GLODAL

OPEN ACCESS

Brief Reports in Language Research

Brief Reports in Language Research, 1, 260103 (2026)
https://doi.org/10.65553/BRLR.260103

Genuine vs. Virtual Social Presence in AI-Assisted Task-Based Interaction: An Exploratory Study on Willingness to Communicate and L2 Performance Across Task Complexity

Gavin Bui*

Department of English, The Hang Seng University of Hong Kong, China

Abstract

This empirical brief report investigates how genuine (human-to-human) and virtual (human-to-AI) social presence interact with task complexity to influence second language (L2) learners' Willingness to Communicate (WTC) and oral task performance. Eighteen Hong Kong Cantonese-speaking university learners of English completed simple (three-element with low reasoning demand) and complex (six-element with high reasoning demand) communicative direction-giving tasks under one of two conditions: interacting either with an AI chatbot (virtual social presence) or with a peer (genuine social presence). Quantitative analyses examined situational WTC and Complexity, Accuracy, and Fluency (CAF) measures, complemented by post-task semi-structured interviews. Descriptive statistics revealed that virtual social presence elicited higher self-reported WTC, a trend interviewee attributed to reduced face threat and lower communicative anxiety. Conversely, genuine social presence was associated with greater syntactic complexity, possibly through peer feedback and interactional modifications, with a tendency towards mid-clause pausing. These findings tentatively suggest that the differential affordances of artificial and human interlocutors: conversational AI may offer a low-anxiety context that supports communicative willingness, whereas human interaction may be more conducive to syntactic elaboration.

Keywords: Task-based language teaching, Willingness to communicate, Artificial intelligence, Social presence, Task complexity

Submission date: August 20, 2026; Revision date: October 26, 2026; Online: November 10, 2026

Copyright © 2026
The authors

(corresponding author: gavinbui@hsu.edu.hk)

Introduction

A central tenet of Task-Based Language Teaching (TBLT) is that second language (L2) acquisition is optimised when learners engage in meaningful, goal-oriented interaction requiring the communicative exchange of information (Bui, 2026a; Ellis, 2003; Long, 2015). For interaction to foster acquisition, however, learners must first be willing to take communicative risks. In second language acquisition (SLA), this disposition is conceptualised as *Willingness to Communicate* (WTC), defined by MacIntyre et al. (1998) as a learner's readiness to enter into discourse at a specific moment with a specific interlocutor. While WTC was originally treated as an enduring personality trait, contemporary research recognises it as a dynamic, situationally contingent construct that fluctuates according to affective states, cognitive demands, and social contextual variables (Aubrey & Yashima, 2023; Bui, 2026c; Cao, 2011; Peng, 2025).

The rapid advent of conversational Artificial Intelligence (AI) and large language model chatbots provides L2 pedagogy with accessible, on-demand speaking partners. Fathi et al. (2024) reported that EFL learners interacting with an AI chatbot outperformed a face-to-face group on speaking skills and WTC, whereas Bui (2026b) showed that learners' initial enthusiasm for AI interlocutors cooled over seven weeks as limitations in social presence, prosody, and interactional authenticity became salient. Wang (2026) similarly found that multimodal generative AI supported task preparation and lowered affective barriers but yielded limited gains in fluency and delivery confidence, with learners continuing to rely on human interlocutors for validation. However, these studies either compared AI with conventional instruction rather than isolating interlocutor presence within a controlled task-based design (Fathi et al., 2024), or did not systematically manipulate task complexity (Bui, 2026b; Wang, 2026); no study has experimentally crossed interlocutor type (genuine vs. virtual social presence) with task complexity to examine their joint effects on situational WTC and CAF measures.

To understand the differentiated affordances derived from human and AI as interlocutors in language teaching and learning, it is helpful to explore the notion of social presence. Social presence in early literature refers to the degree to which an interlocutor is perceived as salient and "real" within an interaction, together with the resulting sense of affective connection and interpersonal relationship that such perceived presence affords (Short et al., 1976). More recently, Garrison et al. (2000) define it as the ability of participants in a community of inquiry to project themselves socially and emotionally as "real" people through the medium of communication being used, while Gunawardena and Zittle (1997) operationalise the construct as the degree to which a person is perceived as a "real person" in mediated communication. To address the above research gaps, this study distinguishes between two distinct forms of communicative presence:

> **Genuine Social Presence:** Grounded in human-to-human interaction, characterised by real-time interpersonal dynamics, reciprocal social accountability, face-concerns, and authentic negotiation of meaning.
>
> **Virtual Social Presence:** Simulated through human-to-machine interaction, characterised by perceived accessibility, conversational patience, and non-judgmental reception, but fundamentally devoid of authentic human consciousness or genuine communicative consequence.

In East Asian English as an Additional Language (EAL) environment, such as Hong Kong, genuine social presence can introduce communicative inhibition. As Wen and Clément (2003) observe, culturally rooted face-saving concerns and evaluation anxiety frequently restrict

learners' willingness to speak in front of peers and instructors. Interacting under virtual social presence may theoretically mitigate these affective barriers via the *online disinhibition effect* (Qiu et al., 2026; Suler, 2004), creating a psychologically secure environment where learners feel emboldened to produce target-language utterances.

Simultaneously, task-based research demonstrates that communicative behaviour cannot be evaluated independently of task design. In Robinson's (2001, 2007) Cognition Hypothesis (CH), manipulating task complexity along *resource-directing* dimensions (e.g., increasing the number of elements and/or reasoning demands) channels learners' attention towards linguistic encoding, typically enhancing accuracy and syntactic complexity at the expense of fluency. Under the CH framework, social presence may become resource-depleting factor that potentially lower linguistic complexity and accuracy. In contrast, Skehan's (1998, 2009) Limited Attentional Capacity Model posits that complex communicative processing may imposes attentional trade-offs between forms and meaning. When a complex task is situated within an anxiety-inducing human-to-human interaction, the compounded cognitive and interpersonal load may suppress WTC. Then, a patient AI conversational partner may buffer against these complexity-induced communicative breakdowns.

To our knowledge, few empirical studies have systematically isolated the interaction between interlocutor presence (genuine vs. virtual) and task complexity (simple vs. complex) within an established task-based paradigm. The present brief report addresses this gap by investigating four interrelated research questions (RQs):

RQ1: How does social presence type (genuine human vs. virtual AI social presence) influence L2 learners' situational Willingness to Communicate?

RQ2: How does task complexity (simple vs. complex map-based tasks) moderate WTC across the two social presence conditions?

RQ3: What measurable differences emerge in learners' oral production profiles (Complexity, Accuracy, and Fluency) across social presence and complexity conditions?

RQ4: How do learners perceive the communicative and learning affordances of genuine versus virtual social presence in task performance?

Methods

Research Design and Context

This study adopted a 2 (social presence types: between-participant) x 2 (task complexity: within-participant) mixed-methods design plus interviews. Participants were randomly assigned to one of the following two social presence groups:

> **Virtual Social Presence (AI Chatbot Condition):** Learners performed tasks with a conversational AI agent (*Doubao*), prompted to act as an English-speaking tourist seeking directions.
>
> **Genuine Social Presence (Human Peer Condition):** Learners performed tasks in dyads, where one student assumed the role of the local resident giving directions and the other acted as an unfamiliar tourist asking for assistance.

In the genuine social presence condition (human to human interaction), each dyad comprised one learner assigned the role of local resident/direction-giver and one assigned the role of unfamiliar tourist. Roles were swapped after the first task, and task order was counterbalanced across dyads. Only the direction-giver's speech was treated as task performance data; tourist

turns, which were minimum consisting of mainly short confirmation or clarification checks like "I see" and "ok, thank you", were retained only as interactional context and were not included in CAF coding.

Task complexity served as a within-subjects factor across both groups, manipulated systematically through the number of spatial elements (landmarks) and degree of reasoning involved in a map direction-giving task. The order of the simple and complex tasks was counterbalanced across participants to offset practice and fatigue effects.

Participants

Participants were 18 undergraduate students (10 female, 8 male) recruited from a private university in Hong Kong. All were native Cantonese speakers with English as an L2. Their English oral proficiency fell within the B2 band of the Common European Framework of Reference (CEFR), evidenced by recent institutional IELTS scores ranging between 5.5 and 6.5 ($M = 6.0, SD = 0.31$). Participants had completed a mean of 13.4 years ($SD = 1.8$) of formal English instruction.

Six participants were assigned to the Virtual Social Presence (AI) condition ($n = 6$), and twelve participants were assigned to the Genuine Social Presence condition, forming six interactive dyads ($n = 12$). The 12 human-condition participants were paired into six dyads by the researcher randomly based on their scheduling availability for ecological validity (tourists and locals are strangers in reality). The AI-condition participants completed the tasks individually. Proficiency was balanced across both conditions, and all participants signed a consent form and, up completion, received an honorarium of HK$100.

Tasks and Instruments

Communicative Tasks

Two information-gap map tasks were developed following established TBLT guidelines (Bui, 2026d; Ellis, 2003; Tavakoli, 2009):

Simple Task (Low Complexity): Required the speaker to direct the interlocutor across a town map incorporating three sequential landmarks (library, café, bus stop). The sequential buildings require low reasoning demand.

Complex Task (High Complexity): Utilised the identical base map but increased the resource-directing cognitive load by incorporating six landmarks (library, café, bus stop, post office, shopping centre, and hospital) along with competing routing pathways. The competing routing requires high reasoning demand.

In the AI condition, the chatbot was programmed to respond strictly via voice synthesis using a standardised tourist scenario prompt, asking clarifying queries only when direct route milestones were ambiguous. In the human condition, peer tourists held a complementary map lacking the routes and were instructed to ask clarification questions freely until they reached the destination. The two map tasks are presented in Appendices 1 and 2.

Situational WTC Scale

Situational WTC was measured immediately post-task using a newly developed 12-item task-specific situational WTC questionnaire (Appendix 3), informed by the MacIntyre et al. (1998) conceptualisation of WTC and by prior WTC measurement (e.g., MacIntyre et al., 2001), rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The instrument measured willingness to initiate and sustain communication specifically during the experimental task

conditions (Cronbach's $\alpha = .84$). A sample item on Willingness to Sustain Communication is "I was willing to keep talking and elaborate with descriptive details when guiding the partner."

Post-Task Semi-Structured Interviews

Individual 15–20 minutes semi-structured interviews were conducted in English immediately after task performance to examine participants' subjective cognitive and affective experiences, perceptions of communicative pressure, and comparative views on interacting with human versus artificial agents.

Data Coding and Analysis

Audio recordings of task performance were transcribed verbatim and segmented into AS-units (Analysis of Speech units) following Foster et al. (2000). Standardised Complexity, Accuracy, and Fluency (CAF, Bui & Skehan, 2018) measures were computed:

Syntactic Complexity: Measured by the subordination ratio (subordinate clauses divided by total clauses) (Bui, 2014).

Linguistic Accuracy: Measured by the percentage of error-free clauses relative to total clauses (Skehan, et al., 2024).

Fluency: Operationalised through Speed (speech rate in syllables per minute), and breakdown fluency (mid-clause pause frequency per 100 words) (Tavakoli & Skehan, 2005).

CAF analyses were based on the direction-giver's AS-units only; tourist backchannels, clarification questions, and AI prompts were excluded. Thus, for CAF, n = 6 in the AI condition and n = 6 in the human condition (one direction-giver per dyad). Situational WTC was analysed at the individual level and therefore included all participants who completed the scale (AI n = 6; human n = 12), including tourists in the human dyads.

Given the exploratory nature and small sample size of this brief report (N=18), null-hypothesis significance testing (NHST) was deliberately omitted to avoid Type II errors associated with low statistical power. Instead, in line with methodological recommendations for small-sample SLA research (Plonsky & Oswald, 2014), we evaluate the strength and practical significance of the comparative trends using descriptive statistics paired with Hedges' g effect sizes, which correct for small and unequal sample sizes (n=6 vs. n=12 in WTC analyses). We interpret Hedges' g using Plonsky and Oswald's (2014) field-specific benchmarks for SLA: $|g| \geq 0.40$ as a small effect, $|g| \geq 0.70$ as a medium effect, and $|g| \geq 1.00$ as a large effect. In addition, qualitative interview transcripts underwent thematic analysis (Braun & Clarke, 2006) to identify recurrent emergent themes, achieving an inter-coder reliability of ICC = 84.72% across independent evaluators.

Results

Data and analysis

Willingness to Communicate: Virtual Safety vs. Genuine Interpersonal Pressure

Descriptive analyses revealed distinct variations in learners' WTC across interlocutor type and task complexity (Table 1).

Table 1. *Descriptive Statistics for Situational WTC Across Conditions*

Task Complexity	Mean WTC (Virtual/AI, n=6)	Mean WTC (Genuine/Human, n=12)	Standardized Effect Size (Hedges' *g*)	Interpretation
Simple	4.02 (.58)	3.58 (.68)	*.644*	Small-to-Medium
Complex	3.76 (.71)	3.26 (.74)	*.652*	Small-to-Medium
Combined	3.89 (.62)	3.42 (.71)	*.655*	Small-to-Medium

Note. Standard deviations in ().

Descriptive statistics tentatively suggest that learners exhibited higher communicative willingness when interacting with the virtual AI agent (M=3.89) than with a human peer (M=3.42), representing a solid, small-to-medium standardized effect size (g = 0.655). The data also seems to reflect an overarching decline in WTC from simple (M=3.80) to complex tasks (M=3.51).

Descriptively, the decline in WTC under high cognitive load was slightly larger in the presence of a human peer. However, this difference was small (g = 0.644 for simple tasks vs. g = 0.652 for complex tasks) and should be interpreted cautiously.

The interviews were consistent with these trends. In the genuine social presence condition, learners described concern about interpersonal evaluation and social face:

> ***"With a human partner, I'm always worried about wasting their time or sounding stupid. That pressure just isn't there with AI." (Participant 4, Human Condition)***

Conversely, participants in the AI condition described a distinct sense of communicative security afforded by virtual social presence:

> ***"I felt much more relaxed talking to the AI. I know it won't judge me if I make mistakes or speak slowly." (Participant 2, AI Condition)***

Some participants described that the interlocutor's machine nature de-escalates social judgment, encouraging uninhibited production, echoing "Online Disinhibition" (Suler, 2004), such as

> ***"I felt like there was a shield between me and the AI. Because it's not a real person, I didn't care about my accent or if my grammar was messy. I just spoke out whatever came to my mind. It felt like playing a solo computer game where you don't have to hide your weaknesses." (Participant 8, AI Condition)***

In addition, this participant added the following comments, showing how why he needed to restrict his output to save "face," which directly suppresses WTC.

> ***"With my classmate, there is this silent pressure to perform well. If I struggle to find the right words on the map, I feel like I'm wasting their time and they might judge my English level. To protect my face, I preferred to keep my answers as short as possible to avoid making mistakes." (Participant 14, Human Condition)***

Speaking Performance: Syntactic Complexity, Accuracy, and Fluency

Performance indices across Complexity, Accuracy, and Fluency displayed marked contrasts between conditions (Table 2).

Table 2. *L2 Performance Measures (CAF) as a Function of Interlocutor and Task Complexity* (*n*= 6 in each social presence group)

Metric	Task Complexity	Mean (SD) [Virtual/AI]	Mean (SD) [Genuine/Human]	Hedges' g	Interpretation
Syntactic Complexity *(Subordination Ratio)*	Simple	.24 (.08)	.30 (.10)	-.612	Small-to-Medium
	Complex	.28 (.11)	.38 (.13)	-.767	Medium
Linguistic Accuracy *(Error-Free Clauses %)*	Simple	68.4 (12.61)	71.20 (11.79)	-.212	Small-to-negligible
	Complex	62.1 (14.22)	65.80 (13.37)	-.247	Small-to-negligible
Speech Rate *(Syllables / Min)*	Simple	118.6 (18.42)	112.41 (16.77)	.325	Small-to-negligible
	Complex	108.2 (21.63)	106.82 (19.21)	.063	Negligible
Breakdown Fluency *(Mid-Clause Pauses / 100w)*	Simple	8.2 (3.40)	9.8 (3.80)	-.410	Small
	Complex	10.6 (4.23)	12.4 (4.62)	-.377	Small-to-negligible

Syntactic Complexity

Descriptive data tentatively point to an association between human interaction (genuine social presence) and higher structural subordination (M = 0.34) compared to AI interaction (M = 0.26) with virtual social presence. In the simple task, learners interacting with human peers produced higher subordination ratios (M = 0.30) than those interacting with the AI agent (M = 0.24), reflecting a small-to-medium standardized effect size (g = -0.612). Under the complex task, this gap widened (M = 0.38, for human vs. M = 0.28 for AI), yielding a medium effect size (g = -0.767). Across both conditions, the complex task elicited greater structural embedding (M = 0.33) than the simple task (M = 0.27), with cognitive task complexity appearing to amplify the syntactic stretching prompted by genuine human interlocutors.

Linguistic Accuracy

Accuracy dropped from an overall mean of 69.8% error-free clauses in the simple task to 63.9% in the complex task, reflecting the expected trade-off under increased cognitive processing load. Across interlocutor conditions, performance was largely comparable. In the simple task, learners in the human condition achieved slightly higher accuracy (M = 71.20%) than those in the AI condition (M = 68.40%), representing a small-to-negligible effect size (g = -0.212). A parallel pattern emerged in the complex task (M = 65.80% for human vs. M = 62.10% for AI; g = -0.247, small-to-negligible). These effect sizes indicate that interlocutor type had a minimal influence on syntactic accuracy relative to the cognitive demands of the task itself.

Temporal Fluency

Temporal fluency displayed contrasting patterns across speed and breakdown measures. Under simple task conditions, learners speaking to the AI interlocutor achieved a faster speech rate (M = 118.60 syllables/min) than those interacting with a human peer (M = 112.41 syllables/min), yielding a small-to-negligible effect favouring the AI condition (g = 0.325). Under complex conditions, this difference narrowed substantially, with speech rates converging (M = 108.20 for AI vs. M = 106.82 for human) and the effect size diminishing to negligible (g = 0.063).

Conversely, breakdown fluency, measured by mid-clause pauses per 100 words, indicated greater online formulation difficulty during genuine human interaction. In the simple task, learners speaking with human peers produced more mid-clause pauses (M = 9.80) than those with the AI agent (M = 8.20), yielding a small effect size (g = -0.410). In the complex task, mid-clause pausing increased overall (M = 11.50 per 100 words vs. M = 9.00 in simple tasks) and remained elevated in the human peer condition (M = 12.40 vs. M = 10.60; g = -0.377, small-to-negligible). This trend tentatively suggests that the interactive accountability of genuine human communication prompted greater real-time linguistic hesitation and formulation effort.

Learner Perceptions: The Trade-off Between Psychological Safety and Communicative Authenticity

Three themes emerged: 1) Affective climate: the AI was described as a low-evaluation, safe space lacking interpersonal warmth, whereas human peers evoked fear of negative evaluation and face-loss but also relational satisfaction when understanding was achieved. 2) Interactional nature: the AI was seen as patient but scripted and uniform, whereas peers produced clarification requests and confirmation checks that demanded immediate reformulation. 3) Perceived learning utility: the AI was seen as a stepping-stone for fluency practice and confidence, and peers as indispensable for pragmatic and interactive competence.

While the AI chatbot created an unhurried, non-threatening environment for speech production, it lacked true *subjectivity* (Jin & Wang, 2025). As one participant observed:

> ***"The AI never gets frustrated if I need to repeat something or take time to think. It just waits and responds... but the AI's responses felt a bit generic. It would say 'I see' or 'okay', but I wasn't sure if it really understood the directions. Real communication involves negotiation, checking understanding, and adjusting your language based on feedback." (Participant 11, Human Condition)***

Another participant highlighted the missing interpersonal satisfaction in virtual presence:

> ***"When I successfully explained the directions and my partner understood, there was a real sense of achievement. The AI always understands, so that satisfaction isn't really there." (Participant 6, Human Condition)***

Discussion and conclusion

Interpretation of the findings

Taken together, these empirical findings suggest how genuine and virtual social presence modulate the balance of affective, cognitive, and linguistic demands during L2 task performance.

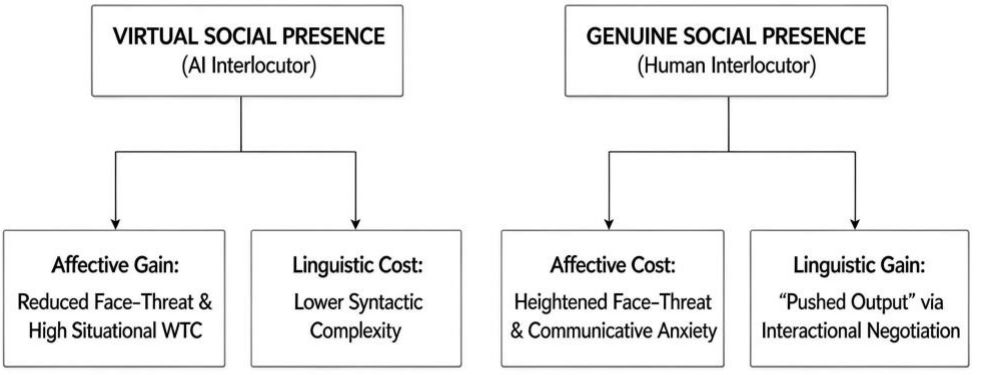


Figure 1. Effects of Social Presence (Virtual Vs Genuine) on WTC and L2 Performance

As shown in Figure 1, the descriptive results appear to reveal that virtual social presence may serve as an affective shield. By removing interpersonal appraisal, it lowers the communicative threshold, fostering higher WTC and, on the simpler task, faster speech. However, we tentatively suggest that this safety may come at a linguistic cost: without an interlocutor who actively demonstrates genuine non-understanding or negotiates meaning, the learner is not compelled to produce more sophisticated syntactic structures.

Conversely, genuine social presence may introduce communicative friction. The interpersonal stakes elevate learner anxiety and lower self-reported WTC, especially when task complexity compounds cognitive load. Yet, this very friction may be productive: the interactional contingencies of genuine human dialogue may push learners to modify their output, leading to more complex linguistic expression.

The higher WTC with the AI interlocutor is consistent with the online disinhibition effect (Suler, 2004). As Participant 8 noted, interacting with a simulated interlocutor creates a 'shield' that de-escalates social threat. This virtual social presence bypasses the traditional face-saving barriers that typically restrict East Asian L2 learners in face-to-face communication (Wen & Clément, 2003). When face-threat is neutralised, learners experience a state of psychological safety that allows them to initiate speech without fear of evaluation. Conversely, the social accountability inherent in genuine social presence introduces communicative friction; as Participant 14 observed, learners may voluntarily restrict their oral output ('keeping answers as short as possible') to protect their social face in front of human peers. All this echoes Bui (2026b), though face-threat and anxiety were not measured directly in this study, and therefore this explanation remains interpretive. WTC declined with task complexity in both conditions; the slightly larger decline with human peers was not reliable, so the proposal that AI buffers learners against complexity-induced declines needs testing with larger samples.

The greater subordination with human peers is compatible with pushed output (Swain, 1995) and the Interaction Hypothesis (Long, 2015): a peer who signals non-understanding may prompt speakers to unpack and elaborate their directions. However, negotiation of meaning was not coded, and the AI and peers followed different clarification protocols (see Communicative

Tasks subsection), so this mechanism remains speculative. More theoretically, the differential effects of social presence on performance highlight how Robinson's (2001, 2007) Cognition Hypothesis and Skehan's (1998, 2009) Limited Attentional Capacity Model function in a complementary, rather than merely competing, manner under different social conditions. Under genuine social presence (human peers), the interactive pressure and need for communicative precision appear to 'push' learners to prioritise structural elaboration, lending support to Robinson's Cognition Hypothesis where task demands prompt syntactic complexity. Conversely, under virtual social presence (AI), the affective shield and absence of evaluation anxiety allow learners to allocate their limited attentional resources differently; freed from face-saving concerns, they can prioritise temporal fluency (faster speech rate) on simpler tasks, supporting Skehan's trade-off predictions. Social presence thus serves as an affective modulator that dictates how learners navigate the trade-offs between cognitive complexity and attentional resource allocation. While the overall effects of task complexity remained mixed, with higher subordination supporting the Cognition Hypothesis, but lower accuracy, slower speech, and more mid-clause pausing fitting Skehan's trade-off predictions, our findings tentatively suggest that virtual presence buffers the cognitive load of simple tasks to support fluency, whereas genuine presence forces syntactic stretching even when cognitive processing demands are high.

Interviewees portrayed the AI as patient and non-threatening but generic, offering no real sign of understanding, whereas human interaction brought accountability and anxiety as well as a sense of achievement once understanding was reached. Because these comparative views came from participants who had experienced only one interlocutor type, they reflect expectations as much as experience.

Pedagogical implications

Rather than viewing AI chatbots as replacements for human interaction, teachers should incorporate them within a principled, sequenced TBLT framework, as also argued in Bui (2027):

AI as a Pre-Task Interactive Rehearsal Space:

AI chatbots can be deployed during the pre-task phase of a task cycle. Anxious or reticent learners can complete simple, fluency-oriented versions of a task with a virtual interlocutor. This initial practice may activate relevant vocabulary and builds communicative confidence with less social risk.

Task Progression from Virtual to Genuine Presence:

Pedagogical sequences should transition learners from virtual to genuine social presence. Once learners establish baseline fluency and familiarity with task demands via AI interaction, they can perform more complex task variants with human peers, where interactional modifications may foster syntactic restructuring.

Fostering Prompt and Interaction Literacy:

To maximise the linguistic utility of conversational AI, learners require explicit instruction in interaction management (Tour & Zadorozhnyy, 2025). Instructing learners to ask conversational AI to challenge their assumptions, request clarifications, or demand alternative routes may help simulate aspects of the pushed output normally provided by human peers.

Limitations and Conclusion

In sum, our exploratory findings tentatively suggest that interlocutor type and task complexity shaped learners' WTC and oral performance in different ways. Specifically, the data tentatively point to an association between the AI interlocutor and higher WTC, whereas the human peer was associated with greater syntactic complexity; higher task complexity lowered WTC, accuracy and speech rate and increased mid-clause pausing in both conditions, while raising subordination. These exploratory findings suggest that virtual and genuine social presence offer complementary affordances for task-based speaking.

This exploratory study is limited by its modest sample size ($N = 18$) and reliance on a single communicative task type and AI tool (*Doubao*). It is further limited by unequal group sizes and symmetrical questioning and clarification mechanisms between the AI and peer interlocutors, which likely influenced oral production profiles. Future SLA research should track these variables with a bigger sample size and investigate how developing AI interactional competencies might prompt conversational agents to offer more authentic negotiation of meaning longitudinally. Ultimately, conversational AI may offer a low-anxiety context for building communicative willingness, whereas human interaction may be more conducive to syntactic elaboration. Larger, preregistered replications are needed to test these possibilities.

References

Aubrey, S., & Yashima, T. (2023). Willingness to communicate in TBLT. In C. Lambert, S. Aubrey, G. Bui (Eds.), *The role of the learner in TBLT: Theory and research methods* (pp.41–57). Routledge.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bui, G. (2026a). Task-based language teaching. In H. Nesi & P. Milin (Eds.), *International encyclopedia of language and linguistics (3rd ed.)* (pp.25-30, Vol. 9). Elsevier. https://doi.org/10.1016/B978-0-323-95504-1.00434-8

Bui, G. (2026b). From novelty to strategy: The trajectory of learner perceptions in human and AI-assisted L2 speaking tasks. *International Journal of TESOL Studies*, *8*(1), 108-127. https://doi.org/10.58304/ijts.260301

Bui, G. (2026c). Effects of task interactivity and willingness to communicate on speech fluency in second language learning tasks. *RELC Journal*. https://doi.org/10.1177/00336882261422018

Bui, G. (2026d). Task-based language teaching: Principles and common misunderstandings. *Conceptual Foundations in Language Research*, *1*(1), 260101. https://doi.org/10.65553/CFLR.260101

Bui, G. (2027). Developing AI literacy in task-based language teaching and learning. In Teng, M. F., Yeo, M., & Wu, J. G. (Eds.), *Fostering AI literacy in the digital applied linguistics era*. Routledge.

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). John Benjamins. https://doi.org/10.1075/tblt.5.03gav

Bui, G., & Skehan, P. (2018). Complexity, fluency and accuracy. In J. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp.1–8). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0046

Cao, Y. (2011). Investigating situational willingness to communicate within second language classrooms from an ecological perspective. *System*, 39(4), 468–479. https://doi.org/10.1016/j.system.2011.10.016

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL Learners' Speaking Skills and Willingness to Communicate via Artificial Intelligence-Mediated Interactions. *System*, *121*, Article 103254.

https://doi.org/10.1016/j.system.2024.103254

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. https://doi.org/10.1093/applin/21.3.354

Garrison, D.R., Anderson, T., & Archer, T. (2000). 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

Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education*, *11*(3), 8–26. https://doi.org/10.1080/08923649709526970

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*, 60(5), 834–851. https://doi.org/10.1093/jopedu/qhaf087

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

MacIntyre, P. D., Baker, S. C., Clément, R., & Conrod, S. (2001). Willingness to communicate, social support, and language-learning orientations of immersion students. *Studies in Second Language Acquisition*, 23(3), 369–388. https://doi.org/10.1017/S0272263101003035

MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

Peng, J. E. (2025). *Willingness to communicate in a second language*. Cambridge University Press. https://doi.org/10.1017/9781009417884

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Qiu, X., Lo, Y. Y., Ge, H., & Bui, G. (2026). Second language learner engagement in computer-mediated interactive oral tasks. *Studies in Second Language Learning and Teaching, 16*(1), 13-41. https://doi.org/10.14746/ssllt.41317

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

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–27). Multilingual Matters.

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.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Skehan, P., Bui, G., Wang, Z., & Shum, S. (2024). Re-examining accuracy measures in second language task-based spoken performance. *Research Methods in Applied Linguistics, 3*(1), 100098. https://doi.org/10.1016/j.rmal.2024.100098

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.

Tavakoli, P. (2009). Investigating task difficulty: Learners' and teachers' perceptions. *International Journal of Applied Linguistics*, 19(1), 1–25. https://doi.org/10.1111/j.1473-4192.2009.00216.x

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), Planning and task-based performance in a second language (pp. 239-273). John Benjamins. https://doi.org/10.1075/lllt.11.15tav

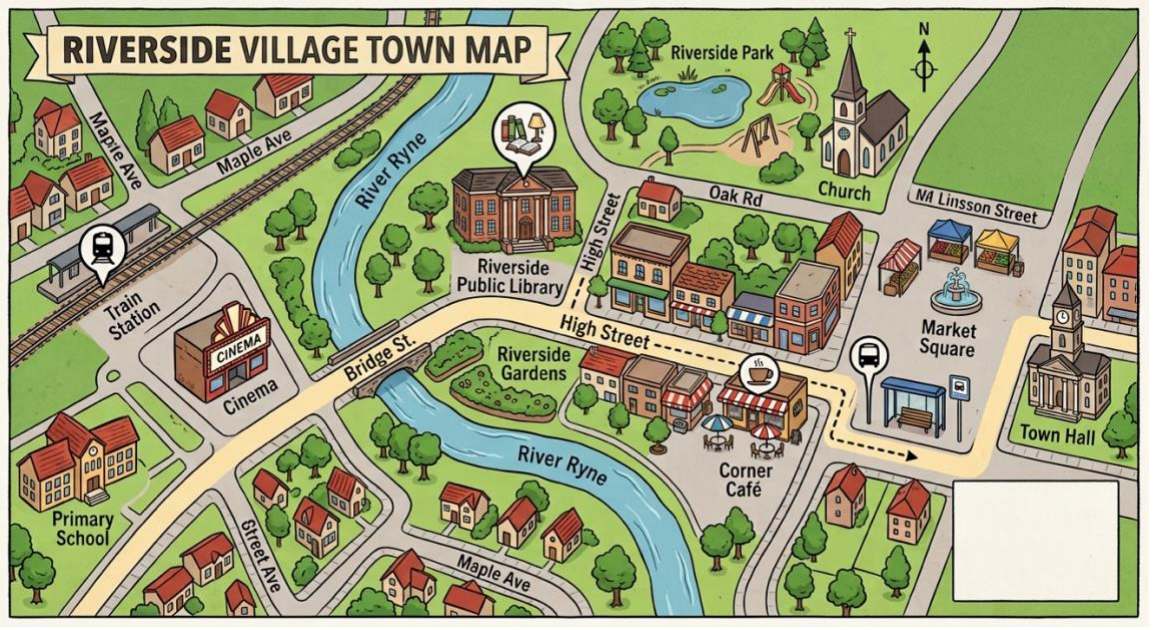
Tour, E., & Zadorozhnyy, A. (2025). Conceptualizing and operationalizing prompt literacy for English language learners. *Journal of Adolescent & Adult Literacy*, 69(3), e70020. https://doi.org/10.1002/jaal.70020

Wang, Y. (2026). Exploring student engagement with multimodal generative AI in task-based Chinese language learning. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2026.2645428

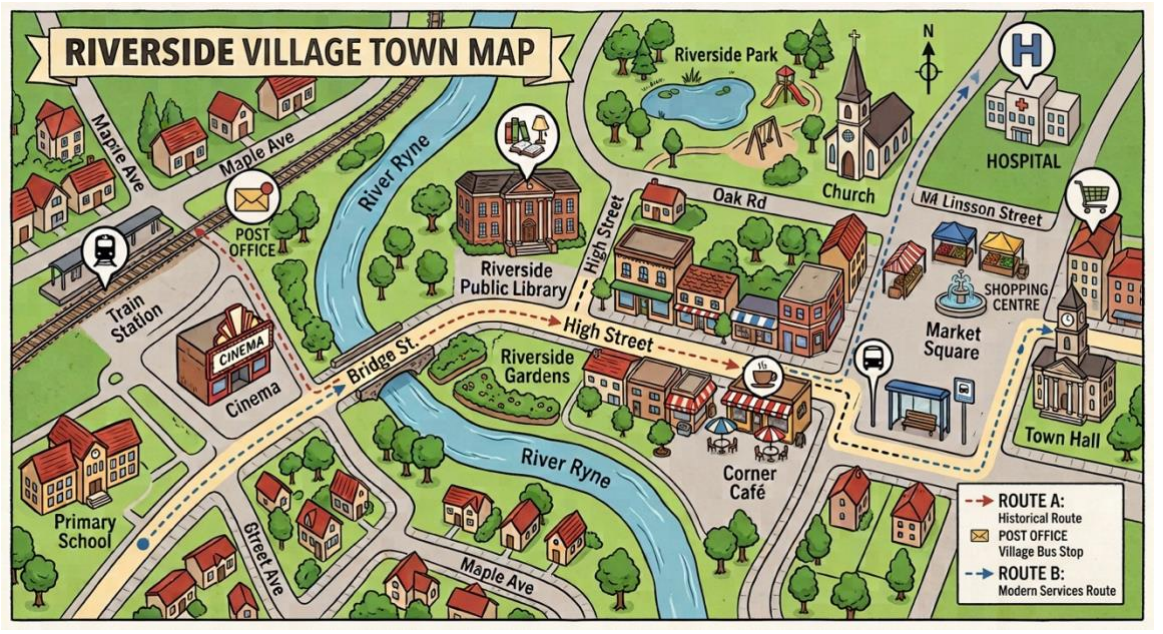
Wen, W. P., & Clément, R. (2003). A Chinese conceptualisation of willingness to communicate in ESL. *Language, Culture and Curriculum*, 16(1), 18–38. https://doi.org/10.1080/07908310308666654

Appendices: Experimental Task Materials

Appendix 1: Simple Task Map (Low Complexity: 3 Landmark + Low Reasoning Demand)



Appendix 2: Complex Task Map (High Complexity: 6 Landmarks + High Reasoning Demand)



Appendix 3: Situational Willingness to Communicate (WTC) Questionnaire

Instructions: Please indicate the extent to which you agree or disagree with each statement regarding your speaking experience during this specific direction-giving task. *(1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)*

No.	Dimension	Item
1	**Initiation**	I was willing to start speaking and give directions as soon as the task commenced.
2	**Initiation**	I felt ready to take the conversational lead without waiting for prompting.
3	**Initiation**	I was eager to introduce the key street names and landmarks immediately.
4	**Elaboration**	I was willing to provide additional descriptive details to make the route clearer.
5	**Elaboration**	I wanted to keep talking even when navigating the more complex sections of the map.
6	**Elaboration**	I tried to use complete and descriptive sentences rather than minimal, short answers.
7	**Interactive Repair**	I was willing to rephrase or repeat my directions whenever my partner seemed uncertain.
8	**Interactive Repair**	I felt comfortable answering clarification questions asked by my partner.
9	**Interactive Repair**	I was eager to actively check whether my partner had understood each step of the route.
10	**Affective Readiness**	I felt relaxed and comfortable communicating in English with my partner during this task.
11	**Affective Readiness**	I was not concerned about making grammatical errors or hesitating while speaking.
12	**General Disposition**	I would gladly volunteer to perform another similar direction-giving speaking task with this partner.

Author bio

Gavin Bui is Associate Professor in the Department of English at the Hang Seng University of Hong Kong. His main research interest is task-based language teaching and learning. His recent publications appeared in *System* (2026), *RELC* Journal (2026), *Language Teaching Research* (2025), and *SSLA* (2024).
Email: gavinbui@hsu.edu.hk
ORCID: https://orcid.org/0000-0002-1567-9074