International Journal of TESOL Studies (2026)
261125, 1-17 https://doi.org/10.58304/ijts.261125

Article

Scaffolding L2 Learners' Critical Thinking Tendency and Achievement Emotions: Students' Experience of a Three-Stage AI-assisted Speaking Pedagogy

Molly Xie Pan
College of Professional and Continuing Education, The Hong Kong Polytechnic University, Hong Kong SAR, China

Di Zou*
The Hong Kong Polytechnic University, Hong Kong SAR, China

Dandan Xie
Shenzhen Polytechnic University, China

Received: 21 May 2026/Received in revised form: 16 September 2026/Accepted: 22 October 2026/Available online: 4 November 2026

Abstract
While artificial intelligence (AI) tools offer rich input for language learning, few studies examine their role in driving higher-order cognitive processes, integrated pedagogical cycles, and broader achievement emotions. Addressing these gaps, this study investigated a scaffolded three-stage AI-assisted speaking pedagogy (preparation, practice, and reflection) among 13 tertiary L2 learners (CEFR A2 speaking equivalent). Grounded in an integrated Community of Inquiry (CoI) and Control-Value Theory (CVT) framework, semi-structured interview data revealed that teaching, social, and cognitive presences form a dynamic, recursive system shaping students' internal appraisals. Specifically, teaching presence established a supportive structural foundation, human-AI social presence functioned as an emotional engine buffering anxiety, and higher-order cognitive presence (integration) acted as a catalyst for high-value appraisals. Ultimately, this study proposes an integrated theoretical model showing how pedagogical scaffolding and dialogic AI interaction may collaboratively transform linguistic friction into empowering cognitive growth, offering practical insights for AI-driven language instruction.

Keywords
AI-assisted speaking, critical thinking tendency, teaching presence, social presence, achievement emotions

*Corresponding author. Email: daisy.zou@polyu.edu.hk

Online First View

1 Introduction

1.1 Challenges of English speaking to L2 learners

English speaking remains one of the most cognitively and affectively demanding skills for second language (L2) learners. Unlike receptive skills such as reading and listening, speaking requires real-time processing, simultaneous retrieval of lexical and syntactic resources, and the ability to maintain coherent discourse under time pressure (Bygate, 1987; Levelt, 1993). For many learners, the transition from knowing a language to actually using it in spoken interaction constitutes a persistent bottleneck, particularly in English as a foreign language (EFL) context where authentic opportunities for oral production are often limited to formal classroom settings (Goh & Burns, 2012). This gap between declarative knowledge and procedural fluency has been well documented in the literature. Learners may demonstrate adequate grammatical competence in written tasks yet struggle to sustain even brief spoken exchanges when required to produce language spontaneously (DeKeyser, 2007).

The difficulty of L2 speaking is compounded by its inherently social and evaluative nature. Speaking exposes learners not only to the cognitive demands of language processing but also to the psychological risk of being judged by interlocutors, teachers, and peers (Horwitz et al., 1986). Foreign language classroom anxiety, first conceptualized by Horwitz et al. (1986), has consistently been found to correlate negatively with speaking performance, as anxious learners tend to engage in avoidance behaviors, produce shorter and simpler utterances, and exhibit reduced willingness to communicate (WTC) (MacIntyre & Gardner, 1994; Liu & Jackson, 2008). More recent work grounded in control-value theory (CVT; Pekrun, 2006) has expanded the emotional landscape beyond anxiety, showing that a range of achievement emotions, including shame, hopelessness, enjoyment, and pride, are shaped by learners' appraisals of their own competence (control) and the significance they attach to the task (value) (Pekrun et al., 2011; Shao et al., 2019). In the context of L2 speaking, low perceived control over one's oral production and high perceived social-evaluative stakes can generate a self-reinforcing cycle of negative emotions that undermines both willingness to participate and the quality of spoken output.

Critical thinking adds another layer of complexity to L2 speaking. Effective oral communication requires learners not only to produce grammatically acceptable utterances but also to organize ideas logically, evaluate evidence, construct arguments, and respond to counter-positions—all while managing the cognitive load of operating in a non-native language (Facione, 1990; Elder & Paul, 2020). Research has shown that L2 learners frequently resort to simplified, formulaic oral production precisely because the simultaneous demands of critical reasoning and linguistic encoding exceed their available cognitive resources (Skehan, 1998). This suggests that speaking instruction designed to promote critical thinking tendency must address both the linguistic and the metacognitive prerequisites for reasoned oral expression. However, the intersection of critical thinking tendency, achievement emotions, and L2 speaking pedagogy remains under-theorized, particularly in technology-enhanced learning environments. While recent studies have begun to explore how critical thinking dispositions relate to learners' emotional experiences in L2 contexts (Ghanizadeh, 2017; Namaziandost et al., 2025), the specific mechanisms through which AI tools can scaffold both critical thinking tendency and adaptive achievement emotions during speaking practice have yet to be further examined.

1.2 Research on AI-assisted English-speaking practice

The rapid advancement of artificial intelligence (AI) technologies, particularly natural language processing (NLP) and large language models (LLMs), has opened new possibilities for L2 speaking instruction. AI-powered tools can now serve as conversational partners, pronunciation coaches, and real-time feedback providers, offering learners individualized, on-demand speaking practice outside the constraints of the traditional classroom (Jeon, 2024; Tai & Chen, 2023). Intelligent personal

Online First View

assistants (IPAs) such as Amazon Alexa and Google Assistant, as well as dedicated language learning chatbots, have been shown to reduce learners' speaking anxiety by providing a non-judgmental interaction space where errors carry low social risk (Dizon, 2020; Moussalli & Cardoso, 2020). Studies report that learners who engage with AI speaking tools tend to produce more output, take greater linguistic risks, and develop higher levels of WTC compared to those limited to peer-only interaction (Ebadi & Ebadijalal, 2022; El Shazly, 2021).

Beyond anxiety reduction, AI tools offer affordances that align with well-established principles of L2 speaking development. The provision of immediate, individualized feedback on pronunciation, grammar, and lexical choice supports the noticing function that is central to output-based theories of acquisition (Swain, 2005). The availability of repeated, self-paced practice aligns with skill acquisition theories emphasizing the role of proceduralization and automatization (DeKeyser, 2007). Furthermore, AI-mediated interaction can extend learners' opportunities for meaning-focused output, a key condition for speaking development according to both interactionist (Long, 1996) and sociocultural perspectives (Lantolf, 2006). Recent studies have also highlighted the potential of LLM-based tools such as ChatGPT to support pre-task planning, vocabulary enrichment, and content organization, thereby strengthening the preparatory phase of speaking tasks (Kohnke et al., 2023; Yan, 2023).

Despite these promising developments, several important gaps persist in the literature. First, the majority of studies on AI-assisted speaking have focused on linguistic outcomes like pronunciation accuracy, fluency, and grammatical correctness, while paying limited attention to higher-order cognitive processes such as critical thinking (Zhang & Zou, 2022; Zou, Huang, & Xie, 2021). Although AI tools can provide rich linguistic input and interactive practice, the manner in which they might scaffold learners' ability to analyze, evaluate, and construct reasoned arguments during oral production remains largely unexplored. This is a significant oversight, given the increasing emphasis on critical thinking as a central competency in language education worldwide (Li, 2016; Yuan & Stapleton, 2020).

Second, existing research has tended to examine AI tools in isolation—a single chatbot or speech recognition application—rather than as part of an integrated pedagogical design that spans the full cycle of speaking preparation, practice, and reflection. This fragmented approach fails to capture how different AI tools with complementary affordances might be orchestrated to support learners across multiple phases of the speaking process. As Li et al. (2025) have argued, technology-supported practice is most productive when paired with metacognitive scaffolding rather than treated as stand-alone exposure.

Third, the affective dimension of AI-assisted speaking, particularly achievement emotions beyond anxiety, has received insufficient empirical attention. While anxiety has been a dominant construct in L2 speaking research (Teimouri et al., 2019), the broader spectrum of achievement emotions posited by CVT, including enjoyment, hope, pride, boredom, and hopelessness, offers a more nuanced lens for understanding learners' emotional experiences during AI-mediated speaking practice. How different phases of AI-supported pedagogy differentially shape these emotions, and how these emotions in turn influence learners' cognitive engagement patterns and speaking outcomes, represents a critical but under-researched area.

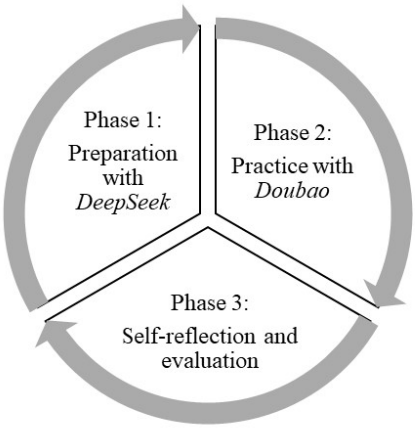
1.3 A three-stage AI-assisted speaking pedagogy

To address the aforementioned gaps, this study proposes and investigates a three-stage AI-assisted speaking pedagogy that integrates distinct AI tools into a coherent instructional cycle: preparation, practice, and reflection, see Figure 1. The model assumes that speaking development depends on support before performance (planning and language resources), during performance (opportunities for output and feedback), and after performance (consolidation and feedforward). This sequencing is consistent with work arguing that technology-supported practice is most productive when paired with metacognitive scaffolding rather than treated as stand-alone exposure (Ghanizadeh, 2017; Namaziandost et al., 2025; Li et al., 2025).

Online First View

Figure 1

A Three-stage AI-assisted Speaking Model



Phase 1: AI-Assisted Preparation (DeepSeek). In this phase, students use DeepSeek for pre-task planning and linguistic preparation. DeepSeek was selected because it is freely available and readily accessible to students in mainland China, and it can provide real-time, individualized support for language learning (Gao et al., 2025). Students are instructed to use the tool to improve content, identify topic-related vocabulary and expressions, explore sentence structures, and refine their outlines. Crucially, students are explicitly instructed not to use DeepSeek to generate complete paragraphs or scripts for later recitation. Instead, DeepSeek is positioned as a tool for content improvement and language upgrading, with learners expected to internalize and selectively use the suggested resources in subsequent speaking practice. Specifically, students are encouraged to first draft a short text (typically around 100 words) based on the speaking task requirements and then use DeepSeek to obtain immediate, individualized feedback on vocabulary, grammar, sentence structures, and coherence. Based on the suggestions, students revise their draft into a "speak-ready" version by selecting a small set of accurate, high-utility expressions, sentence patterns, transitions, and key lexis to carry into subsequent speaking practice. The instructional aim of this phase is not to produce a polished composition but to reduce the barriers of "I don't know what to say" and "I don't have the right words" by strengthening idea clarity and linguistic readiness.

Phase 2: AI-Powered Speaking Practice (*Doubao*). In this phase, students use *Doubao* as a spoken interlocutor through its real-time voice call mode. Students can access the voice mode through the application's interface and practice short dialogues. *Doubao*'s support for adjustable speech delivery (e.g., speed, tone, and expressiveness) is designed to make the interaction accessible to students at different proficiency levels. The pedagogical aim of this phase is to provide students with a low-stakes, accessible speaking partner through which they can apply linguistic resources prepared in Phase 1 and increase opportunities for oral production. Students are encouraged to respond in complete turns, extend their responses beyond minimal answers, and ask simple follow-up questions to sustain interaction. *Doubao* maintains momentum through prompts and follow-up questions and provides brief, selective corrective feedback on salient issues (e.g., tense, prepositions, collocations, and plural forms) without significantly disrupting fluency. Students are advised to recycle corrected forms in subsequent turns, creating a rapid revise-and-retry loop that promotes noticing and uptake.

Phase 3: Structured Reflection via *Xuexitong*. In this phase, students complete structured reflection through online self-evaluation logs on *Xuexitong*, an institutionally adopted learning management platform familiar to participants and capable of storing learners' records longitudinally. Because receiving feedback alone is insufficient unless learners also process and use it to guide subsequent learning (Hattie & Timperley, 2007), this phase is designed to support post-task consolidation and

Online First View

self-regulation. Students are required to reflect on both language learning outcomes and interaction experiences. Reflection is structured as a feedforward loop, requiring students to summarize what they attempted, record useful vocabulary, expressions, and sentence structures, identify recurring problems, and commit to one concrete improvement target for the next cycle. Specific prompts invite students to review vocabulary and sentence structures learned from *DeepSeek*, the perceived usefulness of AI suggestions, whether and how they used these resources in speaking practice with *Doubao*, real-time error corrections received, support received during interaction, recurring problems, an action plan for the next learning cycle, and their emotional responses and confidence changes. The platform enables students to upload audio reflections and maintain a longitudinal record of learning experiences, thereby supporting review of progress and ongoing management of their learning process.

1.4 An integrated CVT–CoL framework

This study addresses the identified research gap by applying an integrated analytical framework that combines Control-Value Theory (CVT; Pekrun, 2006; Pekrun & Perry, 2014) and the Community of Inquiry (CoL) framework (Garrison et al., 1999). These two theoretical perspectives are particularly well-suited for this research because they offer complementary lenses for understanding the cognitive, affective, and social dimensions of the three-stage AI-assisted pedagogy.

CVT provides a parsimonious yet powerful account of how achievement emotions arise from two sets of appraisals: perceived control over achievement activities and outcomes, and subjective value attached to those activities and outcomes (Pekrun, 2006). Within the L2 speaking context, control appraisals encompass learners' perceptions of their ability to plan, produce, and monitor spoken output, while value appraisals concern the importance, interest, and utility learners assign to speaking tasks. CVT predicts that high perceived control combined with high perceived value generates positive activating emotions (e.g., enjoyment, hope), whereas low control and high value produce negative activating emotions (e.g., anxiety, shame). This framework is directly relevant to understanding how the three phases of the AI-assisted pedagogy—preparation (building control through linguistic readiness), practice (reducing social-evaluative threat), and reflection (enhancing perceived value through visible progress)—may differentially shape learners' emotional trajectories.

The CoL framework complements CVT by specifying the environmental conditions that support meaningful learning in technology-mediated settings. Originally developed to explain successful online learning communities, CoL posits that deep learning depends on the interaction of three core presences: teaching presence (the design, facilitation, and direction of cognitive and social processes), social presence (the ability of participants to project themselves socially and emotionally), and cognitive presence (the extent to which learners construct and confirm meaning through sustained reflection and discourse) (Garrison et al., 1999; Garrison, 2016). In the context of this study, teaching presence is instantiated through the structured three-phase design and the instructional guidance embedded in each tool; social presence encompasses both human–AI interaction during DeepSeek preparations and Doubao-based speaking practice; and cognitive presence is reflected in learners' critical analysis, evaluation, and synthesis during preparation and reflection.

The integration of CVT and CoL is theoretically justified on several grounds. First, CVT explicitly recognizes that achievement emotions are shaped by the social and instructional environment, including the quality of instruction, feedback, and social interaction (Pekrun & Perry, 2014). CoL provides the framework for operationalizing these environmental dimensions which is helpful to reveal the cognitive engagement patterns. Second, cognitive presence—a central construct in CoL—aligns with the critical thinking patterns that the three-stage pedagogy seeks to scaffold. Third, both frameworks share an emphasis on the dynamic, reciprocal relationship between environmental conditions and individual psychological processes, making them compatible for examining a multi-phase pedagogical intervention. By applying this integrated framework, the study moves beyond simple evaluation of AI

Online First View

tool effectiveness toward a theoretically informed investigation of how and why the three-stage pedagogy shapes learners' cognitive engagement patterns and emotional experiences.

Research Questions:

(1) How do learners perceive and attribute value to social presence when interacting with AI tools?
(2) How do language learners account for the relationship between cognitive presence and their value appraisals when interacting with AI tools?
(3) How do students perceive and describe the role of structured teaching presence in scaffolding their learning experience within the three-stage AI-assisted pedagogy?

2 Method

2.1 Participants and procedure

Semi-structured interviews were conducted with 13 volunteers (aged 18-19; 9 female, 4 male) drawn from a larger cohort of 91 associate-degree students across three intact classes instructed by the same teacher. The participants shared similar educational backgrounds, including approximately 10 years of prior English instruction. Based on university entrance examination metrics, their baseline proficiency aligned with CSE Level 3 (CEFR A2–B1 equivalent) for listening and reading, and CSE Level 2 (CEFR A2 equivalent) for speaking and writing (MoE & State Language Commission, 2018; NEEA & British Council, 2020).

2.2 Procedure

During Week 1, the instructor introduced the project and facilitated a hands-on workshop to familiarize students with the three-phase AI-assisted speaking framework. This included training on DeepSeek for task preparation, Doubao for contextualized speaking practice, and guided reflection techniques.

From Week 2 to Week 8, students completed seven weekly 45-minute speaking sessions structured around the three-phase cycle: teacher task introduction (≈5 min), AI-assisted preparation and revision via *DeepSeek* (≈15 min), interactive speaking practice with *Doubao* (≈15 min), and post-task reflection submitted via *Xuexitong* (≈10 min). Weekly reflections involved assessing previous goals, evaluating current performance, and establishing targets for subsequent sessions. Beginning in Week 3, each session commenced with teacher-synthesized feedforward based on prior reflections. Finally, students volunteered for semi-structured interviews and completed them in week 9 and 10.

2.3 Interview guiding questions

Semi-structured interviews examined participants' experiences across the three components of the instructional model: DeepSeek-assisted preparation, Doubao-powered speaking practice, and post-task reflection. The interview scheme, partially adapted from Zhang et al. (2025) and supplemented with custom items, elicited insights into perceived benefits, challenges, affective factors (e.g., confidence), workload manageability, interactional fluency, and perceived utility for language development. Core guiding questions focused on overall evaluations of the three-step approach, specific usage strategies for *DeepSeek*, the experience of interacting with *Doubao,* the utility of Xuexitong-based reflection, and perceived challenges along with coping strategies. Sample questions are: "Overall, what is your opinion about using this three-step approach to practice English speaking?" "Which part is the most helpful one?" "Did you encounter any difficulties?" "How did you solve them?". All interviews were conducted in Chinese, with relevant excerpts translated into English for reporting.

Online First View

2.4 Coding schemes and analysis

The qualitative analysis was guided by an integrated framework combining the Community of Inquiry (CoI) model (Garrison et al., 1999) and Control-Value Theory (CVT; Pekrun, 2006), wherein teaching, social, and cognitive presences were regarded as contextual antecedents triggering value appraisals and achievement emotions, see Figure 2. Coding schemes of relevant elements, including teaching presence, social presence, cognitive presence, control appraisals, value appraisals, and achievement emotions were developed based on relevant literature. Table 1 shows the definitions and examples of these elements. Data coding was performed in MAXQDA software. To ensure reliability, one rater coded the entire dataset while the second rater cross-checked all coding. In accordance with the Community of Inquiry (CoI) framework (Garrison et al., 2001), the unit of analysis was at the message level. Each conversational turn produced by a student during the interviews served as a unit of analysis. The coding scheme was deductive with emergent refinement, and all disagreements were resolved through consensus discussions. For instance, when analyzing comments on self-reflection forms, one rater initial coded them as Teaching Presence, whereas the second rater initially disagreed. After a discussion, both raters agreed that because the self-reflection form was a task assigned to students, it constituted part of the pedagogical design supporting independent learning. Consequently, such comments were categorized under Teaching Presence. Overall, the coding process was iterative, aligned with the nature of qualitative inquiry, and all final coded data were cross-checked by both raters. Code intersections were then extracted to identify co-occurrence frequencies and select representative examples.

Table 1

Coding Schemes

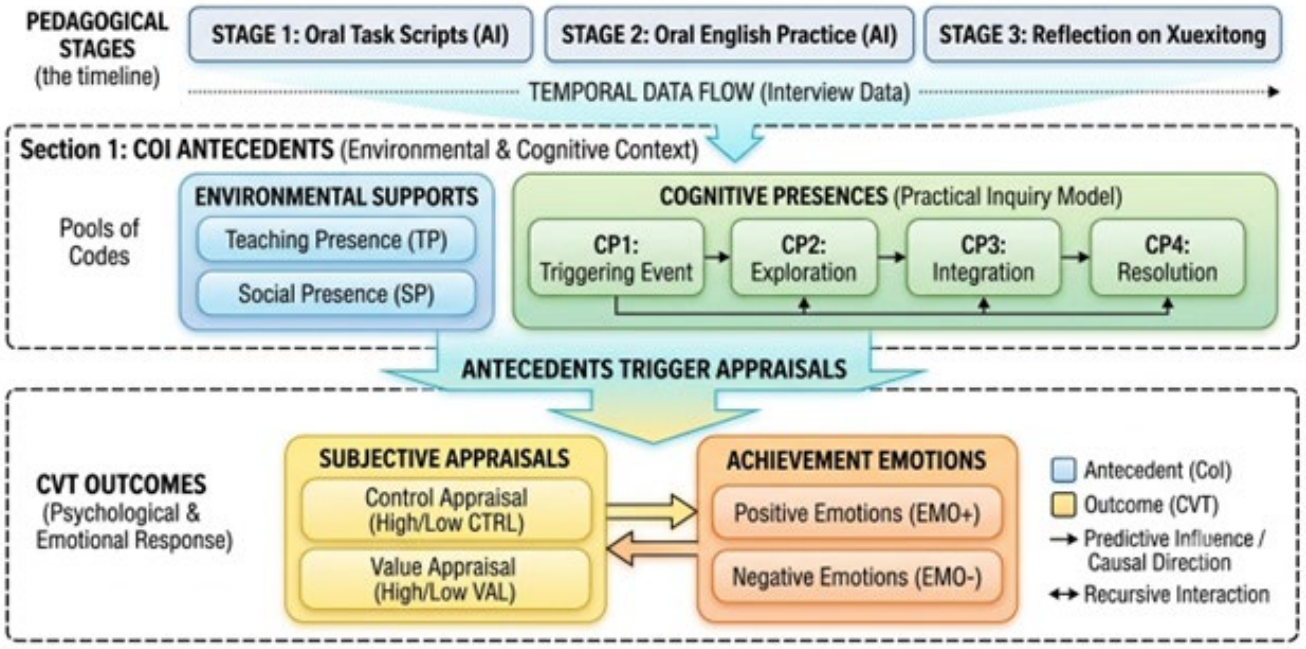
Aspects	Elements	Definitions	Examples
Environmental Supports	Teaching presence	Comments on how the prompt guidelines, scaffolding, or the AI's specific instructions helped or hindered them.	*I think this approach is really effective. Since I hadn't practiced it before, trying it out this semester made me realize just how impactful it actually is.*
	Social presence	How the student perceived their relationship with the AI (e.g., treating it as a peer, an authority, or a cold tool) OR how the AI-assisted tasks impacted their interaction with actual classmates/ teachers.	*You just don't really trust it. You feel like its work capability isn't great—it can only chat with you about simple everyday things. But when it's too simple, you feel it's useless, like 'I don't need to chat with you about this.'*
	Cognitive presence: Triggering Event, Exploration, Integration, Resolution (Garrison et al., 2001)	The process flows from identifying a problem (Triggering Event, TE), to gathering possibilities (Exploration, E), building a unified solution (Integration, I), and testing it out (Resolution, R).	*(Doubao) might misinterpret it as meaning something else entirely and then start talking to you based on that (TE).*

Online First View

Aspects	Elements	Definitions	Examples
The Outcomes	Control Appraisals: High Control (HC) Low Control (LC)	Statements about competence, self-efficacy, or frustration with AI predictability.	*I feel like whenever the topic is fairly easy, I can get on the same page with Doubao fast, and chatting with it is pretty seamless (HC)* *It (Doubao) keeps interrupting. I'm just a bit annoyed (LC + NE).*
	Value Appraisals: High Value (HV) Low Value (LV)	Statements about the perceived utility or relevance of the AI tool.	*I feel like it's really good since it has both words and sentence patterns (HV).*
	Achievement Emotions: Positive Emotions (PE: Enjoyment, Hope, Pride etc.) Negative Emotions (NE: Anxiety, Boredom, Frustration etc.)	Explicit or implicit emotional states linked to the tasks.	*Either way, it probably won't help me improve much (LV).* *I felt pretty happy with it (PE).*

Figure 2

An Overview of the Analytical Framework



Online First View

3 Findings

3.1 Social presence: The engine of student appraisals

The most prominent clusters in the co-occurrence analysis centered on social presence (SP), revealing its centrality in AI-mediated speaking pedagogy. The association between social presence (SP) and high value produced showed the most intersection in the dataset. This finding could be particularly striking for an AI-mediated context, where social presence is mediated not by human peers alone but by human–AI interaction.

Analyzing student reflections within the context illuminated the nature of this social presence. Students repeatedly described the AI tools using relational language: "*it corrects my mistake and it is so smart*" "*I feel there is no pressure at all, and it is completely okay even if I make a mistake*". This relational framing is consistent with Nass and Reeves' (1996) Computers Are Social Actors (CASA) paradigm, which posits that users unconsciously apply human social rules to technological interaction. In CVT terms, this perceived social safety appears to reduce the perceived social-evaluative stakes of speaking, a key antecedent of foreign language anxiety (Horwitz et al., 1986), while preserving or enhancing the subjective value of the speaking task.

The data also revealed a substantial co-occurrence between social presence and low control appraisals. This initially counterintuitive finding carries important theoretical significance. When students reported struggling with the AI tool by experiencing technical friction, struggling to formulate appropriate prompts, or feeling linguistically outmatched, they did not disengage or devalue the activity. Instead, they appeared to lean more heavily into the dialogic, relational dimension of the interaction as a coping mechanism. One participant commented "*Part of the difficulty might be that my vocabulary is not large enough. Once it gets hard, everything stalls. And once I start stumbling, it feels like I am making all kinds of errors. Overall, I still think it is a pretty good experience… Maybe it's because I'm not afraid to speak English with an AI like this. But if you asked me to talk to a real person, first of all, that's just out of the question—I'd be too scared*". This pattern suggests that AI-mediated social presence could serve an "emotional buffering" function. It seems prevents low control appraisals from cascading into the negative deactivating emotions (e.g., hopelessness, boredom) that CVT would predict under conditions of low control combined with low support (Pekrun & Perry, 2014).

A third significant intersection was social presence with resolution and application which indicated that students' ability to apply their learning to a coherent, real-world English language product was deeply tied to their social experience of the AI interaction. This aligns with socio-constructivist accounts of learning, which emphasize that understanding solidifies through dialogic exchange and collaborative meaning-making (Vygotsky, 1978; Lantolf & Thorne, 2006). The integration of *DeepSeek*'s preparatory resources and *Doubao*'s conversational practice facilitated comprehensive language development across multiple modalities. Participants reported that the synergy between the two tools created an effective "feedback loop," which improved their performance in formal writing tests from their perspectives. Beyond writing, participants also commented that this AI-mediated environment seems significantly reduce their foreign language speaking anxiety. One participant noted "*It prepares you to hear people speaking English to you in real life. There has been a qualitative improvement across these core areas: listening, speaking, reading, and writing*".

3.2 Cognitive presence: The indicator of perceived value

The association between cognitive presence, specifically integration (convergence and synthesis in the CoI framework), and high value appraisals yielded fewer occurrences, compared with the social presence clusters. These intersections may suggest that higher-order critical thinking patterns, rather than surface-

Online First View

level engagement with AI tools, could be what makes students perceive the learning experience as genuinely valuable.

The contextual analysis of students' comments indicated the nature of this cognitive integration. Students described active convergence, connecting ideas and synthesis when evaluating interactions with AI tools. For instance, one participant explicitly linked *DeepSeek*'s feedback with his high school knowledge, synthesizing how the tool benefited his language development. Another compared conversations with *Doubao* to speaking practice with classmates, highlighting the linguistic advantages of *Doubao*'s responses. This feedback aligns with what Garrison et al. (2001) characterize as the «triggering event → exploration → integration → resolution" trajectory of critical inquiry. Students who reached the integration phase, i.e. actively connecting ideas, resolving inconsistencies between their own linguistic knowledge and AI suggestions, and synthesizing resources into coherent spoken output, consistently reported the highest levels of perceived learning value.

The relationship between integration and social presence is equally instructive. Critical cognitive synthesis did not occur in isolation. It was deeply embedded in and supported by the dialogic nature of AI interaction. This finding echoes Akyol and Garrison's (2011) argument that cognitive and social presence are interdependent dimensions of deep learning, and extends this principle to human–AI educational interaction. The AI served not merely as an information source but as an intellectual interlocutor that students could question, negotiate with, and ultimately learn from.

Notably, integration rarely co-occurred with low value or low control. These negative associations reinforce the argument that critical thinking tendency could empower and satisfy learners. When students appear to exercise cognitive agency by actively evaluating AI output, exercising selectivity in tool use, and synthesizing disparate resources into coherent spoken production, they seem experience a sense of authentic accomplishment that drives both intrinsic and instrumental value appraisals. This finding resonates with self-determination theory (Ryan & Deci, 2024), which identifies competence and autonomy as fundamental psychological needs that, when satisfied, generate intrinsic motivation. In the context of AI-assisted speaking, the critical thinking tendency may provide the vehicle for satisfying these needs. Students could feel competent when they navigate complex linguistic-cognitive challenges, and autonomous when they exercise judgment over AI-generated resources.

3.3 Teaching presence: The foundation for high value

Teaching presence, i.e. the design, facilitation, and direction of cognitive and social processes in the learning environment (Garrison et al., 1999), emerged as a foundational construct in shaping students' value appraisals. The co-occurrence analysis revealed a trend that teaching presence (TP) seems to be intersected with high value appraisals, making this one of the most salient associations in the dataset. Students consistently attributed their positive engagement to the structural clarity provided by the three-phase design. One participant observed that the preparation phase "*correct my language errors, refine my sentence structure, and teach me natural expressions I can use before speaking*", echoing the CVT prediction that perceived environmental structure enhances control and, in turn, value (Pekrun, 2006).

The specific sequencing of the three phases, i.e. preparation, practice, reflection, appeared to function as a psychological scaffold that segmented what might otherwise be experienced as an overwhelming cognitive demand. Students described the preparation phase (*DeepSeek*) as "*always giving [them] an 'aha!' moment*", the practice phase (*Doubao*) as allowing them to "*talk freely about whatever [they] want*", and the reflection phase (*Xuexitong*) as helping them "*know where [they] went wrong*". This layered support mechanism aligns with what Hattie and Donoghue (2016) describe as "skill, will, and thrill" dimensions of learning, with skill-building through linguistic preparation, will through low-stakes practice, and thrill through visible progress in reflection.

Online First View

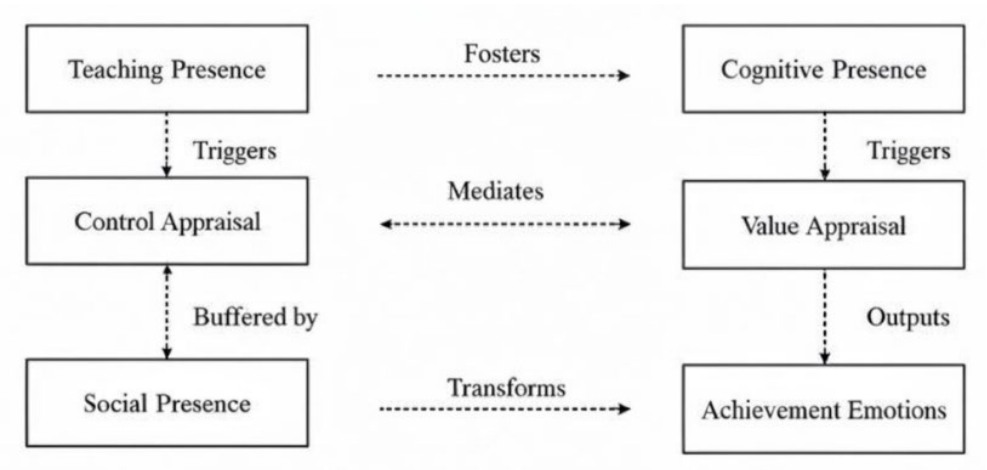
Critically, the dataset revealed minimal co-occurrence between teaching presence and low value appraisals, and similarly few overlaps between TP and low control. This asymmetry suggests theoretical consideration. It appears that when teaching presence is robust, feelings of disengagement and perceived incompetence remain rare. The structured pedagogy appears to function as a protective factor, mitigating the cognitive and affective drift characteristic of unregulated technology use among students (Kirschner, Sweller, & Clark, 2006). As one participant succinctly noted: "*I feel that every stage is closely interconnected*." This finding corroborates Li et al.'s (2025) contention that technology-enhanced practice is most effective when deliberately paired with metacognitive scaffolding.

4 Discussion

This study set out to examine how a three-stage AI-assisted speaking pedagogy shapes L2 learners' critical thinking tendency and achievement emotions through the lens of an integrated CVT–CoL framework. The findings suggest that the three presences including teaching, social, and cognitive presences, may not operate in parallel but tend to form a dynamic, recursive system that alters how students appraise control, value, and ultimately emotional experience. Building on these findings, we propose a tentative conceptual model in which teaching presence could function as the structural foundation, social presence as the emotional engine, and cognitive presence (specifically integration) as the value catalyst, with each reinforcing the others in a cyclical process. Figure 3 shows this tentative conceptual model.

Figure 3

A Tentative Conceptual Model



4.1 Social presence as an emotional buffer (RQ1)

The finding that high social presence co-occurred intensively with both high value and low control suggests that AI-mediated social presence fundamentally alters the traditional CVT appraisal pipeline. In classic CVT formulations, a decline in perceived control typically triggers negative activating emotions such as anxiety and shame (Pekrun, 2006; Pekrun et al., 2011). However, while low control appraisals produced the emotional debilitation predicted by Control-Value Theory (CVT), participants' value appraisals remained surprisingly high.

The mechanism of this buffering effect can be understood through two complementary theoretical lenses. First, from a Computers are Social Actors (CASA) perspective (Nass & Reeves, 1996; Gambino,

Online First View

Fox, & Ratan, 2020), students' unconscious application of human social scripts to AI interaction transforms the meaning of performance struggle. When a human interlocutor would trigger evaluative concern, the AI—perceived as "non-judgmental" and "patient"—converts potential failure into low-stakes experimentation. This aligns with prior findings that AI chatbots reduce L2 speaking anxiety by removing the social-evaluative threat inherent in human–human interaction (Ebadi & Ebadijalal, 2022; Tai & Chen, 2023).

Second, from a sociocultural perspective, the AI functions as what Vygotsky (1978) termed a "more capable other" within the learner's zone of proximal development. When students experience control deficits, either linguistic or technical, the AI's conversational scaffolding maintains interactional momentum, preventing the cognitive shutdown that would otherwise occur. Unlike human peers, the AI provides infinite patience and unlimited retry opportunities, transforming potential failure into a learning loop. This "unconditional interactional support" is a distinctive affordance of AI-mediated speaking practice not available in traditional classroom settings (Jeon, 2022; Kohnke et al., 2023).

4.2 Cognitive integration as a mechanism for value creation (RQ2)

The robust association between cognitive integration and high value appraisals addresses a central question in AI-assisted language learning research: what makes students value AI-mediated practice? The data suggest a counterintuitive answer grounded in cognitive challenge rather than ease. Students did not value the AI because it simplified their work; they valued it because the intellectual friction of critical negotiation with AI, including evaluating suggestions, reconciling contradictions, selecting and adapting resources, made learning outcomes feel authentically earned.

This finding extends self-determination theory (Ryan & Deci, 2024) into the domain of AI-assisted language education. The need for competence is satisfied not by effortless success but by optimal challenge, i.e. tasks that stretch learners without overwhelming them. Integration, the third phase of cognitive presence, by its very nature, represents such optimal challenge in the cognitive domain. It requires learners to move beyond information consumption into active knowledge construction. Students who appeared to engage in integration tended to exercise what Facione (1990) terms the core critical thinking skills of analysis, evaluation, and inference, and this might drive drove their value appraisals.

Furthermore, the co-occurrence between integration and social presence suggests a reciprocal relationship. From students' perceptions, interactions with AI could provide the material for critical synthesis, and critical synthesis could deepen the quality of subsequent AI interaction. This finding extends the CoI framework's original emphasis on the interdependence of cognitive and social presence (Garrison, 2016; Akyol & Garrison, 2011) into the unique context of human–AI learning partnerships. The AI serves simultaneously as an information source, a practice partner, and an intellectual foil which are three roles traditionally distributed across teachers, peers, and learning materials. This multi-role functionality may explain why well-designed AI pedagogy can simultaneously support cognitive engagement patterns and emotional well-being.

3.3 Teaching presence as a scaffold for competence (RQ3)

The strong link between teaching presence and high value, alongside the minimal co-occurrence of teaching presence with negative appraisals, underscores a critical point. Pedagogical design is not merely an environmental variable but the structural anchor that determines whether AI tools empower or overwhelm learners. AI tools, particularly LLMs with their open-ended conversational capacity, can easily induce choice paralysis and cognitive overload (Sweller, 1988; Kirschner et al., 2006). Without deliberate instructional structure, the sheer range of possible interactions with AI may overwhelm learners' limited cognitive resources for managing both linguistic production and tool navigation.

Online First View

The three-stage design—preparation, practice, reflection—represents a form of macro-scaffolding (Hammond & Gibbons, 2005) that could reduce cognitive load at each stage. Phase 1 (*DeepSeek*) offloads linguistic retrieval demands, allowing learners to focus on conceptual organization. Phase 2 (*Doubao*) provides a bounded, low-stakes interaction space that channels productive output. Phase 3 (*Xuexitong*) institutionalizes metacognitive consolidation, transforming ephemeral interaction into durable learning. This structural segmentation transforms what could be an unstructured, anxiety-provoking AI experience into a coherent learning arc in which each phase builds on and justifies the previous one.

We argue that, without deliberate teaching presence of this kind, cognitive presence in AI-mediated learning would likely stagnate at the exploration phase (CP2)—characterized by surface-level interaction, random querying, and passive consumption—without progressing to integration (CP3) and resolution (CP4). Teaching presence, in other words, could convert AI from a digital distraction (Flanigan & Babchuk, 2022) into a cognitive tool. This interpretation aligns with Kohnke et al.'s (2023) call for purposeful pedagogical integration of AI tools rather than ad hoc experimentation, and with Yan's (2023) finding that structured prompting and task framing significantly shape the quality of LLM-mediated learning outcomes.

5 Conclusion

Aiming at addressing important gaps regarding higher-order cognitive processes, isolated tool usage, and broader achievement emotions in AI-assisted language learning, this study evaluated experience of an integrated, three-stage AI speaking pedagogy (preparation, practice, and reflection) among 13 EFL learners. Guided by a combined Community of Inquiry (CoI) and Control-Value Theory (CVT) framework, the qualitative findings of students' accounts suggest that teaching, social, and cognitive presences may not operate in isolation. Analyzing their dynamic intersections helps reveal how critical thinking patterns shape students' control, value and emotional experiences.

Specifically, it appears that teaching presence provides the structural foundation that stabilizes task design, while human-AI social presence serves as the primary emotional engine, creating a judgment-free environment that drives task value. Higher-order cognitive presence, particularly integration, functions as the value catalyst, demonstrating that engaging L2 learners in critical evaluation and synthesis elevates their perceived purpose and agency. Ultimately, this study proposes a tentative conceptual model illustrating how structured pedagogical scaffolding and dialogic AI interaction could collaboratively transform linguistic friction into empowering, value-rich cognitive growth.

To maximize L2 speaking outcomes, AI-assisted pedagogy could be designed as a scaffolded, metacognitive journey that leverages the technology's psychological and analytical strengths. Instructors should design the use of conversational AI tools in their speaking activities, as the low-stakes, judgment-free dialogic AI partners allow L2 learners to make mistakes, experience temporary linguistic friction, and experiment with complex structures without fear of negative social evaluation. This psychological buffer maintains task value and helps transition students from anxiety to agency.

Pedagogy could explicitly shift from teaching students how to generate text with AI to how to critically evaluate and modify AI output. For instance, when preparing oral task scripts, teachers could guide students on "prompt iteration" and "fact/grammar verification" by asking students to identify hallucinated content, critique AI style, or adapt output to their authentic voice. Explicitly scaffolding this cognitive friction helps students move past surface-level exploration into higher-order cognitive presence.

Integrating generative AI tools into language courses might be insufficient on its own. It may require a structured metacognitive anchor (such as *Xuexitong*, Canvas, or Moodle discussion boards) to consolidate learning. Instructors could design reflective tasks that require students to compare their initial

Online First View

AI-assisted scripts, document their speaking practice progress, and reflect on their prompt strategies. This structured reflection could help make their implicit cognitive growth explicit, assisting learners internalize their gains in control and perceive the long-term utility of AI-assisted learning.

This study has two main limitations. First, the small sample of 13 associate-degree students at a single institution and specific proficiency level limits the generalizability of the findings, as learners with higher language proficiency or in different educational contexts may experience different emotional and cognitive dynamics. Second, the reliance on post-hoc, self-reported interview data introduces potential recall or social desirability bias, although it could be useful for capturing retrospective reflections. Future research would benefit from incorporating real-time data streams, such as screen recordings, chat logs, or think-aloud protocols, to trace critical thinking patterns and emotional experience as they dynamically unfold during AI-assisted tasks.

Acknowledgements

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU/RGC Ref No.: 13608125) and a grant from the CPCE Centre for Pedagogic Research (CPCE-CPR), PolyU (Project No.: LC-TF-2025-040(E)).

References

Akyol, Z., & Garrison, D. R. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, *42*(2), 233-250. https://doi.org/10.1111/j.1467-8535.2009.01029.x

Bygate, M. (1987). *Speaking*. Oxford University Press.

DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology, 24*(1), 16–26.

Ebadi, S., & Ebadijalal, M. (2022). The effect of Google Expeditions virtual reality on EFL learners' willingness to communicate and oral proficiency. *Computer Assisted Language Learning*, *35*(8), 1975-2000. https://doi.org/10.1080/09588221.2020.1854311

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, *38*(3), e12667. https://doi.org/10.1111/exsy.12667

Facione, P. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). The California Academic Press.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 71-85. https://doi.org/10.30658/hmc.1.5

Gao H., Hashim H., Md Yunus M. (2025). Assessing the reliability and relevance of DeepSeek in EFL writing evaluation: a generalizability theory approach. *Lang Test Asia*, 15, 33. https://doi.org/10.1186/s40468-025-00369-6

Garrison, D. R. (2016). *E-learning in the 21st century: A community of inquiry framework for research and practice (3rd ed.)*. Routledge.

Garrison, D. R. (2007). Online community of inquiry review: Social, cognitive, and teaching presence issues. *Journal of Asynchronous Learning Networks*, *11*(1), 61-72.

Online First View

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2-3), 87-105. https://doi.org/10.1016/S1096-7516(00)00016-6

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7-23. https://doi.org/10.1080/08923640109527071

Ghanizadeh, A. (2017). The interplay between reflective thinking, critical thinking, self-monitoring, and academic achievement in higher education. *Higher Education*, *74*(1), 101-114.

Goh, C. C., & Burns, A. (2012). *Teaching speaking: A holistic approach*. Cambridge University Press.

Hammond, J., & Gibbons, P. (2005). Putting scaffolding to work: The contribution of scaffolding in articulating ESL education. *Prospect, 20*(1), 6–30.

Hattie, J. A., & Donoghue, G. M. (2016). Learning strategies: A synthesis and conceptual model. *npj Science of Learning, 1*(1), 16013.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern language journal*, *70*(2), 125-132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, *37*(1-2), 1-26. https://doi.org/10.1080/09588221.2021.2021241

Kirschner, P., Sweller, J., & Clark, R. E. (2006). Why unguided learning does not work: An analysis of the failure of discovery learning, problem-based learning, experiential learning and inquiry-based learning. *Educational Psychologist*, *41*(2), 75-86. https://doi.org/10.1207/s15326985ep4102_1

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *Relc Journal*, *54*(2), 537-550. https://doi.org/10.1177/00336882231162868

Lantolf, J. P. (2006). *Sociocultural theory and the genesis of second language development.* Oxford University Press.

Levelt, W. J. (1993). *Speaking: From intention to articulation*. MIT press.

Li, L. (2016). Integrating thinking skills in foreign language learning: What can we learn from teachers' perspectives? *Thinking Skills and Creativity*, *22*, 273-288. https://doi.org/10.1016/j.tsc.2016.09.008

Li, X., Zhang, Y., Zhang, X., Zhou, H., & Yang, Y. (2025). A systematic review on the integration of gamified pedagogy in computer-supported collaborative learning. *Behaviour & Information Technology*, 1-16.

Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal*, *92*(1), 71-86. https://doi.org/10.1111/j.1540-4781.2008.00687.x

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*(2), 283-305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

Ministry of Education of the People's Republic of China, & State Language Commission. (2018). *China's Standards of English Language Ability (CSE)*. Higher Education Press. https://www.hep.com.cn/

Online First View

Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners?. *Computer Assisted Language Learning*, *33*(8), 865-890. https://doi.org/10.1080/09588221.2019.1595664

Namaziandost, E., Rezai, A., Kargar Behbahani, H., & Heydarnejad, T. (2025). Role of academic emotion regulation and self-compassion in EFL learners' academic motivation, growth mindset, and psychological wellness: insights from self-determination theory. *Psychological Studies*, *70*(1), 84-101.

Nass, C., & Reeves, B. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. CSLI Publications.

National Education Examinations Authority, & British Council. (2020). *Alignment of the Chinese Standards of English Language Ability (CSE) with the Common European Framework of Reference for Languages (CEFR)*. Higher Education Press. https://www.britishcouncil.cn/en/exam/ielts/cse-cefr-alignment

Elder, L., & Paul, R. (2020). *Critical thinking: Tools for taking charge of your learning and your life*. Bloomsbury Publishing USA.

Flanigan, A. E., & Babchuk, W. A. (2022). Digital distraction in the classroom: Exploring instructor perceptions and reactions. *Teaching in Higher Education, 27*(3), 352–370. https://doi.org/10.1080/13562517.2020.1724937

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315-341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36-48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Routledge.

Ryan, R. M., & Deci, E. L. (2024). Self-determination theory. In *Encyclopedia of quality of life and well-being research* (pp. 6229-6235). Springer International Publishing.

Shao, K., Pekrun, R., & Nicholson, L. J. (2019). Emotions in classroom language learning: What can we learn from achievement emotion research? *System*, *86*, 102121. https://doi.org/10.1016/j.system.2019.102121

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Routledge.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tai, T. Y., & Chen, H. H. J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, *31*(3), 1485-1502. https://doi.org/10.1080/10494820.2020.1841801

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in second language acquisition*, *41*(2), 363-387. https://doi.org/10.1017/S0272263118000311

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Online First View

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, *28*(11), 13943-13967. https://doi.org/10.1007/s10639-023-11742-4

Yuan, R., & Stapleton, P. (2020). Student teachers' perceptions of critical thinking and its teaching. *ELT Journal*, *74*(1), 40-48. https://doi.org/10.1093/elt/ccz044

Zhang, R., & Zou, D. (2022). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, *35*(4), 696-742. https://doi.org/10.1080/09588221.2020.1744666

Zou, D., Huang, Y., & Xie, H. (2021). Digital game-based vocabulary learning: where are we and where are we going?. *Computer Assisted Language Learning*, *34*(5-6), 751-777. https://doi.org/10.1080/09588221.2019.1640745

Molly Xie Pan is a lecturer at the Division of Languages and Communication, College of Professional and Continuing Education, Hong Kong Polytechnic University. She obtains a PhD in cognitive linguistics. Her research interest includes second language learning and metaphor studies. She serves as the associate editor for *Metaphor and the Social World*. ORCID: https://orcid.org/0000-0003-3053-329X

Di Zou is an Associate Professor in the Department of English and Communication at The Hong Kong Polytechnic University. Her research interests include AI in language education, technology-enhanced language learning, and game-based language learning. She has published more than 200 papers in international journals. Stanford University listed her as one of the World's Top 2% Most-cited Scientists since 2021 for five years. ORCID: https://orcid.org/0000-0001-8435-9739

Dandan Xie is a lecturer in the School of Business Foreign Languages at Shenzhen Polytechnic University. Her research interests primarily revolve around AI in language education and the application of metaphor theories. ORCID: https://orcid.org/0009-0003-2834-979X

Online First View