Exploring the Integration of AI Chatbots into Task-based Language Teaching among Chinese Primary School Students: Associations with English-speaking Fluency and Anxiety

Zhai Xin and Siti Nazleen Abdul Rabu *

Centre for Instructional Technology and Multimedia, Universiti Sains Malaysia, Pulau Pinang, Malaysia
Email: zhaixin@student.usm.my (Z.X.); snazleen@usm.my (S.N.A.R.)
*Corresponding author

Manuscript received January 12, 2026; revised February 27, 2026; accepted April 7, 2026; published September 15, 2026

***Abstract*—This study investigates the persistent challenges in English speaking instruction in lower-tier cities in China, where students often exhibit low English speaking fluency and heightened anxiety associated with limited speaking practice and a reading-centric teaching approach. While Artificial Intelligence (AI) Chatbots (AICT) have shown promise in higher education, empirical evidence regarding their integration into Task-Based Language Teaching (TBLT) at the primary school level remains scarce. To address this research gap and offer a scalable pedagogical contribution, we implemented an 8-week AICT-supported TBLT program involving 68 fifth-grade students. Using a one-group pretest–posttest design, we measured temporal changes in students' English-speaking fluency (assessed via verbal productivity, hesitation, repetition, and self-correction) and their self-reported speaking anxiety. Data were analysed using paired-sample t-tests. The results revealed statistically significant improvements ($p < 0.001$) across all fluency dimensions, with particularly large effect sizes in reducing repetition (d = 0.918) and self-correction ($d = 0.782$), alongside moderate gains in verbal productivity ($d = 0.408$) and hesitation ($d = 0.532$). Additionally, the intervention was associated with a comprehensive decrease in speaking anxiety ($p < 0.001$), with large effect sizes ($d > 0.8$) observed for 11 of the 20 items. These findings suggest that integrating AICT as an interactive tool within a TBLT framework can help create a supportive, low-anxiety speaking environment, potentially offering a practical pathway associated with the support of oral English teaching in resource-limited primary school contexts in China.**

***Keywords*—Artificial Intelligence (AI) chatbots, English-speaking fluency, speaking anxiety, task-based language teaching, primary English education**

I. Introduction

In an era marked by deepening globalisation and digital integration, English proficiency has transcended the boundaries of a single academic discipline and has become a critical competency linking knowledge, technology, and culture. English-speaking ability not only reflects learners' overall ability to use the language but also provides the main channel for negotiating meaning, building genuine emotional connections, and exchanging ideas across cultures. Therefore, developing students' spoken English has foundational value in China's basic education, particularly in the upper grades of primary school, which is a critical stage in language development. Among these grades, the fifth grade represents a pivotal transition point, as students are undergoing a cognitive shift from concrete thinking to more abstract and logical reasoning [1, 2]. At the same time, they are in a highly sensitive period for sound perception, intonation development, and willingness to communicate [3, 4]. The speaking habits and interaction attitudes developed during this phase are likely to exert a lasting influence on students' future language development.

Yet, in many fifth-tier cities in northern China, educational resources are relatively scarce. English teachers' weak oral proficiency and exam-oriented pressures have led to a focus on written skills over spoken practice, emphasising knowledge transmission while neglecting practical communication [5, 6]. Consequently, English teaching tends to overemphasise vocabulary memorisation and grammar rules [7], leaving many fifth-grade students struggling to speak and exhibiting low levels of oral fluency. Language learning anxiety has been recognised as a common and serious barrier to learning in both primary and secondary education [8, 9]. For primary school students, whose minds are still developing, anxiety in traditional teacher-centred classrooms often stems from the fear of making mistakes, being watched by peers, and facing immediate teacher corrections. This anxiety contributes to the phenomenon of "mute English" [10], which undermines confidence, lowers self-efficacy, and ultimately constrains long-term oral English development [11].

Amid these challenges, rapid advances in artificial intelligence have created unprecedented opportunities for innovation in language education. AI chatbots equipped with speech recognition and natural language processing capabilities foster interactive and responsive learning environments by providing real-time feedback. Such systems can provide repeated, personalised, and low-pressure speaking practice, as demonstrated by Safitri *et al*. [12]. By simulating authentic conversations and providing tailored language support, AI chatbots have been found to encourage speech that is both more fluent and accurate [13, 14]. Furthermore, by establishing a safe, nonjudgmental practice space, Al-supported speaking activities have the potential to enhance learners' confidence and reduce speaking-related anxiety [15, 16]. Unlike traditional Computer-Assisted Language Learning (CALL) approaches, which historically rely on static, rule-based branching scenarios and predetermined correct answers, modern generative AI chatbots utilise advanced natural language processing to engage in dynamic, context-aware, and open-ended conversations. This technological shift allows AI to move beyond simple pattern-matching or automated grading; instead, it provides highly personalised, real-time linguistic scaffolding that organically adapts to learners' immediate communicative needs, closely mimicking authentic human

interaction.

However, AI merely serves as a teaching tool and an intermediary resource, and its use must be paired with appropriate instructional strategies to yield positive outcomes. In this regard, Bagasala and Estremera [17] indicated that Task-Based Language Teaching (TBLT) can moderately to substantially promote second language acquisition in authentic classroom settings, particularly by enhancing learners' oral fluency and reducing speaking-related anxiety. The core mechanism of TBLT lies in creating authentic language interaction scenarios through meaning-driven tasks, enabling learners to develop linguistic competence through social negotiation and interaction [18].

AI-assisted TBLT instruction aligns closely with Vygotsky's sociocultural theory [19], which emphasises that cognitive development stems from social interaction and relies on culturally mediated tools. Learners progressively internalise external interactions into their oral proficiency through multi-party collaboration with teachers, peers, and AI chatbots while completing tasks. However, a notable lack of empirical research remains on how AI can be integrated into TBLT for primary school English speaking instruction and how such integration influences both fluency and anxiety, especially in resource-constrained educational contexts.

By addressing these challenges, this study makes a unique and necessary contribution to the intersecting literature on TBLT and AI in language learning. While most existing AI-assisted language research focuses on secondary or higher education populations, this study extends the application of Vygotsky's sociocultural theory to young primary-school learners. It uniquely conceptualises the AI chatbot not as a standalone tutor, but as a "dynamic digital scaffold" operating within a carefully orchestrated triadic partnership between the teacher, student, and AI. Furthermore, it provides a much-needed empirical model for how scalable, low-cost AI tools can be embedded in TBLT task cycles to bridge the gap between structured inputs and spontaneous oral outputs in resource-constrained educational settings.

This study aimed to implement a task-based instructional model integrated with Artificial Intelligence (AI) Chatbots (AICT) and examine its potential association with temporal changes in fifth-grade students' English-speaking fluency and anxiety in resource-limited classrooms. It seeks to provide new pathways for integrating technology with pedagogy and overcoming the challenges of oral language instruction.

II. Literature Review

A. English-speaking Fluency

English-speaking fluency is a key dimension in evaluating oral proficiency. It generally refers to a speaker's ability to express ideas continuously at an appropriate pace without frequent pauses or interruptions. Research on fluency dates back to the late 1970s, when Fillmore *et al*. [20] first proposed that fluency is a multifaceted concept that encompasses speech rate, coherence, semantic density, and creative expression. Since then, various scholars have offered definitions of fluency; however, there is broad consensus that fluency extends beyond mere speed. Lennon [21] identified four core elements of English-speaking fluency: verbal productivity, pauses or hesitations, repetitions, and self-corrections. Building on this perspective, Santos and Ramírez-Ávila [22] provided further empirical support through subsequent studies, reinforcing the importance of these components in assessing spoken-language performance.

Speech verbal productivity refers to the speed at which a speaker produces language and is commonly measured in syllables or words per minute. It reflects the speaker's language proficiency level and the degree of automatisation in language processing [23]. Generally, when semantic coherence is maintained, faster verbal speech productivity indicates greater fluency. However, speed alone does not guarantee fluency, as excessively rapid speech may compromise its clarity. Therefore, fluent speech should balance speed with intelligibility [24].

Pausing or hesitation refers to moments of silence or the use of fillers, such as "uh" or "um" during speech. While moderate pausing can aid listener comprehension, excessive or prolonged pauses are often considered key indicators of disfluency [23]. Frequent hesitation typically indicates difficulties in lexical retrieval or syntactic planning, which reduces overall fluency. Moreover, pausing is considered an observable manifestation of cognitive load, with elevated anxiety levels significantly increasing both the frequency and duration of pauses [25].

Repetition involves re-uttering words or segments that have been previously spoken, often due to uncertainty or difficulty in organising speech. For instance, repeating a word or phrase may serve as a strategy for gaining planning time [26]. Although occasional repetition is natural in spoken interactions, excessive repetition may disrupt the speech flow and diminish communicative efficiency [27].

Self-correction refers to the immediate modification of one's utterance upon recognising an error or inappropriate expression. This may involve correcting mispronounced words or rephrasing sentences to improve clarity. Although moderate self-correction supports accurate communication, frequent interruptions in correction can hinder speech continuity and reduce fluency [26]. Fluent speakers typically make minor adjustments seamlessly, without significantly disrupting the speech flow [28].

English-speaking fluency involves more than smooth and efficient speech production. It also requires minimising unnecessary pauses, repetitions, and corrections. Its development is influenced by several factors, including the degree of language automatisation, emotional states such as anxiety, and prior communicative experience [29]. For instance, students with elevated anxiety levels often speak more slowly and pause more frequently as emotional stress diverts cognitive resources [30]. Therefore, supporting fluency in language classrooms requires an instructional approach that addresses both cognitive processes and emotional well-being.

B. English-speaking Anxiety

Anxiety disorders have emerged as the most prevalent mental health issue among Chinese adolescents aged 10 to 14 years, accounting for 19.6% of the total disability-adjusted life years caused by mental illnesses [31]. This figure reflects

the substantial impact of anxiety on the daily functioning and psychological development of adolescents. A growing body of research has demonstrated a significant negative correlation between student anxiety levels and academic performance, suggesting that higher anxiety typically correlates with lower achievement [32–34].

In language learning, anxiety negatively affects learners' expressive abilities. The lack of speaking practice and interactive environments further contributes to the fear of expressing oneself in English [35]. Dan *et al*. [36] found that students with higher anxiety levels tend to remain silent or avoid English communication with teachers and peers during class, limiting opportunities to develop speaking skills.

Speaking anxiety is regarded as the most common and significant form of anxiety in foreign language learning. It refers to the negative emotions that language learners experience when engaging in English-speaking situations, such as nervousness, uneasiness, and fear. According to Horwitz *et al*. [37], foreign language classroom anxiety is a complex emotional state arising from the unique pressures present in classroom contexts. It consists of three main components: communication apprehension, test anxiety, and fear of negative evaluation.

The concept of communication apprehension was first proposed by McCroskey [38], who described the tension and discomfort that learners feel when participating in or anticipating English conversations. Fear of negative evaluation, as defined by Watson and Friend [39], involves learners' concerns about receiving critical or unfavourable judgments regarding their English performance. Many students hesitate to speak not only because they fear the communication act itself but also because they worry about making mistakes and being ridiculed by others [40]. Beyond these aspects, speaking anxiety may include concerns about pronunciation, intonation, and cultural differences. However, its core feature is characterised by a strong sense of nervousness and, in some cases, fear, when learners attempt to express themselves in English.

Numerous empirical studies have confirmed the detrimental impact of anxiety on English speaking performance. For instance, Daymiel *et al*. [41] found that language anxiety significantly reduced students' lexical range and clarity during free-speaking tasks, thereby undermining their overall speaking performance. Kuswantoro and Novita [42] reported that anxiety often stems from fear of making mistakes and lack of confidence, both of which severely inhibit classroom interaction and verbal expression. Similarly, Que *et al*. [43] observed that students with higher anxiety levels tended to avoid speaking, speak more slowly, and struggle with organising thoughts when completing speaking tasks, all of which negatively affected communicative effectiveness. In another study, Yen *et al*. [44] identified fear of negative evaluation and limited vocabulary as common sources of speaking anxiety among upper-grade students, restricting their fluency and self-confidence in English output. These findings underscore the importance of addressing students' mental well-being, particularly anxiety, in primary English education. Reducing speaking anxiety is essential for enhancing instructional effectiveness and supporting language development.

Theoretically, this anxiety-reducing mechanism can be understood through Krashen's Affective Filter Hypothesis. Krashen [45] posits that negative emotional states, such as high anxiety, low self-confidence, and fear of evaluation, create a psychological block—an 'affective filter'—that prevents comprehensible language input from being successfully processed and acquired. Because AI chatbots are inherently patient and eliminate the fear of peer ridicule or teacher criticism, they effectively lower students' affective filters. In this psychologically safe, low-anxiety digital environment, students are significantly more willing to take communicative risks and experiment with language, thereby creating optimal conditions for fluency.

C. Task-Based Language Teaching and English-speaking Learning

Task-Based Language Teaching (TBLT) is an instructional approach that organises learning around communicative tasks rather than discrete linguistic forms. It traces back to Prabhu's Bangalore Project in India, where he argued that language is best acquired when learners tackle meaning-focused tasks rather than memorising grammatical rules [46]. While Nunan [47] later offered a systematic framework for task typology and sequencing from a curriculum design perspective, and Ellis [48, 49] established TBLT's theoretical foundation within second language acquisition, it was Willis [50] who first translated the theoretical vision into a structured classroom model, providing a practical blueprint for task implementation emphasizing authentic communication, interaction, and form-focused reflection. These contributions form a comprehensive "theory-design-practice" framework that underpins TBLT. At its core, TBLT rests on the belief that language develops most effectively through socially situated interactions when learners engage in meaningful, goal-oriented communication [50, 51].

Willis [50] proposed a three-phase framework that was pivotal in the development of task-based pedagogy, transforming it from a theoretical concept into a structured classroom practice. This model consists of three core stages: Pre-task, Task Cycle, and Language Focus. In the Pre-task stage, the teacher introduces the topic, activates learners' background knowledge, and provides essential language input along with strategic preparation. The Task Cycle stage includes task performance, planning, and reporting, with an emphasis on learners using the target language actively and meaningfully to complete communicative tasks. During this phase, the teacher primarily assumed a supportive role, offering guidance and observational feedback. The Language Focus stage comprises two subcomponents: Analysis and Practice. In the analysis phase, the students were encouraged to reflect on the language forms they used during the task, deepening their understanding through guided discovery. The practice phase follows, with the teacher explicitly highlighting relevant structures and organising targeted practice, such as sentence pattern drills or pronunciation exercises. This final stage aims to reinforce formal accuracy while maintaining communicative relevance, thereby integrating meaning-focused interaction with form-focused instruction [50].

Extensive research has demonstrated that TBLT plays a significant role in advancing learners' English-speaking

skills. By centring instruction on communicative tasks with real-world meanings, TBLT stimulates learners' productive language use and improves fluency, articulation, and pronunciation accuracy [51–53].

D. Applications of AI Chatbots in English Speaking

Recent scholarship from 2023 to 2025 highlights a paradigm shift driven by generative AI and Large Language Models (LLMs) in language education [54, 55]. These advanced LLMs, trained on vast text corpora, have rapidly emerged as powerful tools capable of engaging learners in dynamic, context-aware dialogues and personalising learning at scale. Recent empirical studies emphasise that LLMs not only provide highly personalised, real-time linguistic scaffolding, but also significantly enhance student engagement when properly integrated. Furthermore, contemporary evaluations of conversational generative AI chatbots have demonstrated their positive impact on EFL learners' willingness to communicate and self-perceived communicative competence, validating their use as accessible, low-anxiety practice partners. This new generation of AI vastly outperforms traditional rule-based systems, offering a more authentic simulation of human-to-human interaction.

Advances in generative AI have made conversational chatbots highly responsive, enabling their routine use as practical scaffolding tools for spoken English practices [56]. Existing research suggests that these tools can simultaneously support English speaking development and alleviate speaking anxiety. Specifically, the integration of AICT in EFL classrooms has been positively associated with students' speaking fluency, pronunciation, and participation [57, 58]. Furthermore, AICT helps create a less stressful speaking environment, which is linked to reduced anxiety [59, 60]. Recent studies have also demonstrated that AICT can simultaneously improve speaking fluency and lower foreign language anxiety, with a significant negative correlation between them [61]. By creating this low-pressure, non-judgmental practice space, AI chatbots encourage learners to engage in repeated personalised dialogues without fear of negative peer evaluation, which has been shown to reduce expressive inhibition. However, the current literature primarily focuses on secondary school and university students. This study addresses this gap by examining the potential impact of AI chatbots on primary school students' English-speaking fluency and anxiety.

Furthermore, the interactive nature of AI chatbots strongly facilitates Swain's Output Hypothesis. Swain argued that while receiving comprehensible input is necessary, learners must be pushed to actively produce language (output) to achieve higher levels of grammatical accuracy and communicative fluency. AI chatbots compel students to engage in continuous target-language production rather than passive listening. When a communicative breakdown occurs during a chatbot interaction, the system's immediate personalised feedback pushes learners to notice the gap between their intended meaning and their current linguistic capacity. This 'noticing' function prompts students to test new linguistic hypotheses and consciously modify their subsequent output, transforming static language knowledge into active, automated speaking skills.

E. Sociocultural Theory and AI-assisted English-speaking Learning

Sociocultural Theory (SCT), initially proposed by Soviet psychologist Lev Vygotsky, represents a foundational perspective for understanding how learning occurs and emphasises the social origins of cognition and the mediating role of culture [19]. The theory outlines several core concepts, including the Zone of Proximal Development (ZPD), scaffolding, internalisation, and social interaction, which offer a solid theoretical foundation for reconsidering how AI can be meaningfully integrated into English-speaking instruction.

Vygotsky [19] introduced the ZPD concept to capture the difference between what a learner can achieve independently and what becomes possible with guidance from a more experienced individual. This notion reflects the idea that potential development is not fixed but can be expanded through social support. Within this zone, learners can accomplish tasks that they would otherwise find too difficult, provided they receive appropriate assistance from teachers or more capable peers. As Poehner [62] highlighted, the pedagogical significance of the ZPD lies in its focus on proactive instruction—designing learning experiences that slightly exceed the learner's current abilities, thus creating opportunities for growth that would not emerge through independent efforts alone.

Scaffolding originally came from the construction field, where it refers to temporary structures built to support workers as they perform tasks at elevated heights [63]. Wood, Bruner, and Ross first adapted it to the educational context to describe temporary assistance provided by a teacher or a more capable individual to help learners complete tasks that they are unable to accomplish independently [64]. In language education, scaffolding strategies often take the form of instructional prompts, modelling, task simplification, or sentence starters. For instance, in speaking tasks, a teacher might offer the beginning of a sentence or a pragmatic pattern to help students organise their ideas. As learners gain confidence and proficiency, this support is gradually withdrawn, allowing them to develop greater independence in language production [65, 66]. Importantly, scaffolding is not limited to teacher interventions. More capable peers, instructional materials, or digital tools can also serve as effective sources of support, as long as they provide timely assistance at key points in the learning process [67, 68].

Social interaction is a central component of the SCT. Vygotsky emphasised that human development is rooted in interactions with others and that learning is fundamentally a social process [19]. Through engagement with more experienced individuals—such as adults or peers—children not only acquire new knowledge and skills but also gradually develop higher-order psychological functions. For second language learners, meaningful interactions in authentic contexts provide both the motivation and conditions necessary for language practice, thereby facilitating the acquisition and transfer of linguistic knowledge [69]. Research further indicates that when language classrooms are designed to promote collaborative and communicative interactions, learners become more engaged, produce more language output, and are more likely to internalise and transfer what they have learned [70, 71]. In other words,

social interaction functions not only as a bridge between language input and output but also as a catalyst for cognitive development and language internalisation. Thus, creating a supportive and interaction-rich learning environment is essential for enacting SCT's core principles in educational practice [69, 72].

Internalisation refers to the process by which individuals transform external forms of knowledge, skills, and language into part of their internal cognitive systems through social interaction. According to Vygotsky [19], all higher mental functions originate in social activities and gradually become internalised as individual capabilities. In this developmental path, a function first appears between people as a shared activity and only later becomes part of the learner's independent cognitive structure.

This process is particularly evident in the context of language learning. Children initially acquire language through dialogue with adults or peers, and over time, expressions encountered repeatedly in interactions are absorbed into what Vygotsky called private speech, eventually forming the foundation of their internal linguistic system. As Lantolf [73] explained, internalisation is not a passive reproduction of input but a transformative process through which learners construct meaning and regulate cognition using language. In classroom settings, Leslie [74] found that collaborative exchanges among peers provided valuable linguistic support, and these interactions were especially effective in encouraging learner output, which is an essential step in internalising new language forms. Task-based activities designed to foster meaningful interaction thus serve not only to increase language exposure but also to help bridge the gap between external language use and internal mastery of the language.

From the SCT perspective, incorporating AICT into TBLT advances English-speaking instruction on several fronts. First, by supplying linguistic input that slightly outpaces pupils' current proficiency and pairing it with timely prompts, chatbots operate within learners' zones of proximal development, inviting progress through challenges that are well-calibrated. Second, the instant feedback, model sentences, and vocabulary suggestions they deliver form a constantly adapting, individualised 'digital scaffold' that enables students to complete communicative tasks that they cannot yet manage independently. Third, repeated exchanges with chatbots allow learners to absorb new language and gradually convert it into spontaneous output, marking the shift from other regulations to self-regulation, which defines internalisation. Finally, although chatbots are not human interlocutors, the role-plays and task-based dialogues they support still furnish meaningful social interactions. When teachers orchestrate a triadic partnership among teachers, students, and AI, the communicative landscape of the classroom expands considerably. Collectively, these mechanisms enrich TBLT practices and provide a robust theoretical foundation for cultivating English-speaking fluency in primary school students.

III. Methodology

A. Research Design

This study implemented an eight-week intervention comprising two weekly English sessions, each centred on the primary scripted dialogues ("Talk A" and "Talk B") from the corresponding textbook unit. Lessons were designed within a Task-Based Language Teaching (TBLT) framework, with an AI chatbot integrated as a supplementary instructional tool to support spoken English practices.

A one-group pretest–posttest design was adopted in this study. This choice was guided by two main considerations: (i) the participating class was a cohesive group well-suited for the intervention and (ii) the school administration preferred a unified instructional method to ensure fairness for all students. Assessments were administered before and after the intervention to compare changes in English-speaking fluency and anxiety.

B. Research Objectives and Questions

An eight-week instructional intervention was evaluated in which AICT, delivered through a mobile application called "AI English Teacher," was implemented within a TBLT framework among fifth-grade students at a primary school in a fifth-tier city in China. The intervention aimed to assess the impact of this integrated model on students' English-speaking fluency and anxiety. This study was guided by the following two questions:

(1) To what extent does the integration of AI chatbots enhance English-speaking fluency among primary school students?

(2) To what extent does the integration of AI chatbots reduce English speaking anxiety among primary school students?

C. Population and Sampling

The target population comprised students enrolled in public primary schools operating under resource constraints in northern China. The researcher first selected the school through purposive sampling to ensure that it aligned closely with the study focus and provided a rich classroom setting for observing changes within real instructional constraints. After confirming the site, the researcher employed intact-class sampling and included all students from one naturally formed Grade 5 class (comprising 68 students in total). Because this sample consisted of a single intact class from one specific geographical region, it did not represent a randomised cross-section of the broader student population, which inherently limited the external validity and generalisability of the research findings.

The participants had received formal English instruction since third grade but had no prior experience with AI-based English-speaking applications. Their curriculum was based on the nationally adopted Primary English Textbook published by the People's Education Press (PEP), which integrates task-based learning and interactive dialogue into each unit.

To meet ethical standards and ensure participant protection, the research protocol was reviewed and approved by the participating school's administration prior to data collection. Formal written informed consent was obtained from the students' parents or legal guardians, and verbal assent was provided by the students before the commencement of any research activities. Participation was strictly voluntary, and students and their parents were informed that they could withdraw from the study at any time

without any negative impact on their academic standing or grades.

Strict confidentiality protocols were implemented for data protection. All collected data, including audio recordings of the speaking performance tests and anxiety questionnaires, were immediately anonymised by assigning unique participant identification codes. Raw data were stored on secure, password-protected servers accessible only to the primary researchers. No personally identifiable information was recorded, uploaded, or processed by the Al chatbot application during the intervention, ensuring complete compliance with standard digital data privacy protocols.

D. Research Material

The research utilised a mobile phone application called "AI English Teacher". This educational application was co-developed by Hangzhou Feizhu Technology Co., Ltd.'s Fun Dubbing Team under the pedagogical guidance of a Harvard University Master of Education holder. Its key feature is built-in conversation topics that match the content of PEP textbooks (see Fig. S1 in the Supplementary for an example of this textbook-aligned practice interface). Shanghai Qingpu Concordia Bilingual School and Beijing Normal University Affiliated High School, Changping Campus, have introduced AI tutors into their classrooms.

These AICTs APP are equipped with speech recognition, synthesis, and other functions, including picture-based question answering, though only the conversational practice features were utilised for the purposes of this investigation. Upon receiving spoken English input from the students, the APP can instantly recognise and assess the accuracy of their pronunciation, provide corrective hints for mispronunciation, and suggest additions to incomplete expressions. The avatars communicate with students in a friendly tone, accompanied by simple facial expressions and movements, features deliberately designed to reduce anxiety and lower the affective barrier to spoken practice.

From a technical perspective, the AI chatbot is driven by a hybrid architecture that integrates specific Natural Language Processing (NLP) algorithms with Large Language Model (LLM) capabilities. This underlying LLM infrastructure enables the chatbot to go beyond simple pattern matching in scripted textbook dialogues; it allows the system to comprehend the context, generate dynamically appropriate responses, and provide highly personalised vocabulary scaffolding during open-ended "free talk" scenarios.

Furthermore, a critical technical component of this application is its Automatic Speech Recognition (ASR) system, which has been specifically calibrated to handle the acoustic characteristics of children's voices. Children's speech typically presents significant challenges for standard ASR systems due to higher fundamental frequencies, greater acoustic variability, and developmental articulation patterns. To ensure high reliability, the app's ASR engine is trained on extensive speech corpora of young learners. This targeted acoustic model allows the system to accurately transcribe the speech of fifth-grade students and provide precise real-time phonetic scoring and corrective feedback, thereby minimising technical frustration and ensuring a smooth interactive experience. The APP offers several key functions relevant to this research. First, it provides built-in conversation topics that directly match the content of PEP textbooks, allowing students to practice the core dialogues from their lessons (see Fig. S1 in the Supplementary for an example of this textbook-aligned practice interface). Second, the APP can simulate over 100 real-life scenarios, enabling students to extend their practice to similar but more personalised situational dialogues, such as talking about their favourite idols (Fig. S2 in the Supplementary). Finally, the APP serves as an on-demand linguistic resource, providing direct scaffolding for new vocabulary with definitions, examples, and memory aids in both English and Chinese (Fig. S3 in the Supplementary). Students can use the APP to practice 24/7 conversations with them through tablets or their parents' smartphones.

E. Research Procedures

The research procedure was as follows:

Table 1. The teaching steps of TBLT+AICT

Teaching Steps		How the AI chatbots were used in the TBLT class
Pre-task (5 min)		The English teacher used a screen projector to display the AI English Teacher app on a whiteboard screen, presenting the unit's theme, core vocabulary, and key sentences. The entire class followed the AI chatbots' pronunciation demonstration within the app. The teacher then explained the tasks for the English-speaking lesson and outlined the steps to complete them.
Task Cycle (30 min)	Planning	The students worked in groups of five to discuss their tasks. When they encountered difficulties during the discussion, they could ask the AI chatbots for vocabulary support or explanations of unfamiliar grammatical structures, and may also request example responses or corrections. Subsequently, each group summarised their discussions into key points to prepare for their presentations.
	Report	Each group selected a representative to deliver the English-speaking task to the class. For dialogue tasks, students completed conversations with the assistance of AI chatbots. For paragraph tasks, the teacher recorded the students' speech using AI chatbots. Following the dialogue or recording, the teacher guided the students to review the feedback generated by the AI chatbots, which typically included evaluations of fluency, grammatical accuracy, and other relevant aspects.
	Language focus	Based on the students' presentations and the feedback generated by the AI chatbots, the English teacher guided the class to identify the recurring issues. Students then noted their pronunciation and grammar questions and consulted the AI chatbots again for clarification. The summary explanations from the AI chatbots helped them grasp correct pronunciation, authentic expressions, and other key aspects of language use.
Post-task (5 min)		The teacher invites various students to share their experiences and the benefits of using the chatbot, such as correcting their pronunciation and learning new phrases. The teacher then provides feedback on common concerns and recognises students who exhibit excellent performance or significant growth. After class, teachers assign homework in English-speaking practice, requiring students to complete additional conversational activities at home using the application and submit recordings of the dialogues to the teacher for review.

Pre-test: One week before the intervention, a pre-test consisting of an English fluency assessment and an anxiety questionnaire was administered to all participating students. The assessment took place in a familiar classroom

environment and was supervised by the English teacher and a research assistant. To obtain objective baseline data, students were informed that it was a regular speaking and attitude assessment without being informed of the specifics of the planned intervention.

Intervention: The intervention phase lasted eight weeks and involved task-based instruction supplemented with an AI chatbot. Two 40-minute spoken English classes were conducted each week, with each lesson organised around the theme of the corresponding textbook unit. Table 1 outlines the three-stage teaching procedure for each lesson, comprising the pre-task, during-task, and post-task phases.

Throughout the intervention, the AI chatbot primarily served as a scaffolding aid: when students struggled to complete tasks independently, the chatbot provided prompts or demonstrations to help them complete conversations; when students made errors, the chatbot corrected them immediately; and when students became silent due to nervousness, the chatbot repeated questions or provided prompts in Chinese to guide them in speaking. Researchers actively monitored students' interactions with the chatbot and consulted teachers to measure their adaptation to the chatbot.

Post-test: In week nine (one week after the intervention), the same students underwent an English-speaking fluency test and completed an anxiety questionnaire. To minimise confounding variables, the post-test settings were closely modelled after the pre-test. Notably, the post-test speaking assessment did not use an AI chatbot but instead evaluated students' speech development in a genuine interpersonal communication setting. Consequently, the teachers administered the post-test and assessed the students' fluency performance based on the recordings.

F. Research Instruments

Students' English-speaking fluency was evaluated through a two-part process involving a performance test to elicit speech samples and a scoring rubric to assess them. First, a speaking performance test was administered as a pretest (Supplementary B) and posttest (Supplementary C). The tests were designed to be parallel in structure to ensure comparability, each containing five sections that progressed from structured tasks (e.g., reading sentences and describing pictures) to more spontaneous, open-ended prompts that required students to discuss their personal experiences. This multitask format was intended to capture a comprehensive sample of students' speaking abilities across different communicative demands.

The recorded speech samples from these tests were independently assessed by two veteran English teachers using a structured rubric proposed by Santos and Ramírez-Ávila [22] (see Supplementary D for the complete rubric). This rubric categorises performance across four criteria: verbal productivity (rate), hesitation, repetition, and self-correction. When discrepancies emerged, the raters discussed their evaluations or averaged the scores to reach a consensus. The 'Verbal Productivity' score (adapted from the 'Rate' dimension in the original rubric) reflected the total number of words spoken within the assessment task: a low word count (40–79 words) received one point, a moderate word count (80–119 words) scored 2, and a high word count (120 or more words) earned 3. This study purposefully opted against using standard temporal fluency metrics such as words per minute (WPM) or syllables per second. Because fifth-grade students in this specific resource-constrained setting are beginners who frequently exhibit highly fragmented speech with extended cognitive pauses, strict temporal metrics can easily misrepresent their actual communication efforts. Evaluating overall verbal productivity (total word count) using a structured rubric was deemed a more practical, stable, and developmentally appropriate indicator of their ability to generate meaningful language output.

Hesitation was judged based on the frequency of silent pauses or fillers, such as "um", with higher frequencies resulting in lower scores. Repetition measured how often students reused words or phrases while organising their thoughts, and self-correction captured moments of error correction or restarts. Each aspect was scored on a 1–3 scale, producing a total fluency score out of 12. Prior to the rating, both scorers reviewed sample recordings to align their understanding of the rubric and maintain scoring consistency.

To evaluate students' anxiety when speaking English, the researchers utilised a Chinese-language adaptation of the Foreign Language Classroom Speaking Anxiety Scale (FLCAS), revised to focus on core items relevant to English-speaking expression (see Supplementary E for the full scale). Originally comprising 33 items, Horwitz *et al*. [37]. Originally comprising 33 items by Horwitz *et al*. [37], Apple [74, 75] refined it to 20 items based on Rasch model analysis, all of which focused specifically on English-speaking classroom situations. Additionally, this refined version closely emphasises the instructional context of this study; therefore, the researcher adopted Apple's 20-item version.

A sample item reads, "My heart starts pounding when I think I will have to speak English in class." The cumulative score ranges from 20 to 100, with higher scores indicating more intense speaking anxiety. The instrument was administered one week before and one week after the instructional intervention. To ensure accurate interpretation, the survey was delivered in Chinese and completed in class under the supervision of the English teachers. To establish the reliability of the instrument in the current study context, internal consistency was evaluated prior to the main analysis. The reliability test yielded a Cronbach's alpha coefficient of $\alpha = 0.83$, indicating high internal consistency and confirming that the adapted 20-item FLCAS is a reliable measure of speaking anxiety for this specific sample.

IV. Data Analysis and Results

After data collection, the data were organised and analysed using SPSS 29. Descriptive statistics were first employed on the pretest data to evaluate students' anxiety levels and overall speaking proficiency before the intervention. Next, paired-samples t-tests were utilised to determine whether significant differences existed in anxiety scale and English-speaking fluency scores between the pretest and posttest. Because this study used a one-group pretest-posttest design, paired-samples t-tests were used to assess the statistical significance of mean differences before and after the intervention. If the p-value was less than the specified significance level ($\alpha = 0.05$), the intervention was regarded as

having a significant effect on the indicator. Additionally, Cohen's d effect sizes were calculated to determine the magnitude of the intervention's impact. Cohen's criteria classify effects as small ($d = 0.2$), medium ($d = 0.5$), or large ($d = 0.8$ or above). During the analysis, we verified that the data satisfied the t-test assumptions, such as normality and variance homogeneity. When reporting statistical findings, the researchers included the mean, standard deviation, t-value, degrees of freedom, and p-value.

For the English-speaking fluency score, to ensure scoring objectivity and consistency, two experienced English teachers independently rated the students' speaking recordings using the same rubric. It should be noted that the raters were not blinded to whether the recordings were generated during the pretest or posttest phase. Before formal scoring, a pilot set of 10 recordings was randomly selected for calibration. Both raters scored 10 recordings, and the total scores were further categorised into three levels—high, medium, and low—based on predefined cutoffs. The raters achieved exact agreement on 9 out of 10 recordings (90%), and the remaining case differed by only one adjacent level. Inter-rater agreement was further examined using weighted Cohen's kappa, yielding $\kappa = 0.88$ ($p < 0.001$), indicating excellent agreement and satisfactory rating reliability. After an acceptable agreement was established, the two raters independently scored the remaining recordings to improve efficiency and minimise potential mutual influence. The final speaking score for each student was calculated as the average of the two raters' scores. When discrepancies exceeded a pre-specified threshold (e.g., a difference of ≥X points or ≥1 level), the recording was re-rated through discussion.

A. English-speaking Fluency Outcomes

The results in Table 2 show statistically significant improvements across all four fluency dimensions following the eight-week intervention. However, the magnitude of improvement varied considerably, offering important insights into how AI-assisted TBLT affects different aspects of speaking performance.

Table 2. Paired-samples t-test results for English-speaking fluency

Dimension	Pre-test M (SD)	Post-test M (SD)	ΔM	SDΔ	*t*	*p*	Cohen's *d*
Verbal Productivity	1.07 (0.42)	1.26 (0.51)	0.19	0.448	3.698	< 0.001	0.408
Hesitations	1.11 (0.45)	1.39 (0.58)	0.28	0.528	4.814	< 0.001	0.532
Repetitions	1.29 (0.53)	1.82 (0.67)	0.53	0.571	8.310	< 0.001	0.918
Corrections	1.37 (0.61)	1.89 (0.71)	0.52	0.671	7.079	< 0.001	0.782

Verbal productivity, as measured by the "Rate" dimension, showed modest but significant improvement. The pretest mean of 1.07 increased to 1.26 in the posttest ($t = 3.698$, $p < 0.001$, Cohen's $d = 0.408$). This small-to-medium effect size suggests that the students produced a greater quantity of language after the intervention. However, this measure reflects the amount of speech rather than its speed. The relatively modest gain is unsurprising, given that the underlying automaticity required for faster speech reflects deeply ingrained language-processing capabilities that develop gradually over time.

Hesitation demonstrated moderate improvement, with the mean increasing from 1.11 to 1.39 ($\Delta M = 0.28$, $t = 4.814$, $p < 0.001$, Cohen's $d = 0.532$). This medium effect size indicates that the students reduced their use of fillers and silent pauses, suggesting increased confidence and improved lexical retrieval. However, the moderate effect size reflects the reality that students in under-resourced fifth-tier cities have limited prior exposure to authentic English conversations. Reducing hesitation requires building both linguistic competence and communicative confidence, which are processes that take considerable time.

In contrast, Repetition showed substantial improvement, with the most impressive gains among all measured dimensions. The mean increased from 1.29 to 1.82 ($\Delta M = 0.53$, $t = 8.310$, $p < 0.001$, Cohen's $d = 0.918$), demonstrating a large effect size approaching one standard deviation. This finding suggests that the intervention was particularly effective in helping students organise their thoughts more efficiently and express their ideas more directly. The AI chatbot's immediate feedback helped students recognise unnecessary repetitions and develop strategies for more coherent expressions. Importantly, reducing repetition may be more amenable to conscious self-regulation than increasing speech rate, which explains why students made more dramatic progress in this dimension.

Similarly, Self-Correction demonstrated large improvements, increasing from 1.37 to 1.89 ($\Delta M = 0.52$, $t = 7.079$, $p < 0.001$, Cohen's $d = 0.782$). This medium-to-large effect size indicates that students became more capable of producing accurate language on their first attempt, thereby reducing the need for subsequent corrections. The AI chatbot's consistent corrective feedback throughout the intervention likely helped the students internalise correct forms and patterns, thereby enabling them to produce more accurate initial responses.

These differential effects across the fluency dimensions offer important theoretical and practical insights. The larger improvements in repetition and self-correction compared to rate and hesitation suggest that the metacognitive aspects of fluency, those involving conscious monitoring and adjustment of one's speech, respond more readily to short-term interventions than more automatic processing dimensions. This pattern aligns with the skill acquisition theory, which posits that controlled processes become automatic only through extensive practice over time. The eight-week intervention provided sufficient opportunity for students to develop conscious strategies for reducing repetition and self-correction; however, longer exposure would likely be necessary to significantly impact verbal speech productivity and hesitation patterns.

B. English-speaking Anxiety Outcomes

To investigate the intervention's effectiveness in reducing English-speaking anxiety, we conducted paired-samples t-tests comparing pretest and posttest data across 20 anxiety items. The significance level was set at $\alpha = 0.05$, and

Cohen's d was calculated to measure the actual magnitude of the intervention effects.

The results in Table 3 reveal remarkably consistent and substantial anxiety reduction across all the measured dimensions. All 20 items demonstrated statistically significant decreases ($p < 0.001$), with the majority showing medium-to-large effect sizes. This pattern suggests a substantial overall reduction in anxiety.

The anxiety scale analysis revealed that the most substantial reductions occurred in eleven key areas, all demonstrating large effect sizes ($d > 0.8$). A significant cluster of these items was related to the fear of social comparison, with dramatic decreases in anxiety for statements such as, *I am worried that other students in class speak better than I do* ($d = 1.158$) and *I am worried that my partner speaks better English than I do* ($d = 1.191$). Another area of profound improvement was the fear of negative evaluations from peers. The largest absolute changes were seen in items such as, *I'm afraid my partner will laugh when I speak...* ($d = 1.232$) and *I'm afraid that others in a group discussion will laugh...* ($d = 1.128$). Performance anxiety related to the teacher also decreased significantly. Notably, the item *I tremble when the teacher is about to ask me a question* ($d = 1.388$) had the largest effect size, indicating a reduction in physiological anxiety symptoms. These findings strongly suggest that the AI chatbot's private, non-judgmental interface created a psychologically safe environment for students to practice without the pressure of peer judgment or teacher's observation.

Beyond these, four items registered medium effect sizes ($0.5 \leq d \leq 0.8$), primarily concerning interactions with teachers and general difficulties in expression. Items such as *I'm worried about making mistakes when I speak with the teacher* ($d = 0.625$) and *I'm worried that the teacher will think my speaking is no good* ($d = 0.736$) showed meaningful albeit more moderate reductions. This pattern suggests that while the AI intervention was highly effective in mitigating peer-related anxiety, anxiety stemming from direct evaluation by an authority figure may be more persistent. This indicates that such anxieties might require additional pedagogical strategies beyond those that the AI tool alone can provide.

Finally, five items yielded small but statistically significant effect sizes ($d < 0.5$). The smallest change was observed for the item, *I dislike speaking entirely* ($d = 0.334$), which is likely because students' baseline scores on this item were already low; their issue was not a dislike of speaking but the anxiety associated with it. Similarly, items related to a feeling of panic in specific paired or group settings showed smaller changes. This may be because these structured activities were already perceived as less threatening to students than whole-class performance situations, leaving less room for significant improvement.

Table 3. Paired-samples t-test results for English-speaking anxiety (selected items)

Anxiety Item	ΔM	SDΔ	*t*	*p*	Cohen's d	Effect Size
I'm worried that other students speak better than I do.	0.99	0.853	10.482	< 0.001	1.158	Large
I feel nervous when speaking in front of the entire class.	0.72	0.805	8.089	< 0.001	0.893	Large
I tremble when the teacher is about to ask me a question.	0.87	0.624	12.568	< 0.001	1.388	Large
I am reluctant to express my opinion in a group.	0.78	0.629	11.236	< 0.001	1.241	Large
I'm worried that my partner speaks better than I do.	1.02	0.860	10.782	< 0.001	1.191	Large
I'm afraid my partner will laugh when I speak.	1.73	1.406	11.154	< 0.001	1.232	Large
I'm worried about making mistakes when I speak with a partner.	1.48	1.390	9.615	< 0.001	1.062	Large
Answering a teacher's questions is embarrassing.	0.77	0.775	8.982	< 0.001	0.992	Large
Speaking in front of a group makes me feel self-conscious.	0.93	0.843	9.956	< 0.001	1.099	Large
I'm afraid others in group discussions will laugh.	1.82	1.611	10.213	< 0.001	1.128	Large
I'm worried about making mistakes with the teacher.	0.57	0.917	5.662	< 0.001	0.625	Medium
I'm worried that the teacher will think my speaking is not good.	0.80	1.094	6.664	< 0.001	0.736	Medium
I dislike speaking entirely.	0.12	0.365	3.027	< 0.001	0.334	Small

Note: All 20 items showed statistically significant reductions ($p < 0.001$). The table presents representative items across effect size categories

The overall pattern of anxiety reduction is theoretically significant. The largest reductions occurred precisely where anxiety was most debilitating: fear of peer judgment and social comparison. This suggests that the AI chatbot intervention worked primarily through its unique affordance of providing a completely non-evaluative practice partner. Unlike people, AI chatbots are patient and non-judgmental. This created a safe space for students to practice, make mistakes, and build confidence for real-world interactions. Since all 20 anxiety measures improved, the intervention likely addressed the root cause of the students' anxiety, not just the symptoms. There is evidence consistent with the idea that interacting with the AI chatbot may have supported their overall confidence, potentially leading to a more positive attitude toward speaking English in general.

V. Discussion

This study investigated how an AI chatbot, integrated as a scaffolding tool within a Task-Based Language Teaching framework, might enhance English-speaking fluency and reduce anxiety among fifth-grade Chinese primary school students in an under-resourced educational context. The results of this exploratory study suggest that the AI-assisted TBLT intervention was associated with significant improvements in students' speaking fluency and coherence, alongside substantial reductions in self-reported anxiety. The following sections provide a detailed analysis of these mechanisms and their implications. However, the findings must be interpreted cautiously, given the study's methodological limitations, particularly the absence of a control group that would allow for stronger causal inferences.

A. Analysis of the English-speaking Fluency

The differential effects observed across the four fluency dimensions offer valuable insights into the mechanisms of AI-assisted language learning. The intervention yielded robust improvements in reducing repetition ($d = 0.918$) and self-correction ($d = 0.782$); however, the gains in verbal productivity ("Rate") ($d = 0.408$) and hesitation ($d = 0.532$) were more modest. True speech rate and reduced hesitation

are fundamentally linked to automaticity. This pattern is theoretically significant and warrants further examination.

This finding aligns with cognitive models that distinguish between controlled and automatic language processing. Speech rate and hesitation are fundamentally tied to automaticity, a cognitive state that develops over extensive periods of practice. These skills require rapid lexical retrieval and grammatical processing, which are deeply internalised and thus resistant to short-term change. In contrast, reducing repetition and self-correction can be managed through conscious, strategic control. The AI chatbot's immediate and consistent feedback likely enhanced students' metacognitive awareness of these disfluencies, enabling them to consciously modify their speech patterns within the eight-week intervention period.

From a sociocultural perspective, the chatbot facilitated internalisation by providing repeated, contextualised exposure to correct linguistic forms. The AI's immediate corrective feedback on errors and its acknowledgement of successful utterances operated within each student's zone of proximal development. This scaffolding supported the development of more efficient production strategies, thereby reducing reliance on trial-and-error behaviours that lead to repetition and self-correction.

The substantial reduction in repetitions (d = 0.918) and self-corrections (d = 0.782) provides empirical support for Swain's Output Hypothesis. Within the TBLT task cycle, the AI chatbot acted as an interactive conversational partner that moved students beyond passive comprehension to active production. When students encountered communicative breakdowns, the chatbot's immediate feedback served as a "noticing" mechanism. This forced students to recognize the gap between their intended meaning and their current linguistic output, leading them to consciously modify their speech patterns in subsequent turns to achieve clarity without relying on repetitive hesitation strategies. This "pushed output" transformed static language knowledge into more active, automated speaking skills, directly explaining the improved coherence observed in the post-test result.

Moreover, the AI's consistent feedback may have strengthened the students' internal "monitor", the cognitive mechanism for detecting and correcting speech errors [28]. By consistently highlighting errors, the chatbot likely improved the students' ability to self-monitor, allowing them to prevent mistakes before uttering them rather than correcting them afterward. While the more modest gains in speech rate and hesitation were still statistically significant, they underscore that developing true automaticity requires more prolonged engagement than the duration of this study. The practical implication is that while short-term AI interventions can effectively improve the organisation and coherence of speech, fostering the deep automaticity needed for rapid, pause-free delivery requires sustained, long-term implementation.

B. Analysis of the English-speaking Anxiety

The comprehensive reduction in anxiety observed across all 20 measured items is arguably the study's most significant finding. A closer analysis of these changes is necessary to understand the psychological mechanisms of play. The most substantial reductions were in anxieties related to peer judgment (e.g., fear of being laughed at, d = 1.232) and social comparison (e.g., worry that peers speak better, d = 1.191). This suggests the AI chatbot's primary benefit was the creation of a "zero-judgment practice space."

Unlike human interaction, AI is inherently non-evaluative; it never becomes impatient, never compares a student to others, and treats each interaction as a fresh start. This uniquely safe environment allows for risk-taking and experimentation, as students can make repeated mistakes without the social consequences of embarrassment or criticism from their peers. The remarkably large effect sizes in reducing fear of peer laughter (d = 1.232) and social comparison (d = 1.191) illustrate the mechanics of Krashen's [45] Affective Filter Hypothesis in a digital context. By removing the social-evaluative threat inherent in traditional classrooms, the AI chatbot effectively lowered the students' 'affective filters,' which are typically high in resource-constrained, exam-oriented environments. As this psychological block was bypassed, students became more willing to take communicative risks. This theoretical link explains why the intervention did not just reduce anxiety, but simultaneously allowed for the 'modest but significant' gains in verbal productivity (d=0.408); once the filter was lowered, more comprehensible input could reach the students' internal language acquisition mechanisms.

The implications for foreign language anxiety research are substantial: rather than relying solely on psychological coping strategies to help students manage classroom stress, educators can utilise AI to fundamentally alter the practice environment. By decoupling oral practice from social-evaluative threats, AI provides a crucial intermediary stepping stone, allowing students to build competence in a 'zero-judgment' zone before transitioning to more stressful human-to-human interactions. This process likely involved a positive feedback loop consistent with Bandura's [76] social cognitive theory: practice with the AI lowered anxiety, which enabled more successful communication, which in turn boosted self-efficacy and further reduced anxiety.

Notably, this reduction in anxiety appeared to generalise beyond the AI context, as students also reported feeling less anxious about speaking with teachers and peers. This suggests a fundamental change in their self-efficacy beliefs regarding their speaking abilities, rather than just comfort with the technology. According to Bandura [76], AI provides repeated mastery experiences (successful task completion) and positive reinforcement, which are key sources for developing self-efficacy.

The differential effect sizes also offer insights. The larger reductions in peer-related anxieties compared to the more moderate reductions in teacher-related anxieties suggest that apprehension about authority figures is more deeply ingrained in students. In the Chinese educational context, teacher judgment carries significant weight, and mitigating this specific anxiety may require multifaceted approaches beyond AI practice. Finally, the small effect size for an item like "I dislike speaking entirely" (d = 0.334) is also revealing. This likely showed minimal change because it measures a different construct, fundamental attitude rather than situational anxiety. Most students did not dislike speaking itself; they were anxious about the social evaluation involved. This reinforces the interpretation that the intervention was

most effective in targeting social-evaluative anxiety.

C. Analysis of the AI Chatbot as Dynamic Digital Scaffolding

This study demonstrates that AI chatbots can function as "dynamic digital scaffolds," providing multidimensional support within a task-based framework. This finding extends Vygotsky's concepts of scaffolding and the zone of proximal development to technology-mediated learning and highlights the unique affordances of AI in education.

AI's scaffolding was multifaceted. It provides cognitive scaffolding through vocabulary and sentence model suggestions. It provides effective scaffolding through patient and non-judgmental responses, creating a psychologically safe environment for anxious learners. Finally, it provides linguistic scaffolding in the form of immediate pronunciation and grammar corrections. The term *dynamic* is used because this support was highly adaptive; the AI provided individualised, real-time feedback tailored to each student's specific performance in the moment, a significant advantage over static materials or even a teacher managing a large class.

Furthermore, these findings strongly support Interactionist Theory, which emphasises that language acquisition is driven by communicative interaction and continuous negotiation of meaning. During the task cycles, the AI chatbot functioned as a highly responsive interlocutor. When students encountered communicative breakdowns, the chatbot provided vital interactional modifications—such as repeating prompts, offering linguistic hints, and delivering immediate corrective feedback. This interaction pushed students to notice their linguistic gaps and modify their subsequent output, which explains the significant reductions observed in repetitions and self-corrections. The implications for language acquisition theory are significant: the study demonstrates that human-AI interaction can successfully simulate the essential conditions of human-to-human negotiation of meaning. Consequently, AI chatbots can serve as valid, scalable conversational partners that successfully trigger the interaction-driven mechanisms necessary for second-language acquisition.

However, the AI did not operate in isolation. The human teacher's pedagogical expertise was paramount in designing the learning experience, selecting tasks, introducing the technology, and helping students transfer their learning to authentic human interaction. Therefore, this intervention is best conceptualised as a "triadic partnership" between the teacher, the student, and the AI. In this model, the teacher acts as the instructional architect, the student provides active engagement, and the AI offers scalable, individualised practice and feedback. The positive outcomes observed likely stem not from the AI alone, but from its thoughtful integration into this comprehensive, teacher-led pedagogical design.

This triadic model has broader implications for technology integration. It moves beyond the debate of whether technology can "replace" teachers and instead asks how technology can be used to amplify and extend teachers' expertise. In this framework, AI is not a replacement but a powerful tool that skilled educators can leverage to address persistent instructional challenges, such as providing sufficient individualised speaking practice in resource-constrained classrooms.

From a Sociocultural perspective, the AI chatbot did not merely provide practice; it functioned as a mediating tool that tailored its assistance to the student's Zone of Proximal Development (ZPD). The 'digital scaffolding' provided by the app, such as model sentences and vocabulary hints, offered the 'high support' necessary for students to tackle tasks slightly above their independent proficiency level. The progression from structured textbook dialogues to personalized 'free talk' mirrors the process of internalisation, where the linguistic support initially provided by the AI is gradually absorbed into the student's private speech and eventually their independent cognitive structure.

VI. Research Implications

Despite certain methodological limitations, this study provides valuable insights for educators exploring the integration of AI into primary English education. One key takeaway is the importance of positioning AI chatbots as complements, rather than replacements, for human instruction. The most effective implementation models treat AI as a tool that extends and amplifies teachers' capabilities. Teachers remain essential in setting learning objectives, designing meaningful tasks, offering emotional support, and cultivating a sense of classroom community.

Successful integration also hinges on a triadic partnership between teachers, students, and AI. Educators must guide students in using AI effectively, monitor their interactions to ensure meaningful engagement, and help them transfer skills developed through AI to real-world human communication. This orchestration ensures that AI use is purposeful and pedagogically grounded.

Another notable benefit of AI chatbots is their ability to create a psychologically safe space for language practice. Students often feel more comfortable experimenting in a non-judgmental environment that is free of social consequences. Teachers should frame AI interactions as "mistake-friendly zones", encouraging learners to view errors as valuable learning opportunities rather than sources of embarrassment.

AI tools are most impactful when embedded within structured pedagogical frameworks such as Task-Based Language Teaching (TBLT). These frameworks provide clear objectives and meaningful contexts that give AI interactions coherence and direction. In contrast, the ad hoc use of AI may lack the scaffolding necessary for deep learning.

Moreover, AI chatbots help address a persistent challenge in language education: the limited opportunities for practice. By offering 24/7 access via smartphones or tablets, these tools can significantly increase students' cumulative speaking time, which is particularly beneficial in under-resourced settings where classroom time is constrained.

Ultimately, this study underscores the significance of metacognitive development. Since improvements were most evident in repetition and self-correction, teachers might consider explicitly training students to use AI interactions to build awareness of their fluency patterns. This includes encouraging learners to reduce unnecessary repetitions, improve initial accuracy, and foster more intentional and

reflective language use.

VII. Research Limitations

Although this exploratory study provides promising preliminary evidence regarding the integration of AI-assisted Task-Based Language Teaching, several critical methodological limitations must be explicitly acknowledged to properly contextualise the findings. Foremost among these is the utilisation of a one-group pretest-posttest design without the inclusion of a randomised or matched control group. The complete absence of a comparative baseline inherently weakens any causal claims; consequently, the improvements in spoken fluency and reductions in foreign language anxiety observed in this study can only be interpreted as associative and correlational rather than definitively caused by the artificial intelligence chatbot intervention. This pre-experimental design renders the study highly susceptible to several profound threats to internal validity, which may serve as alternative explanations for the recorded data.

First, the findings are highly vulnerable to the threat of maturation. The participants were fifth-grade students who were undergoing rapid and continuous cognitive, social, and linguistic development. Over the eight-week intervention period, these young learners naturally accumulated broader English exposure, adapted further to their classroom environment, and developed more sophisticated and abstract reasoning skills. Without a control group to isolate the specific impact of the AI chatbot from the simple passage of time, it is entirely plausible that the reductions in repetitive speech and hesitation, as well as the generalised lowering of communicative anxiety, were partially or primarily the result of natural developmental maturation and cumulative academic experience rather than the technological intervention itself.

Second, the study is fundamentally susceptible to the repeated testing effect. The dependent variables were evaluated using a parallel speaking performance test and an identical 20-item Foreign Language Classroom Speaking Anxiety Scale (FLCAS), administered at two distinct temporal points. Exposure to the pretest inherently altered the students' baseline psychological awareness. The objective improvements recorded during the posttest fluency assessment may reflect a practice effect, whereby students simply became more familiar, comfortable, and efficient with the specific mechanics and expectations of the recording process, rather than a true underlying enhancement in spontaneous language mastery. Similarly, completing the anxiety questionnaire during the pretest may have induced metacognitive sensitisation, prompting students to consciously confront and regulate their fears, which could independently lead to lower reported anxiety scores on the posttest regardless of the AI intervention. This is further complicated by the potential for statistical regression to the mean, where initially high anxiety scores naturally lower over time due to statistical variance.

Third, the participants came from a single, intact fifth-grade class (n = 68) in a primary school located in a resource-constrained northern Chinese city. This reliance on one intact class from a single region severely restricts the generalisability of the findings, as several unique contextual factors may have influenced the outcomes. For instance, regarding teacher proficiency, educators in this specific region often face challenges with their own oral English fluency and operate under intense exam-oriented pressure, which has historically limited students' speaking practice. The introduction of the AI chatbot may have produced particularly dramatic improvements in this specific context by filling a profound instructional void; such outsized gains may not be replicated in schools with highly proficient oral English teachers. Furthermore, regarding digital access and school culture, the novelty of utilising mobile applications for interactive learning represents a stark contrast to the traditional reading-focused pedagogy standard in this region. Students in more affluent, technologically integrated schools with established communicative language cultures might not experience the same degree of heightened engagement or anxiety reduction, as the technological and pedagogical shifts would be far less pronounced. Therefore, these findings must be interpreted as being highly context-dependent.

Fourth, the measurement of speaking anxiety relied entirely on student self-reports via the modified FLCAS questionnaire. This method presents a significant limitation when applied to young fifth-grade children, whose metacognitive and self-reflective capacities are still developing. Consequently, their self-assessments may not perfectly reflect their actual psychological state. Furthermore, such self-reported data are highly vulnerable to social desirability bias, as young students may subconsciously report lower anxiety levels to align with perceived researcher or teacher expectations.

Finally, the validity of the outcomes was significantly threatened by the Hawthorne effect and related novelty bias. The introduction of an AI chatbot via mobile tablets represented a stark and highly stimulating departure from traditional instructional paradigms in this resource-constrained setting. The students were acutely aware that they were participating in a novel experiment and were being actively observed by the researchers. This participant reactivity suggests that heightened engagement, increased verbal productivity, and dramatically reported reductions in peer evaluation anxiety may stem primarily from the excitement of using a new digital tool and a subconscious, socially desirable urge to meet perceived researcher expectations. The students may have performed better simply because they knew they were being studied. If the technological novelty wears off in a longer-term longitudinal study, these associative gains may rapidly dissipate.

Consequently, to move beyond associative observations, future research in this domain must employ rigorous experimental architectures, such as randomised controlled trials with active and passive control groups, to neutralise the specific threats of maturation, testing, and participant reactivity, thereby allowing for the definitive investigation of true causal mechanisms in educational technology.

VIII. Future Research Directions

To build on this exploratory work and address its limitations, future research should pursue several key directions. First, rigorous experimental designs are required to isolate the specific contributions of AI technology. Future

research should employ randomised controlled trial designs with comparison groups receiving (1) traditional TBLT without AI, (2) increased speaking practice without TBLT structure, and (3) no intervention. Such designs would enable researchers to disentangle the unique contributions of AI technology from those of increased practice, structured pedagogy, and teacher attention. These studies should involve diverse student populations across multiple schools and contexts to enhance generalisability.

Longitudinal studies are also essential in this regard. Tracking students over one or more academic years would help determine whether initial improvements are sustained, whether continued AI use leads to diminishing or increasing returns, and how long students need to engage with AI tools to achieve stable and lasting gains. To overcome the limitations of relying solely on young learners' self-reports and mitigate potential social desirability bias, future studies may employ data triangulation. Researchers should combine quantitative survey data with qualitative and observational evidence, such as in-depth student interviews and systematic classroom observations by teachers. Using these mixed-method approaches to triangulate the data will provide a more objective, robust, and comprehensive understanding of how AI interventions truly impact foreign language speaking anxiety.

Qualitative methods, such as interviews, focus groups, and classroom observations, can illuminate how and why students experience reduced anxiety and improved fluency. Understanding students' subjective experiences, the AI features they find most helpful, and the psychological processes driving change would inform the design of more effective interventions. A mixed-methods approach that combines quantitative outcomes with rich qualitative data would be particularly valuable.

While the current study evaluated independent changes in speaking performance and psychological states, future research should conduct detailed correlation analyses to examine the direct statistical relationship between individual gains in English-speaking fluency and corresponding reductions in speaking anxiety.

Research on individual differences is another important avenue. Investigating how traits such as personality, prior proficiency, anxiety type, learning style, and digital literacy moderate the effects of interventions could enable personalised and targeted implementation. As AI chatbot technologies continue to evolve, comparative studies are required to examine the different systems, features, and implementation strategies. These studies should explore the ideal frequency and duration of AI interaction, the most effective task types, preferred AI personality and feedback styles, and how best to sequence AI and human interaction for maximum benefits.

Ultimately, teacher professional development research is crucial for the successful implementation of large-scale initiatives. Studies should explore how educators can prepare for AI-augmented instruction, including the necessary knowledge, skills, beliefs, and support structures. Understanding teachers' learning processes, common challenges, and effective professional development models is key to sustaining innovation in practice.

IX. Conclusion

This study demonstrates the potential of integrating AI chatbots as "intelligent scaffolds" within a Task-Based Language Teaching (TBLT) framework to address chronic challenges in primary English education. The findings highlight a critical pedagogical shift: the transition from teacher-centered knowledge transmission to an interaction-driven, triadic partnership involving the teacher, student, and AI.

A significant contribution of this research is its demographic focus on young learners in resource-constrained settings. While previous AI research has focused on higher education, this study provides a scalable, low-cost model that demonstrates how conversational AI can bridge the gap between structured textbook input and spontaneous oral output. By functioning as a dynamic scaffold, the AI provides the "high-support, low-stakes" environment necessary for fifth-grade students to navigate their Zone of Proximal Development without the fear of social evaluation.

Furthermore, the comprehensive reduction in speaking anxiety across all dimensions suggests that the psychological affordances of AI, specifically the creation of a "zero-judgment practice space", may be its most vital contribution to language learning. By decoupling oral practice from peer-related social threats, AI tools allow students to build the foundational confidence required for real-world human interaction.

Ultimately, the future of language instruction in under-resourced contexts lies in the synergistic combination of human expertise and technological scalability. This study offers a blueprint for this synergy, maintaining the teacher's central role as an instructional architect while leveraging AI to provide the individualized, high-frequency practice that traditional classrooms often lack. Subsequent research should now move toward longitudinal experimental designs to validate the long-term sustainability of these associative gains.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Zhai Xin conducted the research, analyzed the data, and wrote the paper. Siti Nazleen Abdul Rabu revised and polished the paper. All authors approved the final version.

References

[1] D. M. Gustina, N. Mariana, and Wiryanto, "Augmented reality-based ethnomathematics learning media to enhance spatial ability in 3D geometry for fifth grade elementary students," *Journal of Innovation and Research in Primary Education*, vol. 4, no. 2, pp. 273–280, 2025. doi: 10.56916/jirpe.v4i2.1229

[2] J. Piaget and B. Inhelder, *The Psychology of the Child,* New York, NY, USA, 1972.

[3] E. C. Lumi, C. N. Wuntu, and O. Rorintulus, "The effectiveness of phonics method in improving English pronunciation to 5th grade students at SD GMIM IV Tomohon," *Santhet: Jurnal Sejarah, Pendidikan dan Humaniora*, vol. 9, no. 3, pp. 824–829, 2025. doi: 10.36526/js.v3i2.5476

[4] E. Kurniati and M. Zaim, "The effectiveness of audio media for English learning based on scripted song at the fifth grade of elementary school," *Ilkogretim Online - Elementary Education Online*, vol. 20, no. 1, pp. 208–216, 2021. doi: 10.17051/ilkonline.2021.01.023

[5] J. Tang, "A policy and practice review of English education in Chinese primary and secondary schools", *Pacific International Journal*, vol. 8, no. 4, pp. 142–149, 2025. doi: 10.55014/pij.v8i4.861

[6] R. Wang, "Influencing factors and countermeasures of Chinese students' oral English proficiency," *Journal of Education, Humanities and Social Sciences*, vol. 13, pp. 158–164, 2023. doi: 10.54097/ehss.v13i.7887

[7] P. Chea, "Exploring the challenges in Chinese EFL learners' spoken English: beyond the surface," *International Journal of English Learning and Applied Linguistics (IJELAL)*, vol. 5, no. 2, pp. 135–155, 2025. doi: 10.21111/ijelal.v5i2.14361

[8] S. Lin, "Examining factors of speaking anxiety among Chinese EFL students in higher education," *Journal of Education Humanities and Social Sciences*, vol. 13, pp. 80–86, May 2023. doi: 10.54097/ehss.v13i.7858

[9] S. Liu, "Analysis of phenotypic factors and research on countermeasures of Chinese secondary school students' anxiety in oral learning," *Lect. Notes Educ. Psychol. Public Media*, vol. 75, no. 1, pp. 64–69, Nov. 2024. doi: 10.54254/2753-7048/75/20241074

[10] J. Hu and J. Liu, "The survey on students' emotional state in oral English learning of junior middle school—taking No.11 junior middle School of Yichang city as an example," *Res. Adv. Educ.*, vol. 2, no. 2, pp. 55–65, Feb. 2023. doi: 10.56397/rae.2023.02.07

[11] Y. Huang, "Exam-oriented education and quality-oriented education in China: A comparative analysis taking English education as an example," *Lect. Notes Educ. Psychol. Public Media*, vol. 36, no. 1, pp. 125–130, Jan. 2024. doi: 10.54254/2753-7048/36/20240439

[12] E. I. Safitri, S. Hidayati, and D. S. Ciptaningrum, "The impact of AI chatbots on english language learners' speaking proficiency: A systematic review," *Journal of Research on English and Language Learning (J-REaLL)*, vol. 6, no. 2, pp. 317–329, Aug. 2025. doi: 10.33474/j-reall.v6i2.23866

[13] J. Taeza, "The role of AI-powered chatbots in enhancing second language acquisition: An empirical investigation of conversational AI assistants," *Edelweiss Appl. Sci. Technol.*, vol. 9, no. 3, pp. 2616–2629, Mar. 2025. doi: 10.55214/25768484.v9i3.5853

[14] T.-Y. Tai and H. H.-J. Chen, "Improving elementary EFL speaking skills with generative AI chatbots: exploring individual and paired interactions," *Computers & Education*, vol. 220, pp. 1–14, Oct. 2024. doi: 10.1016/j.compedu.2024.105112

[15] M. Mingyan, N. Noordin, and A. B. Razali, "Effects of an AI chatbot mobile application on foreign language anxiety among Chinese EFL undergraduates," *International Journal of Academic Research in Progressive Education and Development*, vol. 13, no. 4, pp. 3828–3839, Dec. 2024. doi:10.6007/IJARPED/v13-i4/23894

[16] C. Wang, B. Zou, Y. Du, and Z. Wang, "The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence," *System*, vol. 127, no. 2, 103533, Dec. 2024. doi: 10.1016/j.system.2024.103533

[17] G. L. Bagasala and M. L. Estremera, "A systematic literature review of Task-Based Language Teaching (TBLT) in L2 context," *International Journal for Multidisciplinary Research (IJFMR)*, vol. 7, no. 3, pp. 1–12, 2025. doi: 10.36948/ijfmr.2025.v07i03.44186

[18] M. Michel, "Task-based language teaching TBLT," *ELIA*, vol. 127, no. 23, pp. 317–328, 2023. doi: 10.12795/elia.2023.i23.10

[19] L. S. Vygotsky, *Mind in Society: Development of Higher Psychological Processes*, M. Cole, V. John-Steiner, S. Scribner, and E. Souberman, Eds., Cambridge, MA, USA: Harvard University Press, 1978. doi: 10.2307/j.ctvjf9vz4

[20] C. J. Fillmore, D. Kempler, and W. S.-Y. Wang, *Individual Differences in Language Ability and Language Behavior*, New York, NY, USA: Academic Press, 1979.

[21] P. Lennon, "Investigating fluency in EFL: A quantitative approach," *Lang. Learn.*, vol. 40, no. 3, pp. 387–417, Sep. 1990. doi: 10.1111/j.1467-1770.1990.tb00669.x

[22] J. Santos and M. Ramírez-Ávila, "Santos Improving speaking fluency through 4/3/2 technique and self-assessment," *Teach. Engl. Second Foreign Lang.-TESL-EJ*, vol. 26, no. 4, Aug. 2022. doi: 10.55593/ej.26102a1

[23] P. Soohwan, "Measuring fluency: Temporal variables and pausing patterns in L2 English speech," Ph.D. dissertation, Dept. Elect. Eng., Purdue University, Indiana, United States, 2016.

[24] H. R. Bosker, "The processing and evaluation of fluency in native and non-native speech," Ph.D. dissertation, Utrecht Univ., Utrecht, Netherlands, 2014.

[25] S. Aubrey, "The relationship between anxiety, enjoyment, and breakdown fluency during second language speaking tasks: an idiodynamic investigation," *Front. Psychol.*, vol. 13, Aug. 2022. doi: 10.3389/fpsyg.2022.968946

[26] P. Peltonen, S. Olkkonen, M. Szyszka, and P. Lintunen, "L2 repair fluency through the lenses of L1 repair fluency, cognitive fluency, and language anxiety," *Appl. Linguist. Rev.*, vol. 16, no. 2, pp. 877–899, Mar. 2025. doi: 10.1515/applirev-2023-0011

[27] H. Chang and S. Windeatt, "Fluency issues in L2 academic presentations: Linguistic, cognitive and psychological influences on pausing behaviour," *Lang. Teach. Res.*, Feb. 2024. doi: 10.1177/13621688241233718

[28] W. Levelt, "Monitoring and self-repair in speech," *Cognition*, vol. 14, no. 1, pp. 41–104, Jul. 1983. doi: 10.1016/0010-0277(83)90026-4

[29] N. Segalowitz, *Cognitive Bases of Second Language Fluency*, New York, NY, USA: Routledge, 2010.

[30] T. Houn and S. Em, "Common factors affecting grade-12 students' speaking fluency: A survey of cambodian high school students," *J. -Salam*, vol. 6, no. 1, pp. 11–24, Jun. 2022. doi: 10.37249/assalam.v6i1.360

[31] F. Guo *et al*., "Blue book report | 2024 survey report on adolescent mental health and academic status," *China.com.cn – Psychology*, May 27, 2025.

[32] S. Choudhury and V. K. Chechi, "Tracing the relationship between academic anxiety and academic performance among the upper primary school students," *ECS Trans.*, vol. 107, no. 1, pp. 10085–10095, Apr. 2022. doi: 10.1149/10701.10085ecst

[33] M. N. Khalid, F. Shafiq, and M. Ahmad, "Anxiety and academic performance of undergraduate university students," *Glob. Soc. Sci. Rev.*, vol. VII, no. II, pp. 372–386, Jun. 2022. doi: 10.31703/gssr.2022(vii-ii).37

[34] L. Ăskaree, E. Tariq, and S. Saeed, "Navigating pressure: The link between anxiety and academic performance among university students," *Soc. Sci. Rev. Arch.*, vol. 3, no. 2, pp. 588–603, May 2025. doi: 10.70670/sra.v3i2.644

[35] X. Wang, "Analysis of the causes of chinese high school students' english-speaking anxiety," *J. Educ. Humanit. Soc. Sci.*, vol. 13, pp. 251–256, May 2023. doi: 10.54097/ehss.v13i.7904

[36] T. B. Dan, N. Said, and N. S. Sukor, "English speaking anxiety of Chinese middle school students: A perspective based on students' personality differences," *J. Southeast Asia Psychol. SAPJ*, vol. 12, no. 1, 16, May 2024. doi: 10.51200/sapj.v12i1.5057

[37] E. K. Horwitz, M. B. Horwitz, and J. Cope, "Foreign language classroom anxiety," *Mod. Lang. J.*, vol. 70, no. 2, pp. 125–132, Jun. 1986. doi: 10.1111/j.1540-4781.1986.tb05256.x

[38] J. C. McCROSKEY, "Oral communication apprehension: a summary of recent theory and research," *Hum. Commun. Res.*, vol. 4, no. 1, pp. 78–96, Sep. 1977. doi: 10.1111/j.1468-2958.1977.tb00599.x

[39] D. Watson and R. Friend, "Measurement of social-evaluative anxiety," *J. Consult. Clin. Psychol.*, vol. 33, no. 4, pp. 448–457, 1969. doi: 10.1037/h0027806

[40] D. Sari, "Speaking anxiety as a factor in studying EFL," *English Education Journal*, vol. 8, no. 2, pp.177–186, Apr. 2017.

[41] R. J. Daymiel, J. Cantina, V. R. I. Alipoyo, M. O. Comecilla, A. S. Patay, and J. T. Recamara, "Anxiety in second language in relation to students' speaking performance," *Sprin J. Arts Humanit. Soc. Sci.*, pp. 396–410, Aug. 2022. doi: 10.55559/sjahss.v1i08.41

[42] A. V. Kuswantoro and D. Novita, "Overcoming speech anxiety in english language classrooms," *Pubmedia J. Pendidik. Bhs. Ingg.*, vol. 1, no. 3, 10, May 2024. doi: 10.47134/jpbi.v1i3.419

[43] S. R. Que, P. A. Tiven, and T. Warbanaran, "Exploring anxiety in EFL students' speech production," *MATAI Int. J. Lang. Educ.*, vol. 5, no. 1, pp. 79–86, Dec. 2024. doi: 10.30598/matail.v5i1.16558

[44] Y. V. T. Hai, A. N. Quỳnh, and L. T. Thị, "Study on speaking anxiety of senior students in the process of English language learning at Hanoi Metropolitan University, Vietnam," *Sci. J. TAN TRAO Univ.*, vol. 10, no. 3, Jan. 2025. doi: 10.51453/2354-1431/2024/1210

[45] S. D. Krashen, *Principles and Practice in Second Language Acquisition*, Oxford, New York: Pergamon, 1982.

[46] N. S. Prabhu, *Second Language Pedagogy*, Oxford: Oxford Univ. Press, 1994.

[47] D. Nunan, *Task-Based Language Teaching*, Cambridge, U.K.: Cambridge Univ. Press, 2004. doi: 10.1017/CBO9780511667336

[48] R. Ellis, *Task-Based Language Learning and Teaching*, Oxford, U.K.: Oxford Univ. Press, 2003.

[49] R. Ellis, "Task-based language teaching: Sorting out the misunderstandings," *Int. J. Appl. Linguist.*, vol. 19, no. 3, pp. 221–246, Nov. 2009. doi: 10.1111/j.1473-4192.2009.00231.x

[50] J. Willis, *A Framework for Task-Based Learning*, Harlow, Essex, U.K.: Longman, 1996.

[51] A. Alshahrani and W. Saud, "The effectiveness of implementing task-based language teaching in Saudi secondary schools to improve students' oral fluency," *J. World Englishes Educ. Pract.*, vol. 6, no. 3, pp. 76–91, Nov. 2024. doi: 10.32996/jweep.2024.6.3.8

[52] S. H. M. El-Boraie, "Utilizing a blended strategy based on task-based language learning to develop primary stage pupils' EFL oral communication skills," *Journal of Reading and Knowledge*, vol. 22, no. 254, pp. 79–93, Dec. 2022. doi: 10.21608/mrk.2022.276603

[53] Z. Omar, N. Jamaludin, and M. Arshad, "Task-based language teaching to develop Malay language speaking skills of primary 5 Singaporean students," *Int. J. Educ.*, vol. 14, no. 1, pp. 48–61, Apr. 2021. doi: 10.17509/ije.v14i1.27140

[54] E. Kasneci, K. Sessler, F. Fischer, U. Gasser, and G. Groh, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, pp. 1–13, Jan. 2023. doi: 10.1016/j.lindif.2023.102274

[55] M. Bobula, "Generative Artificial Intelligence (AI) in higher education: A comprehensive review of challenges, opportunities, and implications," *Journal of Learning Development in Higher Education*, pp. 1–27, 2024. doi: 10.47408/jldhe.vi30.1137

[56] N. H. V. Nguyen and V. P. H. Pham, "AI chatbots for language practices," *Int. J. AI Lang. Educ.*, vol. 1, no. 1, pp. 56–67, Sep. 2024. doi: 10.54855/ijaile.24115

[57] N. H. Anh, A. Baharloo, and G. R. Pasaribu, "Integrating AI-based tools to enhance speaking fluency in EFL classrooms in Labuhanbatu Utara," *J. Educ.*, vol. 3, no. 1, pp. 1–10, 2025

[58] S. Shikun, G. Grigoryan, N. Huichun, and H. Harutyunyan, "AI chatbots: developing English language proficiency in EFL classroom," *Arab World Engl. J.*, vol. 1, no. 1, pp. 292–305, Apr. 2024. doi: 10.24093/awej/ChatGPT.20

[59] Z. Susoy, "Reducing anxiety and enhancing performance: The impact of AI chatbots versus human facilitation on EFL speaking assessment outcomes," *Front. Psychol.*, vol. 16, 1745942, Jan. 2026. doi: 10.3389/fpsyg.2025.1745942

[60] D. Ding and A. Muhyiddin B Yusof, "Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: a mixed methods study," *Humanit. Soc. Sci. Commun.*, vol. 12, no. 1, 1223, Aug. 2025. doi: 10.1057/s41599-025-05550-z

[61] N. T. Huong, "The impact of AI chatbot integration on Vietnamese EFL learners' speaking performance, fluency, and confidence," *Lang. Technol. Soc. Media*, vol. 4, no. 1, pp. 70–88, 2026. doi: 10.70211/ltsm.3026-7196.314

[62] M. E. Poehner, "The origins of dynamic assessment: Sociocultural theory and the zone of proximal development," *Dynamic Assessment: A Vygotskian Approach to Understanding and Promoting L2 Development*, Boston, MA: Springer US, 2008, pp. 23–42. doi: 10.1007/978-0-387-75775-9_2

[63] C. S. Adhikari, P. Singh, V. Kumar, Arsalan, Amarnath, Mimoh, S. Yadav, U. Hussain, and M. A. Khan, "Designing and detailing of scaffoldings," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 5, no. 3, pp. 2045–2052, 2019.

[64] D. Wood, J. S. Bruner, and G. Ross, "The role of tutoring in problem solving," *J. Child Psychol. Psychiatry*, vol. 17, no. 2, pp. 89–100, Apr. 1976. doi: 10.1111/j.1469-7610.1976.tb00381.x

[65] A. Shvarts and A. Bakker, "The early history of the scaffolding metaphor: Bernstein, Luria, Vygotsky, and before," *Mind Cult. Act.*, vol. 26, no. 1, pp. 4–23, Jan. 2019. doi: 10.1080/10749039.2019.1574306

[66] K. Wilson and L. Devereux, "Scaffolding theory: high challenge, high support in Academic Language and Learning (ALL) contexts," *Journal of Academic Language & Learning*, vol. 8, no. 3, pp. 91–100, Sep. 2014.

[67] N. Dabbagh, "Scaffolding: An important teacher competency in online learning," *Techtrends*, vol. 47, no. 2, pp. 39–44, Mar. 2003. doi: 10.1007/bf02763424

[68] L. D. León, "Model of models: preservice teachers in a vygotskian scaffold," *Educ. Forum*, vol. 76, no. 2, pp. 144–157, Apr. 2012. doi: 10.1080/00131725.2011.653093

[69] R. Alkhudiry, "The contribution of Vygotsky's sociocultural theory in mediating L2 knowledge Co-construction," *Theory Pract. Lang. Stud.*, vol. 12, no. 10, pp. 2117–2123, Sep. 2022. doi: 10.17507/tpls.1210.19

[70] J. M. Alharbi, "Insight into the role of interaction in language acquisition: Vygotsky's interactionist theory of language," *Arab World Engl. J.*, vol. 14, no. 2, pp. 281–294, Jun. 2023. doi: 10.24093/awej/vol14no2.20

[71] H. Mahn, "Vygotsky and second language acquisition," *The Encyclopedia of Applied Linguistics*, 1st ed. Wiley, pp. 1–8, July 2020. doi: 10.1002/9781405198431

[72] K. Culligan, "The relationship between language and thought: Exploring Vygotsky and sociocultural approaches to second language research," *The Atlantic Journal of Graduate Studies in Education: Special Edition*, pp. 281–294, 2013.

[73] J. P. Lantolf, "Intrapersonal communication and internalization in the second language classroom," in *Vygotsky's Educational Theory in Cultural Context*, A. Kozulin, B. Gindis, V. S. Ageyev, and S. M. Miller, Eds., Cambridge: Cambridge University Press, Sep. 2003, pp. 349–370. doi: 10.1017/CBO9780511840975.018

[74] C. E. Leslie, "Interaction patterns and support for learning in the primary foreign language classroom," *Porta Linguarum Rev. Interuniv. Didáctica Las Leng. Extranj.*, no. 36, pp. 65–82, Jun. 2021. doi: 10.30827/portalin.v0i36.15712

[75] M. Apple, "Using rasch analysis to create and evaluate a measurement instrument for foreign language classroom speaking anxiety," *JALT J.*, vol. 35, no. 1, 5, May 2013. doi: 10.37546/JALTJJ35.1-1

[76] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change," *Psychological Review*, vol. 84, no. 2, pp. 191–215, 1977.