Zone of Proximal Development (Charoenchotphokin, 2024). For AI-supported oral practice, Chen et al. (2025) found that using ChatGPT for eight weeks could assist EFL learners in improving their oral proficiency regardless of their language level. The study also revealed that learners became less anxious about speaking and more willing to experiment with language. Similarly, Tam's (2025) study found that AI applications such as ELSA Speak and Duolingo enhanced pronunciation, fluency, and confidence. However, some learners still experienced barriers such as robotic communication and speech-recognition errors, which imply that AI tools are more effective as supplementary material rather than replacements for human teachers.

The recent application of TAM to AI-supported speaking tools in ELT raises important questions about technology adoption in oral communication contexts. Even though TAM traditionally places perceived usefulness as the dominant predictor of behavioral intention, emerging evidence suggests that this assumption may not hold consistently in AI-mediated speaking circumstances. For example, Shaengchart et al. (2023) investigated ChatGPT adoption among university students in Bangkok and discovered that neither perceived usefulness nor perceived ease of use significantly predicted actual usage. Instead, facilitating conditions and learner perceptions were stronger determinants. This finding directly challenges TAM's primary claim that cognitive evaluations of usefulness drive adoption decisions. In speaking-focused AI contexts, structural and contextual support may override instrumental judgments, implying that the internal belief structure of TAM may not fully explain usage behavior. In a similar vein, Chen et al. (2025) found that GenAI-supported speaking practice resulted in measurable improvements in oral proficiency, together with reduced foreign language speaking anxiety and high linguistic risk-taking. These findings reveal that learners' perceptions of usefulness may extend beyond task performance enhancement to include emotional regulation and confidence development. However, because TAM frames usefulness in terms of productivity, it may not fully capture the social and emotional aspects important to oral communication development. Moreover, evidence from Tam (2025) in Vietnamese EFL contexts further challenges the TAM framework. Although learners improved pronunciation, fluency, and confidence through AI-supported speaking tools, they expressed concerns about robotic AI responses and limited conversational authenticity. These findings suggest that perceived ease of use cannot be solely determined by technical simplicity. Instead, authentic interaction and communication may

influence sustained engagement in ways not explicitly theorized within the original TAM framework. Therefore, this study aims to bridge the gap by exploring Thai and Chinese university students' performances and perceptions of the use of generative AI in oral language learning. The conceptual framework of scaffolding and extended TAM for this study is as follows.

Figure 3

*The Conceptual Framework of Scaffolding and Extended TAM***



Research Methodology

The following section presents the research context and participants, research instruments, data collection, and data analysis.

Research Context and Participants

The aims of the study were to investigate Thai and Chinese students' oral performances and to explore perceptions of the use of generative AI for oral communication. This study is positioned as a comparative case study, in which in-depth data were prioritized over statistical generalizability. The goal was not to produce population-level conclusions but to generate rich and exploratory insights into how AI scaffolding operates across two distinct EFL settings. Accordingly, purposive sampling was employed to select participants

whose contexts were representative of the phenomena under investigation. The study was intended to document and compare the nature of AI-scaffolded speaking development within each context rather than to establish causal claims about AI's effectiveness relative to unaided instruction. These design choices are inherent to the comparative case study methodology and should be taken into account when interpreting the findings. The research context of this study was situated in university-level EFL programs in Thailand and China. Participants were purposively recruited based on the following criteria: (1) they were first-year undergraduate students enrolled in a required EFL course, (2) they were non-English majors with no prior formal training in English as a specialist subject, (3) they had access to the designated AI platform used in their respective context (ChatGPT for Thai participants; Doubao for Chinese participants), and (4) they voluntarily agreed to participate following an informed briefing on the study's objectives and procedures. These criteria ensured that participants were comparable across the two contexts in terms of academic level, language learning background, and exposure to AI-assisted learning.

In the Thai context, the study involved first-year undergraduate students enrolled in a Fundamental English course. They were between 18 and 19 years old and attended a government university in northern Thailand. For the Chinese context, participants were first-year undergraduate students enrolled at a national "211 Project" university in Beijing, P. R. China. All participants were non-English majors and were taking university-level EFL courses as part of their first-year curriculum. In total, 10 volunteers agreed to participate after being informed about the AI-mediated speaking practice.

Participants in both contexts were developing their academic English, particularly speaking. In both contexts, English teaching focuses on grammar accuracy through formative and summative assessments. This limits space for interactive or extended oral practice in authentic environments (Boonkit, 2010; Ork et al., 2024). As a result, students start their university learning with less confidence and less experience in immediate oral communication. Even though institutions in both countries are adopting AI-based tools, differences in educational culture, technological readiness, and teaching approaches may influence how students adopt these technologies for oral learning. By focusing on first-year Thai and Chinese EFL students, the roles of generative AI in this study are positioned in two comparable yet culturally distinct academic contexts. The comparative approach allows an exploration of how students' oral performances and their perceptions of AI-supported oral learning are shaped by their respective educational and sociocultural contexts.

Research Instruments

Oral performance tasks were designed following the concept of scaffolding. The tasks consisted of providing temporary support that gradually decreases as learners gain competence (Wood et al., 1976). This approach is in line with the concept of graduated assistance, where support adapts to the learner's current ability (Van de Pol et al., 2010), and is closely related to Vygotsky's (1978) Zone of Proximal Development. Each speaking topic followed three phases. Phase one involved comprehensive support through sentence models and vocabulary lists generated by AI. Phase two then provided fewer cues with question-and-answer prompts, and phase three involved open-ended responses with minimal assistance to foster learner autonomy. All participants completed the same topic, "Travel Recommendation," to ensure consistency in the data collection. To ensure scoring consistency, all recorded speaking performances were independently assessed by two trained raters: the researchers who were experienced in EFL speaking assessment. Both raters applied the combined analytical framework adapted from Bui and Skehan (2018) and Tuan (2012), evaluating each performance across four dimensions: fluency, vocabulary range, grammatical accuracy, and idea organization, each scored on a 1–5 scale based on clearly defined descriptors as shown in Table 1. In cases of disagreement between the two raters, the scores were discussed until consensus was reached.

Data Collection

This study employed a comparative mixed-methods design to examine the oral performances of ten Thai (N=5) and Chinese (N=5) first-year EFL students and to explore their perceptions regarding the use of generative AI for oral language learning. This research project received ethical approval from one of the researchers' university research ethics committees (Approval No. 009/68).

Prior to the data collection process, the research objectives and instruments were explained to the participants. Additionally, the consent form was explained to all participants, and they were then asked to sign it to confirm their agreement to participate in all data collection procedures. They were also allowed to withdraw from the study at any time.

For data collection, two primary research instruments were used. The first instrument consisted of three scaffolded speaking tasks conducted through generative AI platforms (ChatGPT for Thai EFL students and Doubao for Chinese EFL students). The tasks were structured in three phases: high support, medium support (guided elaboration), and low support (independent performance), to promote increasingly autonomous oral

production while ensuring procedural consistency across the two contexts. All participants were trained to follow the scaffolding process and were pre-tested using a general speaking topic. One limitation worth addressing is that the two groups used different AI platforms. This difference was not a random choice but was based on practical and ethical reasons: ChatGPT is the most widely used and institutionally supported AI tool in Thailand, while Doubao serves the same role in China, where ChatGPT is blocked. Requesting Chinese participants to use ChatGPT would have been both unfair and unrealistic given these access restrictions. It is also important to clarify how the AI tools were used in this study. Neither ChatGPT nor Doubao was used to evaluate or score the learners' speaking performance. Both platforms served only as chatbot tools to provide speaking support through structured prompts. All speaking assessments were carried out by human raters using a standard scoring rubric. To reduce the impact of platform differences on the results, the same prompts were given to both groups at every stage of the task. By keeping the prompts consistent, the study aimed to ensure that both groups received essentially the same type of scaffolding support, even though they used different platforms. The use of different platforms should therefore be considered as a practical constraint shaped by the research context, not as a flaw in the study design. Any differences found between the two groups in speaking performance should be interpreted with this in mind. The second instrument was a structured questionnaire adapted from studies by Chen et al. (2025); Chesney (2006); Davis (1989); Davis et al. (1989); Shaengchart et al. (2023); Tam (2025); Taylor et al. (2023); and Venkatesh and Davis (2000) and grounded in TAM and its extensions. The instrument assessed constructs such as perceived usefulness (8 items), perceived ease of use (5 items), attitude and perceived enjoyment (5 items), output quality and result demonstrability (4 items), social influence and facilitating conditions (4 items), and behavioral intention (3 items). It included five-point Likert-scale items as well as three open-ended questions to capture both quantitative evaluations and qualitative reflections from the participants. All participants were asked to complete the questionnaire after completing the AI-scaffolded tasks. Before administration, the questionnaire was tested for content validity by two experts in English Language Teaching (ELT). Each expert independently evaluated the relevance and appropriateness of each item to its corresponding construct using the Index of Item-Objective Congruence (IOC). Items with IOC values below 0.50 were revised or removed in line with suggestions. Following expert review, the questionnaire was piloted with a small group of EFL students not included in the main study to identify any ambiguous wording or technical difficulties prior to full administration. Internal consistency reliability was then assessed using Cronbach's Alpha, which yielded a coefficient of .925, indicating excellent reliability.

Data Analysis

Data analysis was aligned with the study's research design and extended Technology Acceptance Model (TAM) framework. To answer the first research question, scaffolded speaking transcripts were analyzed using thematic analysis (Braun & Clarke, 2006). Inductive coding was applied to identify recurring themes from scaffolded speaking transcripts. Speaking data were examined for evidence of fluency, grammatical complexity, vocabulary use, and idea organization across the three scaffolding phases. These four dimensions were operationalized using criteria adapted from Bui and Skehan (2018) and Tuan (2012). In addition, comparative coding was used to reveal the similarities and differences between Thai and Chinese participants. To strengthen credibility, member checking was conducted by sharing thematic summaries with selected participants to confirm interpretive accuracy. The detailed analysis of the whole dataset was summarized in Table 1.

Table 1

Dimensional Mapping for Data Analysis Adapted from Bui and Skehan (2018) and Tuan (2012)

Dimensional Mapping for Data Analysis			
Analytical Category	Bui and Skehan (2018) Dimension	Tuan (2012) Dimension	Combined Operationalization
Hesitation/Fluency	Fluency - theoretical basis; explains processing load and automaticity	N/A	Pause-per-word ratio; filled pause count; location of pauses; scored 1-5
Vocabulary Range	Complexity (lexical)- type-token ratio; domain-specific word use	Vocabulary (20 pts)- range, precision, register appropriateness	TTR; presence of specialized lexis; absence of vague hedges; scored 1-5
Grammatical Accuracy	Accuracy - error-free clause ratio; error density	Language Use (25 pts)- agreement, tense, word order	EFC ratio; error type taxonomy; scored 1-5 mapped to Tuan's descriptors
Idea Organization	Complexity - discourse-level elaboration and structural ambition	Organization (20 pts)- sequencing, cohesion, opening/closing	Rhetorical structure; discourse markers; detail specificity; scored 1-5
	Complexity - idea elaboration as linguistic ambition	Content (30 pts)- relevance, depth,	Number of supporting points; specificity of

Dimensional Mapping for Data Analysis			
Analytical Category	Bui and Skehan (2018) Dimension	Tuan (2012) Dimension	Combined Operationalization
		substantive development	evidence; elaboration depth

To answer the second research question, descriptive statistics (e.g., means and standard deviations) from the questionnaire were calculated to identify overall perception patterns among Thai and Chinese students across the key constructs of perceived usefulness, perceived ease of use, attitude and perceived enjoyment, output quality and result demonstrability, social influence and facilitating conditions, and behavioral intention. The five-point Likert scale was interpreted as follows: totally agree (4.20-5.00), agree (3.40-4.19), neutral (2.60-3.39), disagree (1.80-2.59), and totally disagree (1.00-1.79). Qualitative data from the open-ended questions were conceptually coded into the themes of AI benefits, challenges, and human replacement. The coding was rechecked by two experts in applied linguistics, and the inter-rater agreement reached 90%. Finally, the quantitative and qualitative findings were triangulated to provide a comprehensive understanding of students' performance and perceptions of AI-scaffolded oral language learning across the two contexts (Denzin, 2006).

Findings

This section reports the findings, which are divided into two main parts. The first part presents the English oral performances of Thai and Chinese EFL students, which were scaffolded using AI tools, to answer the first research question. The second part demonstrates the findings collected from the extended TAM questionnaire to answer the second research question.

AI-Scaffolded Oral Performance of Thai and Chinese EFL Students

This part reports a comparative analysis of the oral performances of ten Thai and Chinese EFL learners (five from each group) on a "travel recommendation task." For the five Thai learners, the self-reported CEFR levels from Student 1 to Student 5 were A2, B1, B1, A2, and A1, respectively. For the five Chinese learners, their CEFR levels were B1, B2, C1, B2, and A2, respectively. Each learner's oral performance was assessed across four dimensions: fluency, vocabulary range, grammatical complexity, and idea organization based on Bui and Skehan's (2018) Complexity-Accuracy-Fluency (CAF) model and Tuan's (2012) Analytic Scoring Rubric for Speaking.

High AI Scaffolding Environment (Phase 1)

Figure 4

*Pre-Test vs. Phase 1 Oral Performance of Thai EFL Students by Dimension****

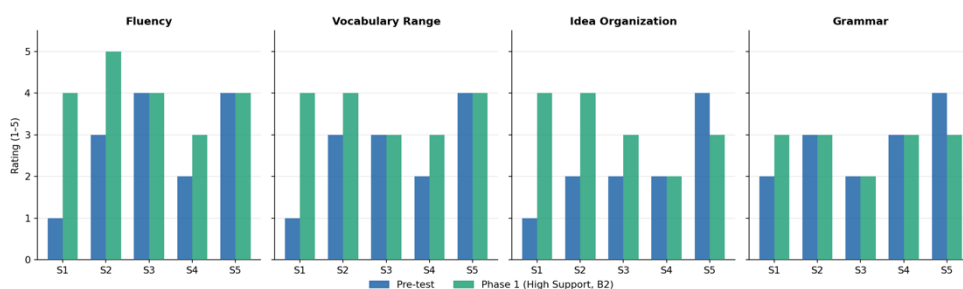
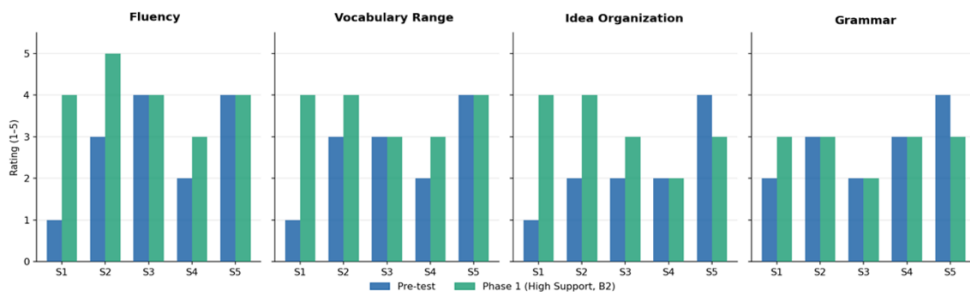


Figure 4 illustrates that high AI support with sentence models and vocabulary lists helped improve learners' oral performance. Responses became more comprehensive, well-composed, and employed a wider vocabulary range, with learners merging cultural and factual information more effectively. Nonetheless, personal patterns varied considerably. *Students 1, 2, 3, and 5* engaged naturally with the AI input, incorporating it meaningfully into their own oral responses. The persistence of spoken disfluency markers in *Student 1's* Phase 1 output ("Um," "ah" "Umm," and "uh") paradoxically shows that the elaborated information was being actively processed and orally produced rather than read from a script. *Student 4* showed an opposite pattern, with incomplete, error-laden prompts, e.g., "What's make this place special," "We can do activity, crying, swimming," where the language was supplied entirely by the AI and confirmed with minimal engagement ("yeb"). This reflects a key design challenge in which high support conditions must be structured to encourage independent thinking, not passive reliance on AI-generated text. Regarding grammatical accuracy, participants showed no improvement. For instance, *Student 3* generated rich descriptions in Phase 1 but maintained the same non-standard structures from pre-test: "It's have," "there have the cool cafe," and "you can like explore." *Student 5*, with the most grammatically accurate performance at pre-test, produced new non-standard constructions in Phase 1 ("It's have a mountain" and "If you love to relaxing"), possibly because of the more spontaneous production conditions.

Figure 5

*Pre-Test vs. Phase 1 Oral Performance of Chinese EFL Students by Dimension****

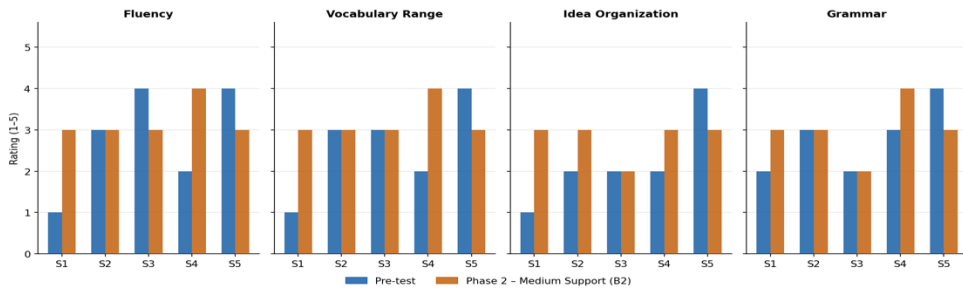


High AI support in Figure 5 reveals strongly facilitative effects among five EFL Chinese learners. Fluency and idea organization showed the most consistent developments. Three patterns are particularly evident. First, the most considerable developments were found among learners with the weakest pre-test. *Student 1's* oral performance increased across three criteria from 1 to 4 in terms of fluency, vocabulary range, and idea organization. This means that B2-level AI-scaffolding functions as an effective interactional resource. The integration of culturally specific references and evaluative framing, e.g., “*I strongly recommend you to pay a visit to Chengdu, one of the most charming cities in southwest China,*” illustrates how AI-generated input can widen the communicative capacity of lower-proficiency learners. Second, *Student 2* reveals an important finding. Although the student improved idea organization and vocabulary range, the evaluation and personal voice evident in the pre-test, such as “*What I love most about Chengdu is how it mixes traditional culture with modern life,*” became less clear under high AI support conditions. For higher-proficiency learners, extensive model language may restrain rather than bolster independent rhetorical expression. Third, grammatical accuracy did not show improvement across learners. To illustrate, *Student 5's* grammar scores decreased from 4 to 3. This pattern illustrates that content-related and lexical scaffolding does not necessarily transfer to morphosyntactic improvement. Even when learner outputs become more developed and structurally organized, non-standard grammatical forms continue to appear.

Medium AI Scaffolding Environment (Phase 2)

Figure 6

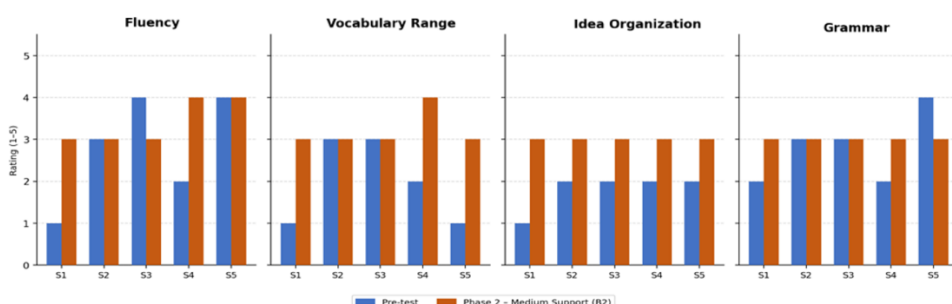
*Pre-Test vs. Phase 2 Oral Performance of Thai EFL Students by Dimension****



In Figure 6, the most noticeable finding was on *Student 4*, whose pre-test demonstrated substantial reproduction of AI-generated text with less personal contribution. The structured Q&A format required active language formulation, producing this student's most grammatically accurate and organizationally coherent performance, with gains across all measured criteria. Next, *Student 1* similarly yielded meaningful improvements across all four criteria, with a shift from a single non-standard clause to a multi-point, coherently organized suggestion. This shows that Q&A AI-scaffolding successfully elicited content knowledge not evident at the pre-test. *Student 2* made steady progress, consistently covering the topic. However, the transactional format seemed to limit the depth of their rhetorical analysis and the range of vocabulary used in the pre-test. *Student 3*'s grammatical challenges remained the most consistent pattern, with non-standard constructions such as "I enjoy to go," "it's very bit the different," and "we not speak in the same language much." This issue persisted across both support conditions despite authentic content knowledge and cultural engagement. *Student 5*'s shift from a formally structured monologue to a personal anecdote reveals that when AI support takes the form of open-ended questions rather than model language, higher-proficiency learners may default to informal production over the audience-directed register appropriate to the genre. This indicates that even stronger learners may benefit from register reminders when direct linguistic modelling is reduced.

Figure 7

*Pre-Test vs. Phase 2 Oral Performance of Chinese EFL Students by Dimension****

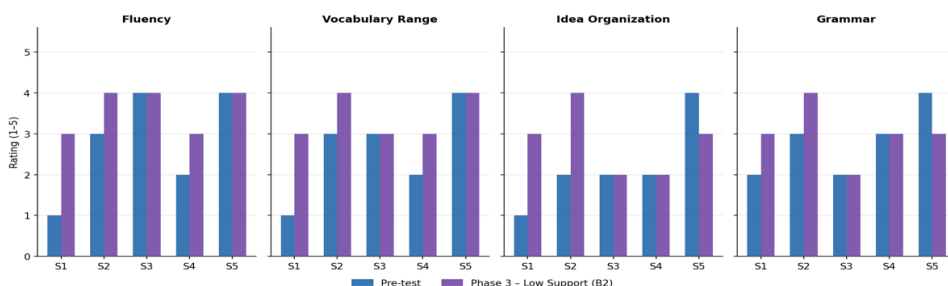


In Figure 7, three patterns among the Chinese EFL learners were observed. First, idea organization showed the most consistent gains, with all five learners improving to a score of 3 in Phase 2. This pattern reflects the organizational scaffolding embedded in the Q&A AI interaction format, which requires learners to produce topically focused and sequential responses. In contrast to the pre-test, the structured interaction appeared to activate organizational capacities. Second, the most distinctive finding was found in *Student 4*, whose pre-test showed minimal independent production and limited development across all criteria. Because the Q&A format required active language formulation without the support of model language, the student achieved the highest fluency and vocabulary range scores (4) recorded across all conditions. The ability to organize budget advice (*“For a student with 500 RMB a day, here’s a way to split”*), providing seasonal guidance and culturally embedded recommendations, reveals a substantial difference from the pre-test. Third, *Student 3* reveals an important finding regarding content breadth. Although idea organization improved in Phase 2, fluency scores dropped slightly from 4 to 3, indicating that the Q&A format may constrain the extended topical development achievable in self-organized monologues. For learners with higher pre-test content scores, medium-support conditions may therefore trade content breadth for organizational coherence. Finally, grammatical complexity remained largely stable across learners at the score of 3, while *Student 5* declined slightly from 4 to 3. Consistent with Phase 1, neither high nor medium AI support produced sustained morphosyntactic improvement. This confirms that content-level scaffolding does not automatically facilitate grammatical development.

Low AI Scaffolding Environment (Phase 3)

Figure 8

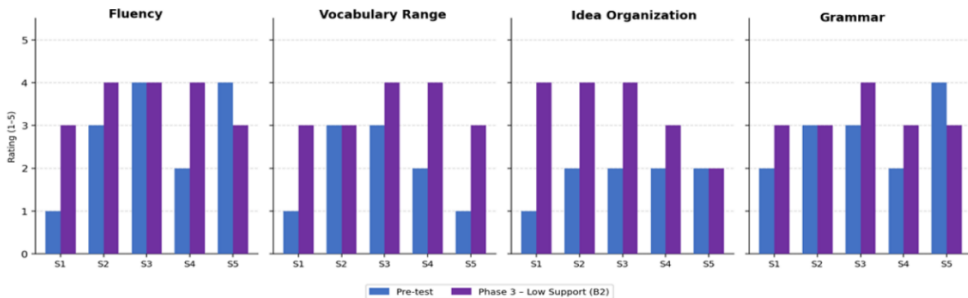
*Pre-Test vs. Phase 3 Oral Performance of Thai EFL Students by Dimension****



According to Figure 8, the greatest improvement was found in *Student 2*, who moved from a low pre-test response to fully organized content, an audience-oriented monologue produced without AI-question scaffolding. Furthermore, *Student 1* similarly demonstrated a clear autonomous improvement in generating a coherent seven-topic recommendation compared to the single-clause pre-test. For example, “*The cost is not too expensive,*” “*so it’s good for students.*” *Student 3* presents the most theoretically important finding. Despite high fluency in culturally embedded stories (“*I blessed something to that old palace, and it’s come true*”) and vivid descriptions (“*It’s crispy outside. The smell is running on my nose*”), morphosyntactic accuracy remained the same across all three phases. Errors such as “*The best activity that I was go*” and “*I’m really enjoy seeing*” existed across the three-phase support conditions. Next, *Student 4*, having demonstrated greater independence in Phase 2, reverted to AI-elicitation strategies under low support by responding with questions (“*Do you know Klong Lan waterfall?*”) rather than delivering an independent monologue. This could suggest that the structure of the interaction format, rather than the quantity of support, is a critical variable, and that removing structural scaffolding without adequate preparation may increase rather than reduce AI dependence. Lastly, *Student 5*’s response shows communicative creativity under the low support phase. The audience-segmentation strategy (“*If you want introvert vibe... if you wanna party*”) represents an organizational innovation absent from all the previous phases. This suggests that reduced AI support may create space for the student’s creativity and communicative style.

Figure 9

*Pre-Test vs. Phase 3 Oral Performance of Chinese EFL Students by Dimension****



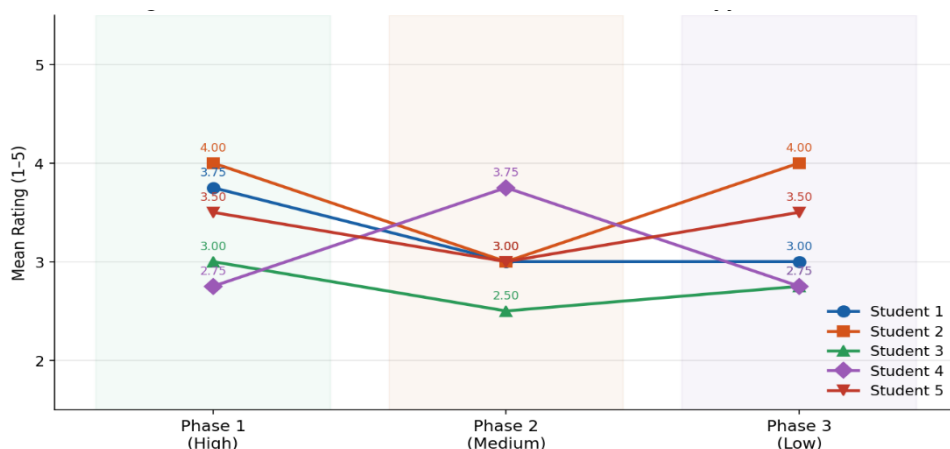
In Figure 9, four main patterns across the Chinese EFL learners can be identified. First, all five Chinese learners maintained or improved their pre-test scores in at least three dimensions during Phase 3. This indicates that repeated exposure to AI-assisted interaction across the two earlier phases helps learners develop and retain communicative resources. In addition, for *Student 1*, the limited pre-test production developed into a coherent, multi-topic, audience-oriented monologue in Phase 3, produced without structural guidance. This demonstrates a clear transfer of previously supported speaking skills into a more independent performance. Next, idea organization shows the most consistent improvement across the learners, with *Students 1, 2, 3*, and *4* achieving scores of 3 or higher. This could suggest that repeated experience with structured interaction, particularly the Q&A AI format in Phase 2, helps these learners develop organizational patterns that they were later able to use independently in Phase 3. Third, *Student 3* produced the most noticeable performance in this condition, achieving a score of 4 across all four dimensions. As the only participant to receive a grammar score of 4 in Phase 3, *Student 3* differed from the broader pattern of limited grammatical development observed in Phases 1 and 2. This finding illustrates that reduced mediation, as well as accumulated experience from earlier phases, may have allowed the learner's existing grammatical knowledge to emerge more fully during independent speaking. Finally, *Student 5* presents an important exception. Unlike the other participants, this student showed a slight decline in content and no improvement in idea organization. This means that low-support conditions may not benefit all learners equally. For learners who rely more heavily on structured guidance, reducing support may limit the breadth and coherence of their spoken outputs.

The following section discusses the findings related to both Thai and Chinese students' AI-scaffolded oral performances.

Thai Students' AI-Scaffolded Oral Performances

Figure 10

*Mean Scores of Thai EFL Students across Three Support Conditions****



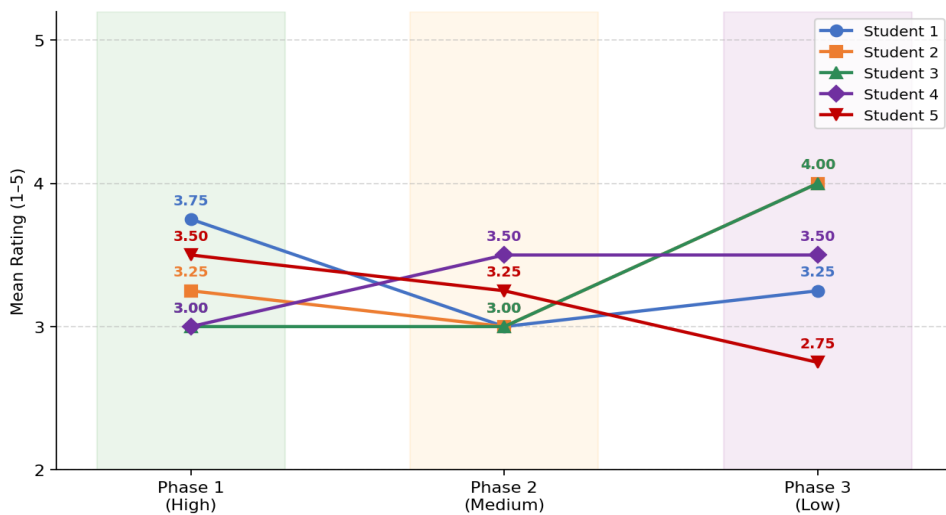
As shown in Figure 10, AI-supported oral tasks generally helped Thai EFL learners perform better, though the benefits differed across learners and conditions. The most consistent pattern was the gap between fluency and grammatical accuracy. All five learners showed improvement in fluency, vocabulary range, or idea organization across at least one phase. However, grammatical progress was limited and inconsistent, and in some cases—such as Student 3—absent across all conditions. This suggests that content-level scaffolding alone is insufficient to drive morphosyntactic development. Support format also proved more consequential than support quantity. For example, Student 4's performance trajectory suggests that the structured question prompts in Phase 2 were more facilitative than either the extensive model language of Phase 1 or the minimal instruction of Phase 3. This suggests that interaction structure may be a more decisive factor than the amount of support provided. However, *Students 1 and 2* showed evidence of cumulative scaffolding internalization, maintaining or improving Phase 1 performance levels as support was progressively reduced. Also, Student 2 achieved peak performance in Phase 3 without any AI prompting. Finally, *Student 5's* Phase 3 output reveals that low support conditions may create space for communicative creativity and individual voice. Rather than limiting performance, the withdrawal of direct linguistic modelling appeared to encourage more personally expressive oral production. This implies that the

sequencing of support reduction is important in AI-assisted speaking pedagogy.

Chinese Students' AI-Scaffolded Oral Performances

Figure 11

*Mean Scores of Chinese EFL Students across Three Support Conditions****



It is evident in Figure 11 that the three-phase data reveal four important patterns among Chinese EFL learners. First, *Students 1 to 4* maintained or improved their Phase 1 scores by Phase 3, showing that repeated interactions across phases help learners carry communicative skills into more independent speaking. This pattern is vividly observed in *Students 2 and 3*, both of whom achieved their highest mean scores (4.00) in Phase 3 without structural support. Second, *Student 4's* progression illustrates the importance of interaction structure over the amount of support provided. While Phase 1 was the weakest condition for this learner, Phase 2 marked a clear improvement. The structured Q&A format encouraged active reformulation (*"If you only have half a day, definitely go for the rickshaw tour in the hutongs"*) in ways that the heavily modeled language of Phase 1 did not. Next, grammatical complexity remained stable at 3 for most participants across all three phases. This reflects the continued separation between content development and grammatical control. The only exception was found in *Student 3*, whose grammar score increased to 4 in Phase 3. Notably, this occurred under the condition with the least linguistic support, which suggests

that extensive content-level guidance may sometimes limit morphosyntactic development for certain learners. Finally, *Student 5* showed a different pattern from *Students 2* and *3*, with performance declining as support was reduced. This finding indicates that lower levels of support do not benefit all learners equally. For students who rely more heavily on structured guidance, the gradual removal of support may restrict them rather than strengthen their communicative performance.

To address the second research question, this section reports on participants' demographics, questionnaire results, and open-ended responses.

Participants' Demographics

Among the five Thai participants, three were aged 20–22, while two were between 17 and 19 years old. Among Chinese learners, three were aged 17–19 and two were aged 20–22. Regarding English proficiency, Thai participants self-reported levels ranging from A1 to B1, including one learner at A1, two at A2, and two at B1. By comparison, the Chinese learners showed slightly higher proficiency levels, with three participants at B1, one at B2, and one at C1. Hence, the Chinese learners seemed to perceive themselves as having slightly stronger English proficiency than the Thai learners.

Regarding AI tool usage, Thai learners reported using between one and three AI tools in their daily lives. ChatGPT was the most frequently used platform, followed by Gemini and Claude. In contrast, Chinese learners reported using between two and four AI tools, primarily domestically developed platforms such as Doubao, DeepSeek, YuanBao, and Kimi. Regarding the purposes of use, Thai learners mainly perceived AI as a means of personalized learning support and language skill development, with one participant emphasizing its role in reducing learning anxiety. However, Chinese learners emphasized efficiency enhancement alongside personalized learning and language improvement, revealing a more productivity-oriented approach to AI use.

Questionnaire Results on Perceptions towards AI Tools Use

Table 2

Mean Scores and Standard Deviations of Thai (n = 5) and Chinese (n = 5) EFL Learners' Perceptions of AI Tools

Item	Thai EFL Learners (n = 5)		Chinese EFL Learners (n = 5)	
	M	SD	M	SD
Part 1: Perceived Usefulness				
1. Overall speaking improvement	4.60	0.49	4.20	0.40
2. Learning productivity	4.20	0.75	4.60	0.49
3. Communicative effectiveness	3.80	0.75	3.60	0.49
4. General study usefulness	4.20	0.40	4.80	0.40
5. Vocabulary enhancement	4.20	0.75	4.40	0.49
6. Fluency improvement	3.80	0.75	2.60	0.49
7. Pronunciation correction	4.00	0.63	3.40	1.20
8. Grammar complexity improvement	4.00	0.63	3.80	0.75
<i>Section Mean</i>	4.10	0.70	3.92	0.93
Part 2: Perceived Ease of Use				
9. Clarity of interaction	3.80	0.75	4.00	1.10
10. Tool controllability	4.40	0.49	3.60	0.49
11. Ease of operation	4.50	0.50	4.40	0.80
12. Advantage over traditional methods	4.20	0.75	4.60	0.80
13. Low mental effort required	4.20	0.75	4.40	0.80
<i>Section Mean</i>	4.21	0.71	4.20	0.89
Part 3: Attitude & Enjoyment				
14. Enjoyment of use	3.80	0.75	4.20	0.75
15. Interest in AI-generated content	4.00	1.10	4.40	0.49
16. Motivation for speaking practice	3.60	1.02	3.00	0.63
17. Reduced speaking anxiety	3.80	0.75	3.60	1.20
18. Stress-free practice environment	4.00	0.89	3.80	1.17
<i>Section Mean</i>	3.84	0.92	3.80	1.02
Part 4: Output Quality & Result Demonstrability				
19. Quality of feedback	4.00	0.63	4.40	0.49
20. Apparent learning outcomes	4.20	0.40	3.40	0.49
21. Communicability of benefits	4.00	0.63	4.20	0.40
22. Naturalness of AI voice/interaction	3.80	0.75	3.20	0.75
<i>Section Mean</i>	4.00	0.63	3.80	0.75
Part 5: Social Influence & Facilitating Conditions				
23. Encouragement from significant others	3.60	0.80	3.80	0.75
24. Peer image/status enhancement	4.20	0.40	2.80	0.40
25. Resource availability	4.00	0.63	4.40	0.80
26. Data privacy security	3.80	0.75	2.80	1.17
<i>Section Mean</i>	3.90	0.70	3.45	1.07
Part 6: Behavioral Intention				

Item	Thai EFL		Chinese EFL	
	Learners (n = 5)		Learners (n = 5)	
	M	SD	M	SD
27. Future use intention	3.80	0.75	3.80	0.40
28. Predicted frequency of use	4.20	0.75	4.40	0.80
29. Recommendation to others	4.20	0.75	3.80	0.40
<i>Section Mean</i>	4.07	0.77	4.00	0.63
Overall Mean	4.03	0.75	3.88	0.94

Both groups of learners appeared to illustrate positive attitudes toward AI-assisted English-speaking practice across all questionnaire items. Thai learners reported a slightly higher overall mean score ($M = 4.03$, $SD = 0.75$) than Chinese learners ($M = 3.88$, $SD = 0.94$), though given the small sample size, this difference should be interpreted as a descriptive tendency rather than a definitive cross-group distinction.

Regarding perceived usefulness, the Thai group tended to show a higher mean ($M = 4.10$) than the Chinese learners ($M = 3.92$). Fluency improvement (Item 6: $M = 3.80$ vs. 2.60) and pronunciation correction (Item 7: $M = 4.00$ vs. 3.40) were rated more positively by Thai learners, suggesting they may have perceived AI tools as particularly beneficial for oral improvement. However, Chinese learners rated general study usefulness (Item 4: $M = 4.80$ vs. 4.20) and learning productivity (Item 2: $M = 4.60$ vs. 4.20) more highly than the Thai group, possibly indicating a tendency to view AI primarily as a broader academic support tool.

In terms of the perceived ease of use, both groups reported almost similar section means (Thai: $M = 4.21$; Chinese: $M = 4.20$), suggesting comparably positive perceptions of AI ease of use. Thai learners appeared to rate greater ease in controlling AI tools to achieve desired outputs (Item 10: $M = 4.40$ vs. 3.60), whereas Chinese learners rated the advantages over traditional methods (Item 12: $M = 4.60$ vs. 4.20) and low mental effort (Item 13: $M = 4.40$ vs. 4.20) slightly higher than the Thai group.

For attitude and enjoyment, Thai learners rated motivation for speaking practice (Item 16: $M = 3.60$ vs. 3.00) and stress-free practice environments provided by AI tools (Item 18: $M = 4.00$ vs. 3.80) more positively than the Chinese group, possibly suggesting a tendency toward stronger affective support from AI interaction. However, Chinese learners appeared to report greater enjoyment (Item 14: $M = 4.20$ vs. 3.80) and found AI-generated content more interesting (Item 15: $M = 4.40$ vs. 4.00) than the Thai group.

Regarding output quality and result demonstrability, Thai learners tended to evaluate learning visibility (Item 20: $M = 4.20$ vs. 3.40) and interaction naturalness (Item 22: $M = 3.80$ vs. 3.20) more positively, while

Chinese learners rated feedback quality slightly higher (Item 19: $M = 4.40$ vs. 4.00) than the Thai group.

The most notable descriptive difference between the two groups was observed in social influence and facilitating conditions (Thai: $M = 3.90$; Chinese: $M = 3.45$). Thai learners appeared to show higher perceptions of peer status enhancement (Item 24: $M = 4.20$ vs. 2.80) and stronger confidence in data privacy (Item 26: $M = 3.80$ vs. 2.80) than the Chinese group. Nonetheless, Chinese learners rated resource availability slightly higher (Item 25: $M = 4.40$ vs. 4.00) than the Thai group.

Finally, both groups appeared to report positive behavioral intentions toward future AI use. Chinese learners tended to report slightly higher predicted frequency of future use (Item 28: $M = 4.40$ vs. 4.20), whereas Thai learners expressed stronger recommendation intentions (Item 29: $M = 4.20$ vs. 3.80).

Open-Ended Responses

The open-ended section invited all learners to reflect on three main issues: (1) the most helpful AI features for speaking development, (2) difficulties encountered, and (3) replaceability of human teachers.

Most Helpful AI Features

Participants from both contexts identified instant feedback and constant accessibility as the most valuable features of AI-assisted speaking practice. Thai learners particularly highlighted AI's ability to identify and correct errors in real time. For example, Thai *Student 2* stated, "*AI know what I'm saying when I speak wrong and fix my word.*" Similarly, Thai *Student 5* further supported using ChatGPT to prepare for classroom presentations: "*That helps about grammar and some words is wrong.*" Chinese participants expressed similar views but often provided more detailed reflections on conversational support and language development. Chinese *Student 1* highlighted the AI's interactive capacity: "*It can practice oral English with me anytime and keep me speaking by asking questions.*" Chinese *Student 2* further emphasized vocabulary support, describing how AI could "*provide me large amount of vocabulary that I can use in my speaking.*" Chinese *Student 3* also pointed to the breadth and evolving nature of AI knowledge, noting that it "*... can provide assistance at any time, with its functions continuously updated and evolving.*" Therefore, both Thai and Chinese learners praised AI not only for cognitive support such as grammar, pronunciation, and vocabulary improvement, but also for motivational benefits including instant feedback, flexible access, and sustained speaking practice.

Difficulties Encountered

The most frequently reported difficulty was the robotic quality of AI-generated speech. This was evident among Thai learners regarding unnatural voice and unclear pronunciation. Thai *Student 1* explained that *“the tone is too robotic, and the pronunciation may not be clear,”* while Thai *Student 4* summarized the issue simply as *“Robot voice.”* Another barrier involved pronunciation recognition. Thai *Student 5* described how communication sometimes failed because the AI could not understand non-native pronunciation: *“My talk is not good and GPT don’t understand and answer me wrong topic.”* Similarly, this aspect was also observed in Chinese *Student 2*’s response, *“Misunderstood pronunciation.”* These responses suggest limitations in AI speech-recognition systems when processing different English accents. In addition, Chinese participants identified limitations related to contextual understanding and natural conversations. Chinese *Student 1* noted that AI *“sometimes forgets what we talked about before,”* reducing the sense of authentic interaction. Chinese *Student 3* further argued that AI struggled with metaphorical language and remained *“rigid”* in communication. In sum, these findings indicate that although AI supports speaking practice effectively, limitations remain in natural voice, contextual memory, and pragmatic understanding.

Replaceability of Human Teachers

All participants from both contexts rejected the idea that AI could fully replace human teachers. Thai learners mainly emphasized the emotional and relational dimensions of teaching. Thai *Student 1* argued that *“AI lacks emotions and feelings,”* while Thai *Student 2* stated directly, *“AI can’t teach emotion.”* Additionally, Thai *Student 5* noted the importance of human empathy, explaining that *“teachers can recognise students’ feelings in ways AI cannot.”* Chinese learners, in contrast, highlighted pedagogical expertise and personalised guidance. Chinese *Student 4* explained that effective AI use requires strong learner autonomy and that students still need *“a human teacher to provide some guidance.”* Chinese *Student 1* similarly argued that teachers provide *“more targeted guidance to improve my speaking.”* Chinese *Student 2* added that teachers help learners speak more naturally and contextually, offering *“more practical advice.”* Overall, both groups viewed AI as a valuable supplementary tool rather than a replacement for human instruction.

Discussion

From Phases 1–3, it can be concluded that AI scaffolding can enhance EFL oral performance, although the outcomes differed across learners and

contexts. Overall, both Thai and Chinese learners showed high gains in vocabulary use and idea organization, especially during high support (Phase 1) and Q&A support (Phase 2). The findings from this study support earlier studies in that AI-assisted interaction helps learners produce longer and more organized oral responses (Charoenchotphokin, 2024; Chen et al., 2025; Tam, 2025). Moreover, the findings also echo Hammond and Gibbons' (2005) idea of graduated support, where assistance is gradually reduced as learners become more independent.

Interestingly, the type of interaction mattered more than the amount of support. Both Thai *Student 4* and Chinese *Student 4* performed best during Phase 2, where the AI tools provided guiding questions instead of giving ready-made model answers. This supports Van de Pol et al.'s (2010) scaffolding framework, especially the ideas of fading support and transferring responsibility to learners. The Q&A format encouraged students to actively create their own language patterns rather than rely on AI-generated responses. This also aligns with Kim and Hwang's (2026) study, which found that different task types create different levels of cognitive demand during speaking. Although learners produced richer content and better organization, grammar development remained unchanged across all phases. This finding supports Bui and Skehan's (2018) argument that speaking development often improves more in fluency and complexity than in accuracy. Thai *Student 3*, for example, gave rich responses but continued to use non-standard constructions in speech. Similar findings were reported by Tam (2025), who found improvements in fluency and confidence but continued grammatical problems. These results suggest that AI works better as a speaking practice partner than as a grammar trainer. Therefore, grammar-focused instruction by teachers is still necessary alongside AI-supported speaking tasks.

Phase 3 showed mixed findings. Some learners, including Thai *Students 1* and *2* and Chinese *Students 1 to 3*, became more independent and creative in their responses. This reflects Wood et al.'s (1976) idea that the goal of scaffolding is learner independence. Chen et al. (2025) also found that AI-supported speaking practice increased learners' willingness to experiment with language over time. Chinese *Student 3* showed this clearly by gaining the highest grammar score in Phase 3 despite receiving the least support. On the other hand, some learners struggled when support was removed. For example, in Phase 3, Thai *Student 4* relied heavily on AI prompts again, while Chinese *Student 5's* overall performance declined. These results support Van de Pol et al.'s (2010) principle of contingency, in that scaffolding should match learners' readiness levels. Lower-proficiency learners may need more gradual support reduction to maintain oral performance. Although Chinese learners generally started with higher proficiency levels than Thai learners, both groups showed similar patterns of development across the three phases.

Thai learners used ChatGPT, while Chinese learners used Doubao, but both platforms produced similar speaking outcomes. However, Thai learners focused more on real-time speaking correction, while Chinese learners emphasized efficiency and broader academic support. This suggests that learners' cultural and educational backgrounds may influence how they use AI tools in language learning contexts. These findings suggest that AI should be understood not just as a tool that delivers practice activities, but as an interactive environment where learners actively shape their own language use. In addition, using AI for speaking practice requires learners to monitor and manage their own learning process more actively. This suggests that AI-mediated practice does more than build language skills since it also develops learner autonomy, meaning the ability to take charge of one's own learning.

The findings from the TAM questionnaire showed positive perceptions of AI-assisted speaking practice. Thai learners reported a slightly higher overall mean score (4.03) than Chinese learners (3.88), which suggests that both groups considered AI useful and easy to use for oral English learning. These findings support earlier research (e.g., Davis, 1989; Granić, 2023) in that usefulness and ease of use strongly influence technology adoption in education contexts. Furthermore, Thai learners rated AI as a tool for improving speaking skills such as fluency and pronunciation, while Chinese learners rated AI more highly for general academic productivity. This difference may be related to the Thai learners' lower English proficiency and stronger need for speaking practice. Chinese learners, meanwhile, appeared to see AI as a normal part of academic study, which may reflect broader digital education practices in China (Zou et al., 2022).

Both groups rated AI as easy to use, but qualitative responses showed some similar concerns. Learners mentioned robotic AI communication, unnatural pronunciation, and limited authentic conversation. These findings agree with Tam's (2025) study on robotic interaction and speech-recognition problems as common limitations of AI speaking tools. The findings suggest that ease of use in speaking contexts extends to how natural and meaningful the interaction feels. Learners also reflected that AI reduced speaking anxiety and created a less stressful practice environment. This supports past studies showing that lower anxiety encourages oral participation in EFL settings (Boonkit, 2010; Ork et al., 2024). However, motivation scores were lower than usefulness scores, especially among Chinese learners, suggesting that students found AI practical rather than highly enjoyable. Finally, learners from both groups rejected the idea that AI could replace human teachers. Thai learners emphasized emotional support and human connection, while Chinese learners emphasized teachers' expertise and personalized guidance. These findings reinforce the idea that AI should be used as a supportive

learning tool rather than a substitute for human instruction (Charoenchotphokin, 2024; Huang et al., 2022; Tam, 2025).

Pedagogical Implications, Limitations and Recommendations

Three practical implications can be identified for EFL teachers. First, AI tools work better when built into a structured, step-by-step speaking task rather than left as an optional add-on. Designing tasks that gradually transfer responsibility back to learners aligns with how scaffolding is intended to work. Second, how AI interacts with learners matters more than how much help it gives. When students are asked to respond to questions rather than copy or confirm AI-generated sentences, they speak more independently and produce richer language. Therefore, teachers should design prompts that push learners to create their own responses. Third, while AI helps learners speak more fluently and build vocabulary, it does not improve grammar on its own, as reflected in the findings. Therefore, teachers still need to give targeted grammar feedback alongside AI practice. A combined approach where AI provides extra speaking time and teachers focus on accuracy works best.

However, this study has some limitations that should be acknowledged for future studies. The study involved only ten participants from the two contexts, so the patterns observed may reflect individual experiences rather than broader trends, and any generalizations drawn from this exploratory sample should therefore be treated with caution. The two groups used different AI platforms (e.g., ChatGPT and Doubao), making it difficult to compare results directly. These platforms differ not only in their training data, but also in the sociocultural and linguistic contexts in which they have been developed and deployed. They may therefore respond to learner input differently, produce varying quality of language feedback, and offer distinct interactional affordances. Consequently, observed differences between the Thai and Chinese groups cannot be attributed solely to learner variables or cultural backgrounds, as platform-specific effects may have also shaped performance outcomes. Future research employing a within-group or counterbalanced design where participants across both contexts use the same AI platform would allow for more controlled and insightful comparisons. In addition, only one speaking topic (i.e., Travel Recommendation) was used across all tasks, and there was no comparison group of students working without AI. The absence of a control group means that improvements observed across the three phases cannot be definitively attributed to AI scaffolding, as factors such as repeated task practice, topic familiarity, and learner motivation may have also contributed to the gains. As a result, it is difficult to determine exactly how much AI contributed to the observed

improvements. Future studies should work with larger groups over longer periods to examine whether speaking gains extend beyond a single task sequence. Moreover, comparing different AI interaction formats across proficiency levels would help clarify which approaches are most suitable for different learners.

Conclusion

This study investigated how generative AI scaffolding supports EFL speaking development and how learners perceive the use of AI tools for oral practice. The three-phase task design showed that AI can give learners more chances to speak beyond the classroom, with both groups showing growth in vocabulary range, idea organization, and speaking confidence. Grammar, however, showed little change throughout, which suggests that AI is better suited as a speaking practice partner than a tool for correcting sentence structures. Regarding the questionnaire results, both groups reported positive attitudes, though they had different priorities. Thai students viewed AI mainly as a way to improve their speaking, while Chinese students found it useful for academic tasks more broadly. Both groups found AI relatively easy to use but pointed out that conversations sometimes felt unnatural and robotic. Interestingly, no student from either context believed AI could replace human teachers. This suggests that learners still value personal connection and guidance in the classroom.

Notes

* Figures 1-2 images were generated by Gemini AI.

** Figure 3 image was generated by ChatGPT AI

*** Figures 4-11 images were generated by Claude AI.

About the Authors

Monthon Kanokpermpoon: An Assistant Professor of Education (English Language Teaching) at the Language Institute, Thammasat University, Thailand. He obtained his PhD in Educational and Applied Linguistics from Newcastle University, UK. His specializations include cognition in language education, curriculum and instruction, and second language learning. ORCID ID: <https://orcid.org/0000-0002-4318-6939>

Thayakorn Klomkaew: A Lecturer in English at the English Program, Kamphaeng Phet Rajabhat University, Thailand. She obtained her PhD in English Language Teaching from Thammasat University, Thailand. Her research interests are corpus-assisted (critical) discourse analysis, language

and gender, and data-driven learning. ORCID ID: <https://orcid.org/0000-0002-6772-1756>

Yuwei Liang: A Lecturer at the College of Foreign Languages, Beijing University of Technology, P. R. China. She obtained her PhD in Education from Newcastle University, UK. Her current research interests include intercultural communication, language and identity, and internationalization of higher education. ORCID ID: <https://orcid.org/0000-0002-6571-5723>

References

- Anthropic. (2026). *Claude Opus 4.7* [Large language model].
<https://claude.ai/>
- Boonkit, K. (2010). Enhancing the development of speaking skills for non-native speakers of English. *Procedia - Social and Behavioral Sciences*, 2(2), 1305–1309. <https://doi.org/10.1016/j.sbspro.2010.03.191>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
<https://doi.org/10.1191/1478088706qp063oa>
- Bui, G., & Skehan, P. (2018). Complexity, accuracy, and fluency. In J. I. L. (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–7). John Wiley & Sons, Inc.
<https://onlinelibrary.wiley.com/doi/10.1002/9781118784235.eelt0046>
- Bulut, B. (2022). Analytic rubric development study for assessment of the speaking and writing skills at primary school level. *International Journal of Progressive Education*, 18(6), 180–193.
<https://doi.org/10.29329/ijpe.2022.477.12>
- Charoenchotphokin, E. (2024). *Using ChatGPT 4.0 to enhance speaking skills through scaffolding prompts for extended utterance construction* [Unpublished Master's special study report]. King Mongkut's University of Technology Thonburi.
- Chen, Y., Ke, N., Huang, L., & Luo, R. (2025). The role of GenAI in EFL speaking: Effects on oral proficiency, anxiety, and risk-taking. *RELJ Journal*, 0(0). <https://doi.org/10.1177/00336882251341072>
- Chesney, T. (2006). An acceptance model for useful and fun information systems. *Human Technology*, 2(2), 225–235.
<http://dx.doi.org/10.17011/ht/urn.2006520>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>

- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35(8), 982–1003.
<https://doi.org/10.1287/mnsc.35.8.982>
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology*, 22(14), 1111–1132.
<https://doi.org/10.1111/j.1559-1816.1992.tb00945.x>
- Denzin, N. (2006). *Sociological methods: A sourcebook*. Aldine Transaction.
- Fryer, L. K., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3), 8–14.
<https://doi.org/10.64152/10125/44068>
- Google. (2026). *Gemini 3 Pro* [Large language model].
<https://gemini.google.com/>
- Granić, A. (2023). Technology acceptance and adoption in education. In O. Zawacki-Richter & I. Jung (Eds.), *Handbook of open, distance and digital education*. Springer. https://doi.org/10.1007/978-981-19-2080-6_11
- Hammond, J., & Gibbons, P. (2005). Putting scaffolding to work: The contribution of scaffolding in articulating ESL education. *Prospect*, 20(1), 6–30.
<https://search.informit.org/doi/10.3316/aeipt.143258>
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning: Are they useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38, 237–257. <https://doi.org/10.1111/jcal.12610>
- Kim, H., & Hwang, H. (2026). Task complexity and genre effects: Examining complexity, accuracy, and fluency in L2 oral production. *Applied Linguistics*, amag033.
<https://doi.org/10.1093/applin/amag033>
- OpenAI. (2026). *ChatGPT* (GPT-5.5) [Large language model].
<https://chatgpt.com/>
- Ork, Y., Chin, P., Ban, T., & Em, S. (2024). Factors causing students' challenges in learning English speaking skills: A review of empirical works. *Cambodian Journal of Educational and Social Sciences*, 1(1), 26–36. <https://doi.org/10.69496/cjess.v1i1.6>
- Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge University Press.
- Shaengchart, Y., Bhumpenpein, N., Kongnakorn, K., Khwannu, P., Tiwtakul, A., & Detmee, S. (2023). Factors influencing the acceptance of ChatGPT usage among higher education students in

- Bangkok, Thailand. *Advance Knowledge for Executives*, 2(4), Article 28. <https://ssrn.com/abstract=4592118>
- Skehan, P. (2011). *A cognitive approach to language learning*. Oxford University Press.
- Tam, T. T. T. (2025). The impact of AI-powered speaking tools on EFL students at Nguyen Tat Thanh University. *MSI Journal of Arts, Law and Justice*, 2(6), 1–20. <https://doi.org/10.5281/zenodo.15675420>
- Taylor, P., Argasvipart, K., Kanokpermpoon, M., Rattanawisadrat, N., Dyamond, B. J., & Hrylytskyy, A. (2023). Stakeholders' perceptions related to technology acceptance of Reading Progress in Microsoft Teams: A case study of a trilingual program at a secondary school in Thailand. *LEARN Journal: Language Education and Acquisition Research Network*, 16(2), 718–736. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/266981>
- Tuan, L. T. (2012). Teaching and assessing speaking through analytic scoring approach. *Theory and Practice in Language Studies*, 2(4), 673–679. <https://doi.org/10.4304/tpls.2.4.673-679>
- Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review*, 22(2), 271–296. <https://doi.org/10.1007/s10648-010-9127-6>
- Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly*, 28(4), 695–704. <https://doi.org/10.2307/25148660>
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. <https://doi.org/10.1287/mnsc.46.2.186.11926>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Zou, C., Li, P., & Jin, L. (2022). Integrating smartphones in EFL classrooms: Students' satisfaction and perceived learning performance. *Education and Information Technologies*, 27, 12667–12688. <https://doi.org/10.1007/s10639-022-11103-7>