

Article

From Novelty to Strategy: The Trajectory of Learner Perceptions in Human and AI-Assisted L2 Speaking Tasks

Gavin Bui

The Hang Seng University of Hong Kong, Shatin, Hong Kong SAR, China

Received: 26 November 2025 / Received in revised form: 10 February 2026 / Accepted: 1 March 2026 / Available online: 5 March 2026

Abstract

This research explores the dynamics in perceptions of L2 English learners in task-based speaking interactions with human interlocutors in comparison to AI (chatbot “Doubao”) communication partners over seven weeks. Findings from three rounds of interviews indicated enthusiastic initial adoption of the AI talk pal, driven by its ability to stimulate output, its ease of access, and its lower affective barriers to communication. However, with continued use, their fervent zeal had cooled to being “lukewarm” and finally “cool” due to AI’s lack of social presence, speech prosody, responsive interaction, and appropriate feedback, whereas interaction with real communication partners was consistently preferred for their authentic nature. In light of AI’s limitation, learners adapted to develop prompt literacy which enabled them to deliberately employ AI for particularised tasks, such as increasing fluency and acquiring vocabulary. Based on the findings, a dynamic model for AI acceptance in L2 task interaction is proposed to advance our understanding of the role of AI in task-based learning and teaching. It suggests that a hybrid pedagogy, which sequences AI and human interaction to leverage their complementary strengths, is essential for effective task-based language learning.

Keywords

Task-based language teaching and learning, L2 interaction, AI chatbot, AI-assisted interaction, Learner perception of AI, AI literacy

1 Introduction

Task-based language teaching (TBLT) has established itself as a prominent pedagogical approach that foregrounds meaning-centred communication and real-world language use as central to language learning processes (Bui, 2025; Ellis et al., 2019). The contemporary language learning landscape, however, is increasingly characterised by digital transformation and the integration of artificial intelligence (AI) technologies that offer new possibilities for interactive practice (Moorhouse & Wong, 2025). The rapid diffusion of conversational agents and AI tutoring in language education has

created unprecedented access to on-demand L2 speaking practice, yet their pedagogical value depends on how interaction unfolds during tasks (Chapelle & Sauro, 2017; Okonkwo & Ade-Ibijola, 2021). Within task-based language teaching, L2 acquisition is promoted when meaning-oriented activities elicit negotiation of meaning, “pushed” output, and timely focus on form (Bui, 2025; Ellis, 2024; Long, 2015), which are mechanisms established primarily in human-to-human contexts. Research on online human interaction shows that technology-mediated exchanges can reproduce many of these benefits, though outcomes vary with modality and design (Satar & Özdener, 2008). Whether AI-assisted interaction can reliably instantiate these processes is subject to empirical inquiries. Intelligent computer-assisted language learning (ICALL) and automatic speech recognition (ASR) offer both the potential of adaptive, immediate feedback and persistent concerns about diagnostic accuracy, dialog management, and sociopragmatic appropriateness (McCrocklin, 2019; Chapelle & Sauro, 2017). Given institutional pressures to scale speaking opportunities and learners’ interest in low-stakes practice, comparative evidence is urgently needed to determine when AI can serve as an effective interlocutor, where limitations constrain learning, and how tasks and feedback should be engineered to align with TBLT principles (Bui, 2025; Chapelle & Sauro, 2017). This study attempts to shed light on these issues by comparing human and AI-assisted interaction in task-based L2 English speaking over seven weeks, with self-reports of experiences both in the school and outside of classroom.

2 Literature Review

2.1. Task-based language teaching and the role of interaction

Task-based language teaching (TBLT) situates language learning in meaning-focused goal-oriented activity, with task performance conceived as the locus of second language (L2) development when learners are engaged in interaction that prompts attention to form and opportunities for output (Ellis, 2003; Long, 2015). Long’s (1996) Interaction Hypothesis argues that modified input and feedback arising from breakdowns in communication promote negotiation of meaning and comprehensible input. They also provide negative evidence and create conditions for noticing and intake. All this is conducive to L2 acquisition. In parallel, Swain’s (1995) Output Hypothesis posits that “pushed” output during communicative activity contributes to development by prompting syntactic processing, hypothesis testing, and metalinguistic reflection. These theories have significantly influenced task design. Tasks should promote genuine information exchange, create communicative gaps to elicit negotiation of meaning, and be structured to afford negotiation and focus on form without compromising primary meaning-focused engagement (Bui, 2024; Long, 2015).

Early research has explored how human-to-human interaction in tasks supports L2 speaking performance and development. Negotiation of meaning through clarification requests, confirmation checks, and comprehension checks are associated with greater comprehensibility and opportunities for learning (Long, 2015). Furthermore, different types of feedback, such as recasts and prompts, are shown to be associated with learner uptake and morphosyntactic development under certain conditions (Kong & Bui, 2019). Meta-analyses of corrective feedback have synthesised robust effects for oral feedback on L2 development, moderated by factors such as feedback type, target feature, and measurement type (Lyster & Saito, 2010). Within a TBLT framework, this evidence underscores the value of tasks that create frequent opportunities for interaction and negotiation of meaning, thereby generating focus-on-form episodes that drives L2, especially syntactic, development. At the performance level, the presence of responsive human interlocutors can scaffold fluency (e.g., through planning between turns and construction in interaction), complexity (e.g., through deeper idea development and pushed language), and accuracy (e.g., through feedback and reformulation) (Bui & Skehan, 2018; Michel, 2017).

2.2. Technology- and AI-enhanced interaction in L2 learning

Over the past two decades, technology-mediated interaction has expanded the modalities through which negotiation and feedback can occur. Research on computer-mediated communication (CMC), both synchronous text/voice and video, has shown that learners engage in negotiation in online environments and that such interaction can promote L2 development similarly to face-to-face exchanges, with modality-specific affordances (Qiu et al, 2025a, 2025b; Smith, 2003). Prior literature has compared text and voice chat and reported differences in interactional density, planning time, and anxiety, with text chat often affording more reflection and noticing (e.g., due to persistence of the written record) and voice chat aligning more closely with oral performance demands (Jepson, 2005; Satar & Özdener, 2008). Telecollaboration studies demonstrate gains in intercultural communicative competence and pragmatics through sustained interaction with L2 or advanced L2 interlocutors in authentic tasks (Akiyama & Saito, 2016). Taken together, previous research suggests that technology-assisted (online) interaction generally benefits L2 learning even though such outcomes vary with task design, partner pairing, and the nature of feedback (Chapelle & Sauro, 2017).

More recently, developments in ICALL and conversational agents have introduced possibilities for AI-assisted interaction in task-based learning. ICALL systems offer adaptive, systematic analysis of learner input and tailored feedback, especially in writing and grammar (Heift & Schulze, 2007). In speaking, ASR has been used to create interactive tasks for pronunciation and fluency practice. McCrocklin (2019) documented improvements in segmental/suprasegmental features and learner awareness when ASR feedback is integrated into communicative practice. Belda-Medina and Calvo-Ferrer (2022) conducted a four-week intervention study with 176 undergraduates from Spain and Poland, they found that the integration of conversational AI agents in language learning was well received by the students, especially in terms of the perceived ease of use and attitudes. However, the short duration and the ad hoc nature of the AI module in the learning process may not fully capture the dynamic changes of learner perception over time. Generally speaking, the emergence of conversational AI-chatbots for education allows easier access to simulated speaking partners that can provide immediate responses, high availability, and reduced social anxiety (Xing & Saeed, 2026). Specifically in SLA, early work on chatbots and agent-mediated tasks point to potential for increased practice opportunities and reduced affective barriers; however, the use of AI-agents also raises concerns as to how to ensure feedback accuracy, comprehensibility, and opportunities for noticing of forms (Chapelle & Sauro, 2017; Heift & Schulze, 2007).

2.3 Issues in AI-assisted interaction and coping strategies

With the advent of more capable large language models, AI-assisted interaction has gradually improved in approximating human interlocutors. AI chatbots have learned (or trained to) sustain topic continuity, offer contextualised recasts or prompts, and adapt task difficulty according to learner needs. That being said, the issue of interlocutor authenticity (van Lier, 1996) cannot be ignored. Early social presence theory (Short et al., 1976) argues that the extent to which one perceives the presence of a real human is crucial for communication. More recently, this theory has been applied to computer-mediated communication or virtual environments where the degree of social presence can impact the quality of virtual interactions and outcomes (Bickle et. al., 2019). Current AI agents, however, appear to be unable to demonstrate a sufficient level of sociopragmatic appropriateness (Bui & Wong, 2021), turn-taking nuances, and the co-construction of meaning (Long, 2015). These shortcomings may result in diminished or absent social presence, potentially undermining L2 learners' motivation to sustain conversational practice with AI.

Furthermore, AI systems, if they were to outperform human interlocutors in L2 learning tasks, must be engineered to include capacities of negotiation of meaning, pushed output, and focus on form,

depending on specific interactional contingencies (e.g., communication breakdowns, selective feedback). Emerging findings on ASR-supported oral tasks and AI-chatbot-mediated practice point to design principles consistent with TBLT: tasks should be goal-oriented; agent feedback should be salient, timely, and specific; and interaction should require learners to plan, repair, and reformulate, but not merely produce rote responses (McCrocklin, 2019; Chapelle & Sauro, 2017). From a philosophical perspective, Jin and Wang (2025) argues that, while AI use in education is helpful at the utilitarian level, it may not have full potential to replace human teachers / partners because it lacks subject-ness, and therefore could not be treated with the same level of responsibility and sentiment as humans.

One of the suggestions for coping with the limitations of AI-assisted L2 task interaction is prompt literacy (Tour & Zadorozhnyy, 2025). Prompt literacy refers to a holistic set of capabilities required for dynamic and iterative interaction with generative AI, which include the ability to formulate, critically evaluate, and iteratively refine prompts (Gattupalli et al., 2023). Prompt literacy is important because the quality of the prompt directly determines the relevance, accuracy, and creativity of the AI's output. For language learners, effective prompts can generate tailored linguistic input, more authentic conversational practice, and assistance with grammatical accuracy and lexical complexity (Cope & Kalantzis, 2024; Godwin-Jones, 2023). Otherwise, learners may receive generic or overly complex responses that hinder rather than help their language development. Tour and Zadorozhnyy (2025) suggest four practices training L2 learners in prompt literacy: code-breaking (using clear, specific language); participating in the text (adding context and meaning); using the text (tailoring prompts for a specific audience or purpose); and analysing the text (critically examining prompts and outputs for bias or inaccuracy). This structured approach helps learners refine their prompts in stages to achieve more effective and meaningful AI interactions.

2.4 Research gaps and research questions

In summary, decades of human interaction research have much emphasised the importance of negotiation of meaning, pushed output, and feedback in task-based speaking performance and development (Ellis, 2003; Long, 1996; Qiu et al, 2025a). Technology-mediated human-to-human interaction can reproduce many of these benefits, with modality-specific affordances and constraints (Blake, 2000; Qiu et al, 2025b). In contrast, AI-powered interaction offers additional scalability and adaptivity, but its effectiveness depends on how well tasks and agent behaviours instantiate engaging input processes and output functions, and on safeguards against inaccurate or inauthentic input and feedback. Taking the above review together, three research gaps emerge. Firstly, most existing studies emphasised L2 task performance rather than learners' lived experiences and perceptions of human versus AI interlocutors within task-based teaching and learning context. Secondly, prior research is predominantly cross-sectional, with little research on how experiences and perceptions on AI-assisted interaction evolve with extended exposure as novelty effects fade and agent behaviours become familiar. Last but not least, there is little fine-grained evidence on the task, agent, and learner factors that shape these perceptions and their changing trajectories. These gaps motivate the following research questions (RQs):

RQ1: What are the participants' experiences and perceptions of interacting with human and AI-chatbot interlocutors, and what factors underlie them?

RQ2: How, if at all, do these experiences and perceptions evolve over time?

In this article, "experiences" are operationalised as participants' self-reported and researcher-observed behaviours (i.e., what they did) during the two types of task-based interaction. "Perceptions" refer to their affective responses to, and evaluative comments on, these interactive learning processes (i.e., what they thought).

3 Research Methods

3.1 Participants

This research adopted a purposive sampling strategy to recruit participants. Four learners of L2 English were invited by the researcher, who was their teacher at the time of research, to participate in this longitudinal study. They were selected from a cohort of 43 new Mainland Chinese students (36 MA and 7 BA) admitted that academic year to the Department of English at a private university in Hong Kong. They were invited because they had previously expressed to the researcher their strong need to improve oral English in order to meet academic requirements (e.g., class discussions and presentations). Also, they perceived their English proficiency as relatively weak compared to local Hong Kong students. The participants included two MA students in language teaching (one male, 28, labelled with the pseudonym of MM; one female, 34, as MF) and two Year 1 BA English students (one male, 18, as BM; one female, 18, as BF). All were from mainland China and spoke Putonghua (Mandarin) as their L1. According to entrance scores, the two MA students had overall IELTS 6.5 but 5.5 in Speaking; the two BA students had Gaokao English scores of 106 and 109, approximately equivalent to IELTS band 5. All signed informed consent. There was no monetary compensation; instead, they received guided L2 speaking training at no cost. Pseudonyms were used in records to protect privacy.

3.2 Study design

The study ran for seven weeks, from Week 2 to Week 8 of the semester, and did not last longer to allow participants to focus on midterm and final course assignments due after Week 8. The design compared Human-to-Human versus Human-to-AI interaction within task-based L2 English speaking, with repeated exposure and three interview time points (weeks 2, 4 and 8) to capture perceptions over time. The participants received peer feedback and AI generated feedback for their task performance. Teacher (the researcher) feedback was provided each time after the interviews, targeting pronunciation, fluency, clarity, interactional strategies, and task fulfilment.

3.3 Procedures

Week 2 (pre-test):

Each pair (MA pair; BA pair) completed two interactive speaking tasks in English with their human partner: (1) an information-gap “spot the differences in two pictures” task; (2) a reasoning-gap “schedule a meeting for five teachers” task based on teaching timetables. Individually, each participant then completed the same two tasks with the Doubao AI chatbot. The order of AI versus human interaction was counterbalanced across participants to mitigate practice and novelty effects. Immediately after the Week 2 tasks, semi-structured interviews were conducted in Putonghua focusing on experiences, perceptions, and preferences regarding AI versus human interaction. Both the task performance and interviews were audio-recorded.

Weeks 3–8 (practice and follow-up):

After Week 2, participants were provided four speaking tasks of different types and increasing task complexity each week. Appendix 2 provides detailed description of all the tasks used from week 3 to week 8, with task types and task difficulty explained. The task selection criteria were based on decreasing task familiarity (Bui, 2014; Bui & Teng, 2019) and increasing cognitive demands (Robinson, 2001). They were required to complete the same tasks both with their human partner and with the AI chatbot outside class time on at least two separate days of the week. The sequencing (partner-first vs. AI-first; task order)

was counterbalanced across pairs and weeks to reduce practice effects on the learning experience and outcome. All task interactions were recorded by participants and uploaded to a designated Moodle site managed by the researcher.

Follow-up interviews were conducted in Putonghua at Week 4 (mid-study) and Week 8 (post-study) after completion of the week's tasks, probing evolving experiences and perceptions. Appendix 1 provides the interview guide.

3.4 Data analysis

A total of twelve interviews at weeks 1, 4 and 8 from the four participants were transcribed verbatim in Chinese and translated into English using ChatGPT 5.0. The bilingual researcher verified the translations against the original sound files to ensure that original meanings and affective nuances were captured. Then, the transcription was coded according to Glaser and Strauss' (1999) qualitative data analysis guide with the following steps.

Open coding: Line-by-line coding of the English translations identified salient concepts related to experiences and perceptions of human versus AI interaction (e.g., comfort, authenticity, interactivity, feedback credibility, efficiency, anxiety).

Axial coding: Codes were clustered into categories and subcategories, exploring relationships among conditions, interactions, and consequences (e.g., how task type, interlocutor, and week influenced perceived authenticity or confidence).

Selective coding: Core categories were integrated to articulate an emergent account of learners' experiences and perception change over time, and to identify contributing factors.

A constant comparative method was used across time points (Week 2, Week 4, Week 8) and interlocutor conditions (human vs. AI) to examine shifts in perceptions and to refine category properties. Researcher field notes were used as analytic memos to contextualise interview data and support interpretation. An audit trail of coding decisions was maintained.

An initial 5% random sample of the interview corpus was independently coded by the author and a colleague specialising in the same research area. 75.40% agreement on the assignment of main themes was reached. Discrepancies were compared and examined by both coders, with all disagreements resolved through active discussion and a revised coding guideline. Another 5% of the data was then coded independently, which achieved 91.30% percent agreement on main themes. After another round of discussion to reach consensus on remaining issues, the author proceeded to code the remainder of the dataset.

4 Findings and Discussion

4.1. RQ1: What are the participants' experiences and perceptions of interacting with human and AI-chatbot interlocutors, and what factors underlie them?

Across the three interviews, participants reported three recurring themes.

Theme 1: Initial engagement: Novelty, accessibility, and lowered affective barrier

The initial encounters with the AI interlocutor were characterised by a cluster of positively perceived attributes that distinguished it quite significantly from human interaction. While a pronounced novelty effect was universally reported, this was intertwined with immediate appreciation for the AI's accessibility and its capacity to create a low-anxiety practice environment than those from human interaction.

The novelty itself was a significant motivator. As MF, the female MA student, recounted,

“The first time I used the chatbot, I was excited. It felt like a new way to practice: no waiting, just instant conversation. I was curious to see how quickly and how well it could understand me.” (MF, 1stInt)

This sentiment was confirmed by BM, the male BA student, who described the experience as “like testing a new mobile game,” highlighting the experience of exploration and low-stakes experimentation that the AI technology was able to offer.

Beside novelty upon their encounter with the AI chatbot, these participants also valued AI’s accessibility and convenience. The AI’s asynchronicity and constant availability break practical barriers to practice. BF during the first interview noted, *“With AI I can practice at midnight and there is no need for scheduling with a partner. It fits into any small gap in my day.”* This observation highlights AI technology’s potential to provide far more opportunities and greater convenience for speaking practice than is possible with a peer or teacher, a factor considered crucial for fluency development (Bui, et al, 2019; Ellis, 2003; Okonkwo & Ade-Ibijola, 2021).

Another powerful initial theme was the immediate reduction in task-specific anxiety (An & Li, 2024a) and in particular, foreign language speaking anxiety (Alamri & Qasem, 2024; Horwitz et al., 1986). Participants frequently compared conditions of AI-mediated interaction with the social pressures of human interaction. MM expressed a common relief:

“It was fun, and I spoke more because I wasn’t scared to make mistakes in front of AI” (MM, 1stInt). MF provided a more detailed comparison, stating, *“With a human partner, especially at the beginning, I was constantly aware of my accent and my stuttered English. I felt I was wasting their time. With the AI, that pressure vanished. It felt like a judgment-free zone”* (MF, 1stInt)

This perception aligns with the notion of the online disinhibition effect (Qiu & Bui, 2022; Suler, 2004), where the non-human nature of AI as an interlocutor mitigates the pressure of peer or teacher evaluation. Such evaluation often inhibits output in the critical early stages of language practice. AI is able to lower affective filter and allows learners to speak more freely, a crucial condition for the “pushed output” Swain (1995) identified as essential for acquisition. It appears that the initial phase was not only about technological curiosity but also about a confluence of factors such as motivational novelty, practical accessibility, and perceived affective benefits. All this collectively positioned AI as a uniquely non-threatening and readily available interlocutor as compared with human interlocutors from the participants’ perspectives.

Theme 2: Perceived benefits of AI chatbot as a task partner beyond the initial stage

Besides the afore-mentioned initial novelty and constant availability of AI, the participants also stressed the perceived higher English proficiency of AI and its potentially more trustworthy feedback than that of peer or even teacher interlocutors. Several contrasted AI’s language competence with peers,

“The chatbot’s grammar and vocabulary were stronger than my classmate’s, so I felt I could trust its corrections more.” (BF, 1stInt) and *“It gave more examples, explanations and synonyms than my classmates could.”* (MM, 2ndInt)

Trust in perceived expertise is a known determinant of reliance on automation (Lee & See, 2004). For L2 learning, rich exemplars can support noticing and uptake (Kong & Bui, 2019; Schmidt, 1990, 2001), although quality varies. For example, MF stated,

“I trust the AI’s grammar check more than my friend’s because it’s based on a lot more data” (MF, 1stInt), while BM highlighted its resourcefulness, noting, *“When I’m stuck, the AI can*

always provide a wider range of vocabulary and sentence structures to use, which is really helpful” (BM, 2ndInt).

The students also reported reduced anxiety and increased comfort with AI. In terms of the perceived comfort, MM expressed his feeling,

“I felt less nervous with AI as I know it won’t be judgemental. My English is not that good and I don’t speak fluently because I need to pause a lot to think of the right words. The Doubao AI chatbot seems to understand whatever I say, no matter how poor my English sentences are. It seldom requests clarifications or interrupt my speech” (MM, 2ndInt), MF contrasted: “With a human partner, especially in front of the whole class, I worried about my accent and being slow” (MF, 2ndInt)

Although foreign language anxiety is a well-established construct (Horwitz et al., 1986; MacIntyre & Gardner, 1991), task-specific anxiety has just begun to attract attention in task-based research literature (An & Li, 2024a, 2024b; Aubrey, 2025). Participants in this research reported reduced task anxiety and increased comfort with the AI partner as compared with human interlocutors. Lower social risk with AI supports fluency attempts and self-initiated output (Suler, 2004; Swain, 1995), though limited social cues may reduce opportunities for pragmatic negotiation, which will be discussed further below.

Theme 3: Constant comparisons between peer, teacher and AI feedback

Participants anchored credibility in teacher assessment. They felt that teacher feedback is most reliable while AI-generated feedback was sometimes perceived as irrelevant or hallucinated. BM said,

“I still rely on teacher feedback the most. It’s targeted and more credible. AI feedback is quite robotic as it always starts with some praise and then negative feedback which may or may not be true” (BM, 2ndInt). BF added, “Sometimes the AI’s feedback didn’t match what I was talking about, so I wasn’t sure if I should follow its comments” (BF, 3rdInt)

This differentiated credibility reflects learners’ perceived hierarchy in power and precision of expert formative feedback (Hattie & Timperley, 2007; Lee, 2017). Reports of inapplicable or fabricated comments from AI align with AI hallucinations documented in the literature (Ji et al., 2023). This finding suggests a need for critical evaluation and trust calibration (Lee & See, 2004).

The students contrasted the interaction between humans and AI in terms of authenticity and social presence. Human interaction was perceived as “more real” since the students regarded the talk among humans as more authentic,

“With my classmate, it was like an actual conversation, where I responded to the tone and feelings. Regardless of the perfection in pronunciation and the related topics, I am still conscious that I am chatting with an AI and not a human” (BF, 3rdInt)

Another reflected,

“The chatbot was quite helpful, albeit somewhat mechanical at times—it would overlook some cultural nuances. It is hugely knowledge-rich, but it would lack the local and contextual understanding that my human friend would provide” (MF, 2ndInt)

Early social presence theory proposed (Short et al., 1976) and authenticity in pedagogy (van Lier, 1996) have long highlighted the fact that the social company in contextualised interaction affects their sociolinguistic development, given the richer affective and pragmatic cues (Kasper & Rose, 2002). While AI systems can handle turn-taking to a large extent, they lack the ability to identify the appropriate social codes in terms of pragmatics. All this contributes to the perception of “inauthentic” conversations among the learners.

In addition, the participants noticed the distinctive interactional pacing and turn management in human and AI conversations. MM expressed,

“AI responded quickly, sometimes a bit too long. I had to cut in. My classmates, like all others, would pause normally. I could jump in, and ask a question” (MM, 2ndInt). BF added, *“The voice of Doubao’s AI chat feature is very flat and unemotive. Together with the lack of facial expressions and eye contact, at times, when chatting with the AI, my mind tends to wander, making it rather easier to get distracted than when chatting with a human ”* (BF, 3rdInt).

Differences in turn-taking rhythm reflect AI’s insufficiency to detect extra-linguistic cues. The current AI technology commonly available to learners only receives either audio or textual input, with a lack of sensitivity to social cues such as facial expression and body language that characterises human-human interactions. This may cause untimely response to the participant’s change of tone or emotion, which in turn may affect learners’ sustained interest in topic development and negotiation of meaning (Long, 1996). In addition, prolonged AI turns can increase cognitive load for processing while reducing opportunities for learner output. BF’s comments suggest challenges in task engagement during AI-assisted task interaction.

Although state-of-the-art text-to-speech systems achieve remarkable clarity, many default voices are optimised for neutrality rather than affective attunement. This was noticed by the learners,

“The voice stayed very flat; prosody didn’t change even when I shared something exciting or difficult” (MM, 2ndInt) and *“It didn’t cheer me up or slow down when I struggled; it just kept the same tone”* (MM, 3rdInt).

This lack of emotion in AI voice can make practice feel “mechanical,” especially in sensitive moments when learners seek validation or encouragement. In L2 acquisition, prosody is central for signalling discourse structure, focus, and stance (Wennerstrom, 2001). Prosodic variation helps listeners segment input, identify salient items, and infer pragmatic meaning (Cutler et al., 1997). When AI interlocutor prosody is too flat, learners lose cues that typically reduce processing effort and guide uptake. Research in speech synthesis also documents the challenge of expressive prosody; while neural text-to-speech (TTS, e.g., WaveNet) improves naturalness, robust, context-sensitive emotional prosody remains difficult (Skerry-Ryan et al., 2018). The result is a mismatch: learners expect an interlocutor to signal empathy and emphasis through voice, but the AI’s prosody stays unvaried, decreasing perceived social presence (Nass & Brave, 2005).

Moreover, the students encountered problems related to technology and interfaces. Some students suffered from friction influencing talk,

“Sometimes the connection lagged and I lost my train of thought (when using AI)” (MM, 3rdInt). Another contribution quoted, *“If the chatbot didn’t hear me well, it replied off-topic”* (BF, 2ndInt) and *“Doubao often delayed; the lag broke my flow and made the talk feel strange”* (BM, 2ndInt).

It appears that technical reliability influences AI’s usefulness and flow (Venkatesh et al., 2003) as perceived by the students. Latency and speech recognition mistakes in AI interaction responses can generate cognitive load and emotional turbulence that hinder the conditions for effective practice and output.

4.2 RQ2: How, if at all, do these experiences and perceptions evolve over time?

The three time points revealed a clear trajectory and fluctuation in learners’ affect, trust, and preferences toward AI interlocutors, while their views of human interaction remained comparatively stable. Building

on the original observations of novelty waning, trust recalibration, improving prompt literacy, growing comfort with human partners, and a shift to hybrid strategies, participants' week-by-week accounts add specific reasons for changing attitudes: the emotionless and flat prosody of the AI voice, the absence of extralinguistic cues such as gesture and facial expression, unfamiliar or "strange" interactional patterns with delayed responses (specifically with Doubao), and a perceived lack of emotional support. In contrast, perceptions of peer interaction, which is considered helpful and familiar in interactional pattern, remained consistent across interviews, though peers were seen as less proficient for scaffolding and high-quality feedback compared to teacher input.

Theme 1: Trajectory of Affect and Trust: From Excitement to Lukewarm Engagement to a Colder Stance

Consistent with the novelty effect described earlier, participants framed the first interview around high anticipation and curiosity about AI practice. One learner summarised the arc succinctly:

"At first, chatting with AI felt exciting and limitless; by Week 8 my enthusiasm almost disappeared" (MM, 3rdInt)

The second interview (Week 4) was characterised as "lukewarm," with sustained use but diminishing emotional engagement:

"By Week 4, I felt lukewarm; I still used the chatbot for oral English practice but not as much" (MM, 2ndInt).

By Week 8, the affect had cooled further as limitations in affective and interactional qualities became salient:

"I know it's there and I can use it anytime, anywhere, but I don't often think of it as a practice partner unless my partner or teacher emails remind me that I need to do the tasks and upload it to Moodle. There isn't much motivation or sense of urgency, it seems." (MM, 3rdInt).

This trajectory is consistent with well documented trends in the reliance on automation and technology in general, namely the periods of optimism, followed by the recalibration and development of trust in the system when encountering boundary conditions and instances of error (Lee & See, 2004). Theories of technology acceptance (Venkatesh et al., 2003) would then predict that ease of use and perceived use are the primary drivers in the adoption phase, although the long-term usability and fit with social and perceived needs would establish the performance and relationship with the technology tools and systems themselves. In L2 learning, the role and impact of affective factors such as motivation, anxiety, and perceived efficacy play pivotal roles (Bui, 2023; Dörnyei, 2005). AI's affective signals are often ineffective or misaligned when the tone and the response and delay in the prosody represent a lack in perceived social presence. Then, perceived social presence drops, which weakens satisfaction and engagement (Qiu & Bui, 2022). Learners have to recalibrate their trust in AI after experiencing hallucinations or irrelevant feedback (Ji et al., 2023). All this results in re-anchoring credibility in teacher feedback and adopting selective reliance on AI.

Theme 2: Social presence and extralinguistic cues as reasons of attitude change

A major source of attitude cooling during the changing perception over AI use was the perceived absence of support and social presence from the AI. As a student was quoted,

"The AI voice was emotionless and flat; it didn't feel supportive" (MM, 2ndInt).

Participants pointed out the lack of nonverbal communication, such as nods and smiling, making it difficult to read the intention and feel encouraged because,

“Without the gestures and the facial expressions, I don’t think I could have determined whether it understood and cared” (MM, 2ndInt). In contrast, human peers offered reliable paralinguistic reassurance: *“My classmates are familiar; their nods and smiles helped me relax. I need these kinds of responses to sustain a conversation”* (MM, 1stInt)

These perceptions map cleanly onto social presence theory (Bickle et al., 2019; Short et al., 1976). Channels that convey more nonverbal information (e.g., facial expressions, gesture, prosody) support affective attunement, empathy, and clarification of intent. A lack or reduction of social presence may impose additional cognitive effort and diminish trust, comprehension, and satisfaction (Bickle et al., 2019). The absence of kinesic and facial signals is not trivial in L2 settings: a sizable body of research shows that gestures and facial cues scaffold comprehension, disambiguate reference, and support lexical retrieval and pragmatic alignment (Kendon, 2004; Gullberg, 2006). In second-language listening, visual cues from the speaker, namely gestures and facial expression, enhance comprehension and recall (Sueyoshi & Hardison, 2005). Without these cues, learners lose immediate feedback on the success of their message, and their ability to calibrate tone, politeness, and stance suffers (Kasper & Rose, 2002).

Theme 3: Improving AI literacy through task practice to maximise AI benefits

In the face of both AI’s benefits and limitations, the learners displayed agency in the use of AI interaction in L2 task-based learning and increased their AI literacy, especially in their prompting skills to mitigate the afore-mentioned insufficiencies of AI. The learners reported adapting the prompts to manage AI performance,

“I learned to assign the chatbot roles, such as ‘be a strict IELTS examiner when you are giving me feedback on this task,’ and the tasks worked better”, said MF (3rd Int). BM noted, *“When I asked it for short answers and follow-up questions, the exercise seemed like an actual conversation”* (BN, 2ndInt)

This is indicative of the use of metacognitive strategies (Oxford, 2011) and the user-design approach in the scaffolding (Vygotsky, 1978). The emergence of prompt literacy signals learners’ agency in co-constructing interactional affordances, improving alignment with goals.

The participants also aimed at personalising the task fit and the preferences. They utilised AI for the purpose of brainstorming, fluency and sample phrases/sentences while taking the advantage of humans for accuracy/pragmatics. Participants differentiated function by task. BM mentioned,

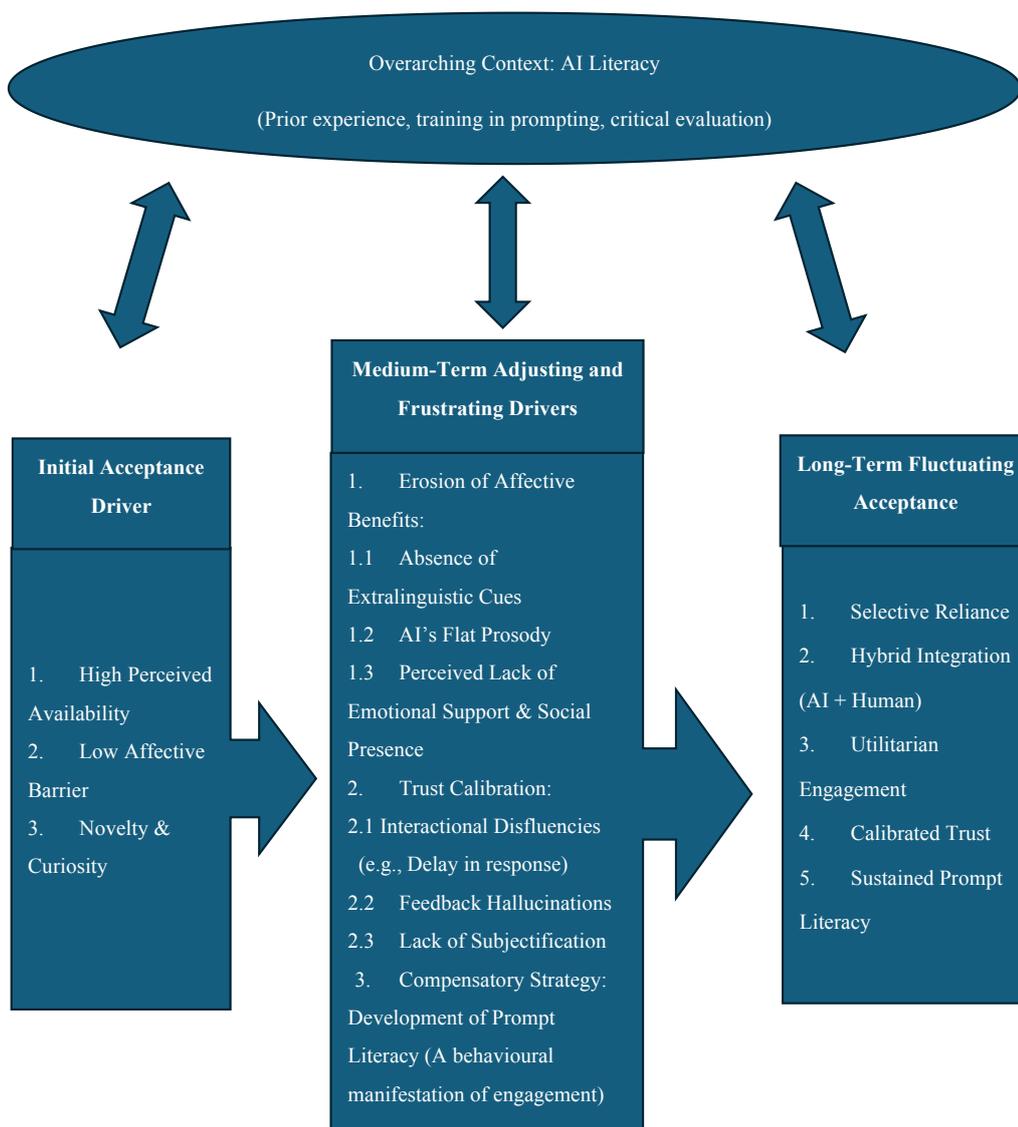
“AI was great for brainstorming and building vocabulary quickly” (BM, 2ndInt) while MF said, *“I can learn the idioms sometimes from Doubao, but in terms of smoothing out the tone and polite expressions, I would like a human conversation partner or teacher—that just feels much more natural. AI answers are quite generic and formal”* (MF, 3rdInt)

These needs and preferences align with the following complementary strengths of each type, where AI can provide quick lexical variation and fluency practices with low stakes (Okonkwo & Ade-Ibijola, 2021), and when paired with human interaction, it can provide better opportunities for pragmatic appropriateness and receiving feedback (Kasper & Rose, 2002). The emerging prompt literacy skill enables learners to compensate for AI weaknesses.

4.3 A general pattern of evolving perceptions of AI in L2 learning

Based on the above findings and discussion, a preliminary model of AI acceptance in task-based language learning is summarised in Figure 1, which includes an overarching context of AI literacy and three stages of evolving acceptance that reflect learners’ trajectory of L2 learning with human and AI partners.

Figure 1

A Preliminary Model of AI Acceptance in Task-Based Language Learning

The overarching AI literacy component contextualises the entire process, influencing every stage. A learner's AI Literacy, which comprises their prior experience with AI, and crucially, any formal training in prompting and critical evaluation, forms the background that shapes their AI interaction journey. As the double arrows indicate, each stage of the learning journey will potentially reinforce their AI literacy skills that feeds into the next stage of learning.

The three stages of AI acceptance are characterised by initial excitement to realising the complexity of AI use in context, with its pros and cons exposed, and to finally learn to adjust their learning strategies and expectations. As Figure 1 indicates, learners have enthusiastic acceptance for AI interlocutors due to accessibility and low anxiety.

After repeated use, however, AI interlocutor's initial appeal is undermined by specific, experiential limitations. Firstly, the "Low Affective Barrier" is eroded by the Absence of Extralinguistic Cues (no nods, smiles) and Flat Prosody (emotionless voice) of AI. These issues limit AI's provision of the emotional support and encouragement that characterised human interaction, which results in the "colder stance" and a perceived lack of social presence (Bickle, et. al., 2019). Secondly, the "High Perceived Usefulness" is challenged by Interactional Disfluencies (e.g., response delays) and other technical issues.

Natural conversational flow may be disrupted and occasionally, irrelevant feedback (“hallucinations”) will be given. All these factors push a recalibration of trust, where learners learn to rely on the AI less and fall back on teacher feedback. More importantly, learners’ perceived lack of subjectification (Jin & Wang, 2025), which significantly reduces their responsible attitude towards, and responsive use of, the AI Chaobot. Prompt literacy emerges as a key behavioural and metacognitive strategy. It is a direct response to the AI’s limitations. Learners actively try to compensate for the lack of social presence and interactional disfluencies by engineering the interaction through better prompts (e.g., “ask me follow-up questions,” “give short answers”).

The final component “Long-Term Fluctuating Acceptance” describes a more stabilised, utilitarian relationship with AI interlocutor. Learners developed selective reliance on AI. With the initial excitement fades away, learners gradually develop a more practical preference for a hybrid mode of learning. AI is then chosen strategically for specific tasks, such as expanding vocabulary or drilling fluency and pronunciation. Task involving more nuanced, pragmatic functions will be reserved for human partners. This utilitarian engagement is sustained by the continued application of prompt literacy to navigate its limitations.

5 Potential Pedagogical Implications

The role of human interaction, which is indispensable in task-based language teaching, cannot be replaced by AI chatbots as of the present time (Jin & Wang, 2025). As the study shows, human interlocutors will provide sociopragmatic richness, authentic emotional and nonverbal (gestures, intonation, smiles) cues, and the ability to co-construct meaning through natural turn-taking that are irreplaceable. These latter elements are crucial for the development of pragmatic competence, for learner confidence building through relational support, and for creating conditions that would actually allow negotiation of meaning and pushed output to take place. On the other hand, a hybrid pedagogical model is suggested in which AI and human interaction are sequenced strategically. AI should be used for low-stakes, high-frequency practice aimed at building fluency and increasing lexical resources, capitalising on its availability and low anxiety. This should be systematically followed by human-interactive tasks with peers or teachers. These in-person tasks focus on refining pragmatic appropriateness, receiving nuanced feedback on accuracy and tone, and engaging in the authentic, unpredictable negotiation that drives deeper acquisition.

To best take advantage of AI, educators need to proactively train students in prompt literacy (Tour & Zadorozhnyy, 2025) and critical evaluation skills. In more detail, this involves teaching learners how to successfully engineer their interactions with AI: assigning roles to the tool. For example, “act as a strict examiner”; managing turn-taking by instructing the chatbot to give shorter answers and ask follow-up questions; and scoping feedback requests such as “correct only my grammar.” Students also need to be directed to critically assess AI-generated feedback for possible inaccuracies or hallucinations, anchoring their ultimate trust in teacher guidance. It is through such metacognitive and strategic skills that learners will shift from being passive users to active agents who can capitalise on the strengths of AI, including instant vocabulary support and brainstorming aid, but mitigate its limitations, leading to enhanced self-regulation of their learning and more effective completion of oral tasks.

6 Conclusion

This research traced the trajectory of learner perceptions during their engagement with human as well as AI conversants in L2 speaking tasks over a seven-week period. Though novelty, unparalleled availability, and substantial decreases in speaking anxiety initially welcomed the arrival of AI, permitting enhanced output, perceptions have subsequently cooled to take into consideration interactional, affective

realities. Findings reveal learner preference for engagement with human interlocutors for their facility for authentic communication, substantial social presence, and pragmatic support, while being adept at selectively employing AI for particular purposes, such as fluency training, lexical enhancement, with prompt literacy being an indicator of substantial learner agency in optimising such interactions. Overall, there is evidence for an optimal blended model for L2 learning tasks, taking advantage of what each, namely human conversants and AI, uniquely offers.

Several limitations exist in this investigation, including having only one type of AI chatbot, named Doubao, while other types with varying strengths, for example, enhanced synthetic voices, could result in varied perceptions. Another consideration is having only four participants, which, although contributing to in-depth information, could not result in generalisation for other groups, for example, varied levels, cultures, or schools. Future research needs to be undertaken on a more extensive group to provide an in-depth analysis regarding AI's contribution to second language acquisition.

Acknowledgement

The author would like to thank the four volunteers for their active participation and Dr. Maggie MA for her useful suggestions. DeepSeek (V3.2) was used to proofread the language of parts of the manuscript, but the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix 1 Interview Guide in brief (full version available upon request)

Opening

- Brief purpose reminder; confirm consent/confidentiality.
- Reconstruct the most recent tasks:
 - Human: What happened? Key moments? What was easy/difficult? How did you feel?
 - AI: What happened? Key moments? Any offtopic/confusing replies? How did you feel?

Core domains (ask at each timepoint)

- Comfort, anxiety, authenticity:
 - How comfortable with your human partner? Why?
 - How comfortable with AI? Why?
 - How “real/natural” did each feel? Why?
- Learning/process:
 - In which setting did you practice more effectively (fluency, accuracy, vocabulary, pragmatics)? Why?
 - Examples of clarification/repair/reformulation. What triggered noticing or selfcorrection?
- Interaction/usability:
 - Compare interactivity (turntaking, speed, followup questions, coherence).
 - Did you change how you prompt the AI? How/why?
 - Any technical/interface issues that affected speaking?
- Preferences/contexts:
 - For similar tasks, which do you prefer and why?
 - Tasktype preferences (e.g., informationgap vs. reasoninggap; brainstorming vs. planning).

- Evidence (optional, timelimited):
 - Play 30–60s clip; ask what you were thinking, why you responded that way, what you'd change.
- Feedback/external factors:
 - Effect of the researcher's feedback on your next tasks.
 - Other influences (mood, workload, familiarity, environment).

Change over time

- Week 4: Compared to Week 2, what changed with human and with AI? Why?
 - Quick ratings (1–10): comfort (human, AI); perceived usefulness (human, AI). Why?
 - Any preference shifts? In which situations?
- Week 8: Across all weeks, what stayed the same and what changed? Any turning points? Perceived improvement (how, where noticed)?
 - Ideal future mix of human and AI; final ratings (1–10): comfort, anxiety, usefulness.

Tailored followups

- “Last time you mentioned X; did it change? Why?”
- “In your recording you did Y with human but not AI; why?”

Appendix 2 Detailed Task Description for each week

Week	Task 1	Task 2	Task 3	Task 4
Week 3	Spot the Difference (simple) Find differences in a familiar classroom scene in two pictures	Giving Directions on a Map Guide a partner to a location on a simple map.	Describe and Draw Describe a simple abstract shape for partner to draw.	Schedule a Meeting (Simple) Find one common free slot in two simple timetables.
Week 4	Spot the Difference (Detailed) Differences in more detailed, real-life photos of a kitchen).	Picture Sequencing (Daily Routine) Arrange 4 picture cards in logical order to tell a simple story.	Deciding on a Restaurant Choose a restaurant from options based on shared preferences.	Giving Directions (with Obstacles) Navigate a map with road closures or “no entry” points.
Week 5	Spot the Difference (Complex Scene) Find subtle differences in a crowded festival scene.	Narrative Task (Picture Story) Create a story from a set of related but unsequenced pictures.	Planning a Weekend Trip Plan a one-day itinerary for a city visit within a budget.	Schedule a Meeting (with Conflict) Reschedule a meeting when the first-choice time is unavailable.
Week 6	Spot the Difference (Abstract Art) Describe and compare two abstract images with complex patterns.	Giving Directions (with Landmarks) Give directions using prominent landmarks and estimated times.	Pros and Cons Discussion Discuss the advantages and disadvantages of two holiday destinations.	Deciding on a Gift Choose a birthday gift for a friend from a catalog with conflicting information.

Week	Task 1	Task 2	Task 3	Task 4
Week 7	Picture Story (Inferring Emotions) Sequence a story and infer characters' feelings and motivations.	Ranking Task (Survival Items) Rank items for a desert island survival scenario and justify choices.	Resolving a Complaint Role-play: one is a customer with a complaint, the other is a manager finding a solution.	Schedule a Meeting (Multiple Constraints) Find a common slot for five people considering personal preferences, not just availability.
Week 8	Narrative Task (Create a Story) Collaboratively invent a story based on a single picture prompt.	Ethical Dilemma Discussion Discuss the pros and cons of a simple ethical dilemma (e.g., lying to protect a friend).	Giving Complex Directions Give directions to a destination that involves multiple transport modes.	Spot the Difference & Explain Why Find differences and discuss <i>why</i> they might exist (e.g., different times of day).

References

- Akiyama, Y., & Saito, K. (2016) Development of comprehensibility and its linguistic correlates: A longitudinal study of video-mediated telecollaboration. *The Modern Language Journal*, 100(3), 585-609. <http://dx.doi.org/10.1111/modl.12338>
- Alamri, W., & Qasem, F. (2024). Foreign language speaking anxiety in relation to gender: Students and teachers' perceptions. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2423440>
- An, H. & Li, S. (2024a). Anxiety in task-based language teaching: A state-of-the-art review. In S. Li (Ed.), *Individual Differences and Task-Based Language Teaching* (pp. 52-83). John Benjamins. <https://doi.org/10.1075/tblt.16.02anh>
- An, H. & Li, S. (2024b). (2024b). Task-specific writing anxiety and self-efficacy are separate from general L2 writing anxiety and self-efficacy and they have differential associations with the effects of written corrective feedback in pre-task and within-task planning. *System*, 126, 103480. <https://doi.org/10.1016/j.system.2024.103480>.
- Aubrey, S. (2025). The relationship between task-specific anxiety, enjoyment, and use of planned content across discourse stages of second language learners' spoken task performances. *Learning and Individual Differences*, 120, 102677. <https://doi.org/10.1016/j.lindif.2025.102677>
- Bickle, J. T., Hirudayaraj, M., & Doyle, A. (2019). Social presence theory: Relevance for HRD/VHRD research and practice. *Advances in Developing Human Resources*, 21(3), 383–399. <https://doi.org/10.1177/1523422319851477>
- Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology*, 4(1), 111–125. <https://doi.org/10.64152/10125/25089>
- Bui, H. Y. G. (2014). Task readiness: Theoretical framework and empirical evidence from topic familiarity, strategic planning, and proficiency levels. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 63–93). Amsterdam: John Benjamins. <https://doi.org/10.1075/tblt.5.03gav>
- Bui, G. (2023). A Dual-Motivation System in L2 and L3 learning: A theoretical framework and pedagogical application. *Languages*, 8(1), 69. <https://doi.org/10.3390/languages8010069>
- Bui, G. (2024). There is more to the task: A response to Ellis (2024). *International Journal of TESOL Studies*, 6(4), 14-19. <https://doi.org/10.58304/ijts.20240402>

- Bui, G. (2025). Task-based language teaching. In H. Nesi & P. Milin (Eds.), *International encyclopedia of language and linguistics* (3rd ed.) (pp.1-6). Elsevier. <https://doi.org/10.1016/B978-0-323-95504-1.00434-8>
- Bui, G., Ahmadian, M., & Hunter, A. (2019). Spacing effects on repeated L2 task performance. *System*, 81, 1–13. <https://doi.org/10.1016/j.system.2018.12.006>
- Bui, G., & Skehan, P. (2018). Complexity, fluency and accuracy. In J. Liantas (Ed.), *The TESOL encyclopedia of English language teaching* (pp.1–8). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118784235.eelt0046>
- Bui, G., & M. F. Teng. (2019). Task planning and task-readiness. In J. Liantas (Ed.), *The TESOL encyclopedia of English language teaching*. Hoboken, NJ: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118784235.eelt0972>
- Bui, G., & Wong, C. H. (2021). From linguistic skills to pragmatic competence: The role of functional adequacy in task-based language teaching and learning. *Asian Journal of English Language Teaching*, 30, 61-76. https://cup.cuhk.edu.hk/image/catalog/journal/jpreview/AJELT30_61-76_full.pdf
- Chapelle, C. A., & Sauro, S. (Eds.). (2017). *The handbook of technology and second language teaching and learning*. Wiley-Blackwell. www.doi.org/10.1002/9781118914069
- Cope, B., & Kalantzis, M. (2024). A multimodal grammar of artificial intelligence: Measuring the gains and losses in Generative AI. *Multimodality & Society*, 4(2), 123–152. <https://doi.org/10.1177/26349795231221699>
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature Review. *Language and Speech*, 40(2), 141-201. <https://doi.org/10.1177/002383099704000203>
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition* (1st ed.). Routledge. <https://doi.org/10.4324/9781410613349>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Ellis, R. (2024). Task-based and task-supported language teaching. *International Journal of TESOL Studies*, 6(4), 1-13. <https://doi.org/10.58304/ijts.20240401>
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-based language teaching: Theory and practice*. Cambridge University Press. <https://doi.org/10.1017/9781108643689>
- Gattupalli, S., R., Maloy, W., & Edwards, S. A. (2023). Prompt literacy: A pivotal educational skill in the age of AI. *College of Education Working Papers and Reports Series*. <https://doi.org/10.7275/3498-wx48>
- Glaser, B., & Strauss, A. (1999). *Discovery of grounded theory: Strategies for qualitative research* (1st ed.). Routledge. <https://doi.org/10.4324/9780203793206>
- Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology*, 27(2), 6–27. <https://doi.org/10.64152/10125/73501>
- Gullberg, M. (2006). Handling discourse: Gestures, reference tracking, and communication strategies in early L2. *Language Learning*, 56(1), 155–196. <https://doi.org/10.1111/j.0023-8333.2006.00344.x>
- 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>
- Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues* (1st ed.). Routledge. <https://doi.org/10.4324/9780203012215>

- Jepson, K. (2005). Conversations—and negotiated interaction—in text and voice chat rooms. *Language Learning & Technology*, 9(3), 79–98. <https://doi.org/10.64152/10125/44033>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Ishii, E., Bang, Y., Madotto, A., & Fung, P. N. (2023). A survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 248. <https://doi.org/10.1145/3571730>
- Jin, H., & Wang, S. (2025). Why not replace people with machines? Exploring Artificial Intelligence’s (in) capacity for education as subjectification. *Journal of Philosophy of Education*, qhaf087. <https://doi.org/10.1093/jopedu/qhaf087>
- Kasper, G., & Rose, K. (2002). *Pragmatic development in a second language*. Blackwell.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511807572>
- Kong, A., & Bui, G. (2019). Disambiguating recasts with prosodic and extra-linguistic cues in task-based interactions among young learners. *Language Teaching for Young Learners*, 1(2), 187–206. <https://doi.org/10.1075/ltyl.18008.kon>
- Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Springer. <https://doi.org/10.1007/978-981-10-3924-9>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate Reliance. *human factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392
- 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.
- Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.
- Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32(2), 265–302. www.doi.org/10.1017/S0272263109990520
- MacIntyre, P. D., & Gardner, R. C. (1991). Language anxiety: Its relationship to other anxieties and to processing in native and second languages. *Language Learning*, 41(4), 513–534. <https://doi.org/10.1111/j.1467-1770.1991.tb00691.x>
- McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation*, 5(1), 98–118. <https://doi.org/10.1075/jslp.16034.mcc>
- Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50-68). Routledge. www.doi.org/10.4324/9781315676968-4
- Moorhouse, B. L., & Wong, K. M. (2025). *Generative artificial intelligence and language teaching*. Cambridge University Press.
- Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human–computer relationship*. MIT Press.
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033. <https://doi.org/10.1016/j.caeai.2021.100033>
- Oxford, R. L. (2011). *Teaching & researching: Language learning strategies* (1st ed.). Routledge. <https://doi.org/10.4324/9781315838816>
- Qiu, X., & Bui, G. (2022). “We are responsive on Zoom, but...”: L2 learners’ perceptions of and attitudes towards speaking tasks in the physical and virtual classroom settings. *International Journal of Computer-assisted Language Learning and Teaching*, 12(2), 1-12. www.doi.org/10.4018/IJCALLT.291535

- Qiu, X., Lo, Y. Y., Ge, H., & Bui, G. (2025a online first). Second language learner engagement in computer-mediated interactive oral tasks. *Studies in Second Language Learning and Teaching*. <https://doi.org/10.14746/sslst.41317>
- Qiu, X., Bui, G., Yao, X. (2025b online first). Task type, task elements, and L2 grit: Predictors of ESL learners' behavioural engagement in computer-mediated oral tasks. *System*, 103925. <https://doi.org/10.1016/j.system.2025.103925>
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. <https://doi.org/10.1093/applin/22.1.27>
- Satar, H.M., & Özden, N. (2008), The effects of synchronous CMC on speaking proficiency and anxiety: Text versus voice chat. *The Modern Language Journal*, 92, 595-613. <https://doi.org/10.1111/j.1540-4781.2008.00789.x>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139524780.003>
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. John Wiley.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Skerry-Ryan, R.J., Battenberg, E., Xiao, Y., Wang, Y., Stanton, D., Shor, J., Weiss, R.J., Clark, R.A., & Saurous, R.A. (2018). Towards end-to-end prosody transfer for expressive speech synthesis with Tacotron. International Conference on Machine Learning. <https://doi.org/10.48550/arXiv.1803.09047>
- Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal*, 87(1), 38–57. <https://doi.org/10.1111/1540-4781.00177>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, 12(17), 8427. <https://doi.org/10.3390/app12178427>
- Sueyoshi, A. and Hardison, D.M. (2005), The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(4), 661-699. <https://doi.org/10.1111/j.0023-8333.2005.00320.x>
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 125–144). Oxford University Press.
- Tour, E., & Zadorozhnyy, A. (2025). Conceptualizing and operationalizing prompt literacy for English language learners. *Journal of Adolescent & Adult Literacy* 69(3), e70020. <https://doi.org/10.1002/jaal.70020>
- Van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy and authenticity*. Longman.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press. <https://doi.org/10.1093/oso/9780195143218.001.0001>
- Xing, C., & Saeed, M. A. (2026). A systematic review on artificial intelligence (AI) technologies in ESL/EFL speaking skills. *International Journal of TESOL Studies*, 8(2), 240-270. <https://doi.org/10.58304/ijts.250908>

Ziegler, N. (2016). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*, 38(3), 553–586. <https://doi.org/10.1017/S027226311500025X>

Dr. Gavin Bui is an Associate Professor and Associate Head in the English Department at The Hang Seng University of Hong Kong, where he also serves as the Director of the MA in English Language Teaching and Assessment programme. His primary research interest lies in task-based language teaching and learning. His recent publications have appeared in international journals such as the *RELC Journal* (2026), *System* (2026), *Studies in Second Language Learning and Teaching* (2025), and *Studies in Second Language Acquisition* (2024).

ORCID: <https://orcid.org/0000-0002-1567-9074>