



Localizing Integrated Performance Assessment in Chinese Tertiary EFL Contexts: Effects of the LT-IPA Model on English Speaking Proficiency and Willingness to Speak

Lingyi Liu^a, Punchalee Wasanasomsithi^{b,*}

^alingyiliuchristine@outlook.com, Graduate School, Chulalongkorn University, Thailand

^bpunchalee.w@chula.ac.th, Language Institute, Chulalongkorn University, Thailand

*Corresponding author, punchalee.w@chula.ac.th

APA Citation: Liu, L., & Wasanasomsithi, P. (2026) Localizing integrated performance assessment in Chinese tertiary EFL contexts: Effects of the LT-IPA model on English speaking proficiency and willingness to speak. <i>LEARN Journal: Language Education and Acquisition Research Network</i>, 19(2), 629-664. https://doi.org/10.70730/JROV8402	
Received 02/12/2025 Received in revised form 17/04/2026 Accepted 21/04/2026	ABSTRACT China represents one of the world’s largest English as a foreign language (EFL) contexts. Despite decades of curriculum reforms, tertiary English teaching and assessment in China remain dominated by traditional, discrete-point written tests, resulting in limited spoken language use and consequently weak speaking proficiency and low willingness to speak (WTS) among undergraduates. This mixed-methods, quasi-experimental study addressed these challenges by incorporating localized tasks into the ACTFL’s Integrated Performance Assessment (IPA) framework, referred to as the Localized Task-IPA (LT-IPA) Model. Conducted at a private college in southwest China, the study aimed to (1) investigate the effects of the LT-IPA Model on Chinese undergraduates’ English speaking proficiency, and (2) examine their levels of WTS. Data were collected from 29 first-year undergraduates

	through speaking pre- and post-tests, a WTS questionnaire, and semi-structured interviews. Findings revealed significant improvement in participants' English speaking proficiency following 16 weeks of LT-IPA implementation, accompanied by high levels of WTS. These findings provided empirical evidence for the feasibility of localizing the IPA framework in tertiary EFL contexts in China, thereby extending its theoretical applicability beyond Western settings. The study also offered insights into how pedagogical factors can shape students' WTS in classrooms. Keywords: integrated performance assessment, localized tasks, English speaking proficiency, willingness to speak, Chinese undergraduates
--	---

Introduction

English has become a key medium of global communication, playing a crucial role in diplomacy, academia, and international business. Among the four language skills, speaking is particularly important, as it enables real-time interaction and effective collaboration across cultures. In China, English education has long been prioritized as part of the nation's strategy for modernization and internationalization. Since the 1980s, the Chinese Ministry of Education (MOE) has introduced successive curriculum reforms promoting communicative language teaching (CLT) and task-based pedagogy, aiming to cultivate learners' communicative competence (Fang, 2010; MOE, 2020). Nevertheless, despite decades of reform, the improvement of Chinese undergraduates' English speaking proficiency has remained limited (Gan, 2013; Wei & Su, 2012).

Two persistent problems characterize this gap between policy and practice: (1) university students' speaking proficiency lags significantly behind their reading and writing skills, and (2) students demonstrate a low willingness to speak (WTS) English in class. While national policies advocate communicative competence, classroom instruction continues to be dominated by teacher-centered lecturing and discrete-point testing, emphasizing grammar, vocabulary, and written production over oral communication (Gan, 2013). Assessment systems such as the College English Test (CET) further reinforce this imbalance, as their written components determine most academic outcomes, whereas the speaking test remains optional and marginal. Consequently, students prioritize exam-oriented learning and rarely practice spontaneous speaking, leading to the phenomenon often described as "deaf-and-dumb English" (Fang, 2010).

This deficiency is also reflected in international proficiency tests. Chinese candidates consistently underperform in speaking compared to other skills: the average IELTS speaking score in 2022 was 5.6, lower than writing (5.8), reading (6.4), and listening (6.1) (IELTS.org, 2022). Similar trends appear in TOEFL iBT results (ETS.org, 2023). Beyond scores, surveys indicate that only a minority of Chinese EFL learners feel confident engaging in extended conversations (Wei & Su, 2012). The lack of authentic communicative opportunities, coupled with a fear of making mistakes and a reliance on teacher authority (Amoah & Yeboah, 2021; Gan, 2013), further discourages classroom participation. Thus, the exam-oriented learning system, traditional pedagogy, and limited language environment jointly constrain Chinese undergraduates' English speaking proficiency and willingness to speak English.

The Localized Task-Integrated Performance Assessment (LT-IPA) Model was proposed in this study to address these challenges. The LT-IPA Model builds on two complementary pillars. The first pillar, localized tasks, derives from TBLT principles that prioritize meaningful communication through goal-oriented tasks (Ellis et al., 2020). Unlike standardized textbook tasks, localized tasks are designed around learners' linguistic, contextual, and cultural realities (Adair-Hauck et al., 2013; Ellis et al., 2020; Gilmore, 2007; Kramsch & Hua, 2016). They promote engagement and authenticity by connecting classroom activities with students' everyday communication contexts, making language learning more relevant and meaningful. The second pillar, IPA, provides a holistic evaluation framework consisting of three interrelated communicative modes: interpretive, interpersonal, and presentational (Adair-Hauck et al., 2013). These modes mirror authentic language use, requiring learners to comprehend input, engage in interactive exchanges, and produce output for real audiences. The integration of localized tasks into the IPA framework ensures that instruction and assessment reinforce each other, enabling students to practice and demonstrate language skills in realistic communicative sequences.

This integration is particularly suited to China's EFL context for three reasons. First, it aligns with national educational policies that emphasize communicative competence, student-centered instruction, and task-based pedagogy (MOE, 2020). The LT-IPA Model operationalizes these principles by engaging students in active meaning-making and authentic communication. Second, it addresses the need for contextual adaptation. As IPA originated in Western educational systems, localizing its implementation ensures linguistic, contextual, and cultural appropriateness for Chinese learners, enhancing familiarity and reducing anxiety. Third, the model's feasibility and flexibility make it applicable in typical Chinese classrooms. Teachers can adapt IPA task

designs, select locally relevant topics, and incorporate familiar materials without disrupting institutional requirements or assessment structures.

Empirical evidence from other EFL contexts supports the model's potential. Studies in Korea and Thailand have demonstrated that localized task-based instruction and integrated performance assessment significantly improve learners' speaking ability and engagement (Kim et al., 2017; Plengkham & Wasanasomsithi, 2023). Given the shared educational challenges, such as teacher-centered instruction, exam orientation, and limited oral practice, similar pedagogical outcomes are expected in China.

In sum, despite strong policy advocacy for communicative English teaching, Chinese university classrooms continue to produce learners who can read and write but struggle to speak. This study argues that the LT-IPA Model offers a practical and theoretically grounded solution by bridging the gap between policy intentions and classroom realities. Through integrating localized, context-sensitive tasks into a performance-based assessment framework, the model aims to enhance students' speaking proficiency and foster their willingness to speak English in class. Accordingly, this study addressed two research questions:

1. What are the effects of the LT-IPA Model on Chinese undergraduates' English speaking proficiency?
2. What are the levels of Chinese undergraduates' willingness to speak (WTS) English in class after implementing the LT-IPA Model?

Literature Review

Integrated Performance Assessment (IPA)

Developed by the American Council on the Teaching of Foreign Languages (ACTFL), the Integrated Performance Assessment (IPA) represents a shift from traditional discrete-point testing to performance-based, communicative assessment (Adair-Hauck et al., 2013). Grounded in the National Standards for Foreign Language Learning and the ACTFL Proficiency Guidelines, IPA evaluates learners' ability to communicate meaningfully across the interpretive, interpersonal, and presentational modes within a unified thematic framework. Anchored in backward design (Wiggins & McTighe, 2005), it ensures alignment among learning objectives, assessment, and instruction. Its emphasis on authenticity engages learners in purposeful, real-world communicative tasks that cultivate linguistic proficiency, cultural understanding, and learner autonomy (Adair-Hauck et al., 2006).

Each of the three IPA modes serves a distinct yet interconnected communicative function. The interpretive mode develops comprehension

through engagement with authentic written, oral, or visual texts that demand both literal and inferential comprehension. The interpersonal mode focuses on spontaneous, two-way interaction that requires negotiation of meaning. Finally, the presentational mode requires learners to organize and deliver information to an audience, thus promoting purposeful communication and higher-order thinking (Adair-Hauck et al., 2013). Together, these modes form a developmental sequence from comprehension to interaction to production, reflecting authentic language use.

Central to IPA implementation are analytic rubrics that define performance expectations along a continuum ranging from “not meeting expectations” to “meeting expectations” (with distinctions between minimal and strong performance), and “exceeding expectations.” Based on the IPA manual (Adair-Hauck et al., 2013), five dimensions are assessed for interpersonal speaking: language function (the ability to achieve communicative purposes such as requesting, explaining, or persuading), text type (the length and complexity of discourse, ranging from words and phrases to extended, connected speech), communication strategies (the use of negotiation strategies such as clarification or breakdown repairing), comprehensibility (the degree to which speech is understood by the listener), and language control (vocabulary range, grammatical accuracy, and fluency). For presentational speaking, impact replaces communication strategies and refers to the overall effectiveness of the message in engaging and influencing the audience (Adair-Hauck et al., 2013).

A defining feature of IPA is its recursive feedback loop, which integrates assessment, learning, and learner reflection into a continuous cycle. This loop consists of five stages: (1) clarifying expectations through rubrics, (2) modeling target performances, (3) guided practice and preparation, (4) task performance, and (5) feedback and reflection (Adair-Hauck et al., 2013). Feedback in IPA is not limited to teacher evaluation but involves co-constructive processes in which learners actively participate. After task performance, learners engage in peer feedback, self-assessment, and guided reflection using rubric criteria, while teachers provide descriptive feedback to further pinpoint areas for improvement and offer actionable strategies. This dialogic process allows learners to interpret feedback, reflect on their own performance, and set goals for subsequent performances. The outcomes of one cycle then inform the next, creating an iterative process that continuously supports learning development (Adair-Hauck & Troyan, 2013).

Studies have shown that systematic implementation of IPA leads to measurable improvements in speaking performance and is perceived positively by both teachers and students (Altstaedter & Krosch, 2018; Martel, 2018; Plengkham & Wasanasomsithi, 2023; Zapata, 2016).

Localized Tasks

Originating from task-based language teaching (TBLT), localized tasks represent an adaptive evolution of Western pedagogical frameworks to better suit the sociocultural and institutional realities of East Asian contexts. TBLT, rooted in the communicative language teaching (CLT) movement, foregrounds meaning-making and authentic communication over structural drills (Ellis et al., 2020). However, direct transplantation of Western-oriented TBLT has often proven problematic in non-Western settings, where educational traditions, assessment systems, and learner expectations differ substantially (Littlewood, 2011). In response, localization has emerged as a pedagogical imperative, one that “thinks globally and acts locally” (Xu, 2018), to ensure tasks remain linguistically, contextually, and culturally relevant. Localized tasks thus bridge global communicative frameworks with local realities, providing learners with opportunities to use English meaningfully within familiar and resonant contexts (Liu & Guo, 2020).

The defining features of localized tasks can be conceptualized along three interrelated dimensions: linguistic relevance, contextual authenticity, and cultural appropriateness. Linguistic relevance ensures that task content aligns with learners’ language levels and communicative goals, following the CALL-IT principle (Context, Age, Linguistic Level, Importance of Task) in the IPA framework (Adair-Hauck et al., 2013). Contextual authenticity requires that tasks mirror real-life communicative situations and employ unedited, authentic language use rather than simplified textbook discourse (Gilmore, 2007; Shrum & Glisan, 2010). Cultural appropriateness emphasizes integrating learners’ cultural identities and local communication norms into instructional design, thereby transforming English from a purely global medium into a vehicle for expressing local thought and identity (Kramsch & Hua, 2016). Together, these dimensions reflect a shift from global adoption to local adaptation (Xu, 2018), reinforcing that effective pedagogy must emerge from within its own linguistic, contextual, and cultural realities.

Empirical research in East Asian EFL contexts emphasizes the need for localization of Western instructional approaches. Studies reveal that large class sizes, exam-oriented instruction, and outdated textbooks hinder the implementation of TBLT (Liu & Guo, 2020; Liu et al., 2021; Ren & Han, 2016). Teachers frequently modify form-focused materials to create task-supported learning environments, demonstrating a pragmatic transition from adoption to adaptation (Butler, 2011). Learners’ attitudes toward task-based learning also evolve positively when tasks are tailored to their linguistic needs and cultural realities (Kim, 2019; Kim et al., 2017). Collectively, these findings support that localized TBLT enhances learner engagement and learning outcomes when grounded in authentic and contextually resonant practice.

Localized tasks are conceptually and pedagogically compatible with the IPA framework, which similarly emphasizes authenticity, relevance, and contextualization (Adair-Hauck et al., 2013; Martel & Bailey, 2016). Therefore, incorporating localized tasks into the IPA framework provides a viable pathway toward communicative, contextually responsive English education in East Asian higher education contexts.

Speaking Proficiency

Speaking involves conveying messages through oral discourse, using spoken language in real-time communication. As a productive skill, it facilitates individuals' interaction with others in a variety of contexts (Hughes & Reed, 2016). According to the ACTFL Proficiency Guidelines (Malone et al., 2024), speaking proficiency refers to an individual's ability to effectively achieve communicative goals in spoken language across diverse situations. It is evaluated through four core criteria summarized by the acronym FACT (Functions, Accuracy, Context and Content, and Text Type). By contrast, speaking performance concerns spoken language use that has been explicitly taught and practiced in familiar contexts (Sandrock & Swender, 2012). While both proficiency and performance employ the same evaluative dimensions (FACT), the former emphasizes consistent, spontaneous communication across unpredictable contexts, whereas the latter reflects practiced language use within instructional boundaries. In this sense, performance can serve as an indicator of proficiency development, although the two constructs are not equivalent (Malone et al., 2024).

ACTFL further conceptualizes speaking proficiency as encompassing two complementary modes of communication: interpersonal and presentational speaking (Malone et al., 2024). Interpersonal speaking involves spontaneous, two-way interaction in which meaning is negotiated in real time. In contrast, presentational speaking refers to one-way, organized discourse directed at an audience without immediate feedback. Aligned with this conceptualization, IPA provides a task-based performance assessment framework through which these two modes of communication can be elicited and observed in classroom contexts, thereby linking classroom-based performance assessment with broader proficiency-oriented evaluation (Adair-Hauck et al., 2013).

Willingness to Speak (WTS) English

The construct of Willingness to Speak (WTS) English originates from the broader notion of Willingness to Communicate (WTC), first proposed by McCroskey and Baer (1985) in the field of first language (L1) communication

and later extended to second language (L2) contexts (MacIntyre et al., 1998). While L1 WTC was conceptualized as a relatively stable personality trait reflecting individuals' predisposition to engage in communication, L2 WTC incorporated both enduring and situational influences, such as motivation, self-confidence, intergroup attitudes, and contextual variables. As MacIntyre et al. (1998) have argued, the ultimate pedagogical goal of L2 education should be to cultivate learners' willingness to seek and engage in communication opportunities. Building upon this framework, Riasati (2018) has refined the concept into Willingness to Speak (WTS), a construct specifically addressing learners' readiness to participate in spoken English communication within EFL classroom settings. This refinement narrows the scope from general L2 communication to the more contextually bound EFL classroom environment, where speaking opportunities are constrained and shaped by instructional and affective factors.

Theoretically, WTS embodies a dynamic interaction between personal and situational factors. While enduring influences such as personality traits or general self-confidence provide a stable basis for communication (MacIntyre et al., 1998), contextual or situational factors, such as topic familiarity, interlocutor effect, and classroom atmosphere, play a more immediate and malleable role (Riasati, 2018; Riasati & Rahimi, 2018). Because of their transient and manipulable nature, situational factors have become a central focus of recent WTS research.

Empirical findings indicate that learners' WTS fluctuates across time and situations, suggesting that instructional design and classroom climate significantly influence their speaking behavior (Cao & Philp, 2006; Riasati, 2012). Hence, creating supportive language learning environments is essential for fostering WTS in EFL classrooms. Similar evidence from Asian EFL contexts suggests that learners' reticence is primarily situation-specific, often triggered by misalignments between pedagogical practices and learners' linguistic competence (Cheng, 2000). These findings highlight the situational nature of WTS and reinforce the teacher's central role in shaping conducive classroom environments. By strategically promoting contextual conditions that encourage communication and mitigating those that hinder it, teachers can effectively transform learners' situational unwillingness into active participation (Riasati, 2018). Empirical studies also demonstrate a positive relationship between WTS and learners' speaking proficiency. Riasati (2018), for example, has reported a correlation ($r = .64$) between learners' WTS and their IELTS speaking scores, emphasizing that greater willingness leads to more active participation and improved speaking performance.

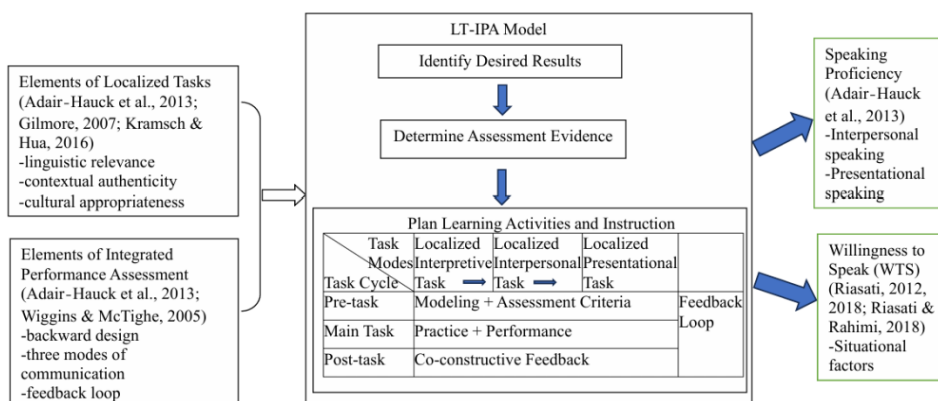
The measurement of WTS has evolved from self-report questionnaires (Cao & Philp, 2006; McCroskey, 1992) to classroom-specific scales developed by Riasati and Rahimi (2018), which better capture learners'

willingness to engage in spoken activities. Non-verbal indicators, such as hand-raising during participation, also serve as observable manifestations of WTS (MacIntyre et al., 1998). These methodological advancements provide a more comprehensive understanding of learners' readiness to speak, reinforcing the view that WTS is not a fixed trait but a context-sensitive construct, shaped by pedagogical intervention.

The conceptual framework of this study incorporates localized tasks into the IPA framework. It is grounded in four interrelated components: backward design principles, the three IPA task modes, key features of localized tasks, and the IPA feedback loop. The LT-IPA Model was developed to enhance English speaking proficiency and foster WTS among Chinese undergraduates.

Figure 1

The Conceptual Framework of the LT-IPA Model



Methodology

This study employed a mixed-methods quasi-experimental design, combining both quantitative and qualitative data to ensure methodological triangulation, as Creswell (2024) has pointed out that “bringing quantitative and qualitative research together addressed offsetting strengths and weaknesses of the two methods” (p. 204). Quantitative data were collected through a single-group pretest-posttest design and a WTS questionnaire, while qualitative insights were obtained from semi-structured interviews.

Participants

This study was conducted at a private undergraduate college in Kunming, Yunnan Province, China. Participants were 29 first-year undergraduates from one intact class enrolled in a compulsory English Listening & Speaking (Level 2) course over a 16-week semester. An intact group of students was recruited due to the natural constraints of classroom-based research, where random assignment to groups was not feasible. The participants, at an intermediate level of English proficiency, consisted of five males and 24 females. They had previously completed two spoken English courses as part of their college curriculum, and most had begun formal English learning at Grade 4 or later in primary school.

Instruments

Four instruments were employed: the IPA speaking pretest and posttest, the corresponding IPA rubrics, a WTS questionnaire, and a semi-structured interview protocol.

IPA Speaking Pretest and Posttest

The IPA speaking pretest and posttest were developed to assess participants' speaking proficiency before and after the intervention. Adapted from the ACTFL's IPA speaking test format by the researchers and validated by a panel of experts in English language instruction, the tests were aligned with topics from the English Listening & Speaking (Level 2) curriculum and calibrated to the intermediate proficiency level. The pretest and posttest were administered in Weeks 2 and 15, respectively, using parallel forms with different topics to minimize practice effects. Each test included interpretive, interpersonal, and presentational tasks incorporating localized features of linguistic relevance, contextual authenticity, and cultural appropriateness. All performances were recorded and independently rated by two trained raters following standardized procedures to ensure fairness and reliability.

IPA Rubrics

The IPA rubrics for intermediate learners, adopted from Adair-Hauck et al. (2013), were used to assess participants' interpersonal and presentational speaking performance. An analytic scoring approach evaluated five criteria on a four-point scale, yielding a total score of 20 for each task. The following two tables present the scoring rubrics.

Table 1*IPA Rubrics for Interpersonal Speaking Tasks*

Criteria	Exceeds Expectations (4 points)	Meets Expectations- Strong (3 points)	Meets Expectations- Minimal (2 points)	Does Not Meet Expectations (1 point)
Language Function				
Text Type				
Communication Strategies				
Comprehensibility				
Language Control				

Table 2*IPA Rubrics for Presentational Speaking Tasks*

Criteria	Exceeds Expectations (4 points)	Meets Expectations- Strong (3 points)	Meets Expectations- Minimal (2 points)	Does Not Meet Expectations (1 point)
Language Function				
Text Type				
Impact				
Comprehensibility				
Language Control				

IPA Speaking Test Raters

Two raters independently evaluated all IPA speaking test recordings to ensure scoring consistency. Before formal rating, a calibration session was conducted in which the raters reviewed the analytic rubric descriptors and scored sample recordings to align their interpretations. Following calibration, all recordings were scored independently by both raters. Inter-rater reliability was assessed using the intraclass correlation coefficient (ICC). The final score for each participant was calculated as the average of the two ratings to enhance reliability and reduce individual bias. When necessary, additional discussions were conducted to resolve discrepancies and ensure consistent application of the rubric.

WTS Questionnaire

Adapted from Riasati's (2018) Willingness to Speak (WTS) model and validated by a panel of experts in English language instruction, a WTS questionnaire was administered in Week 16 to examine participants' levels of

WTS in class and the underlying factors influencing their WTS. The instrument encompassed seven dimensions, namely topic of discussion, localized tasks, task type, interlocutor effect, IPA task modes, IPA backward design, and IPA feedback loop. The questionnaire consisted of 20 items rated on a five-point Likert scale (from 1 = strongly disagree to 5 = strongly agree). Because mean scores were reported to two decimal places, interpretation was based on score ranges rather than discrete integers. To determine the interpretation intervals, the scale range ($5 - 1 = 4$) was divided by the number of response categories (5), resulting in an interval width of 0.80. Based on this calculation, the following interpretive ranges were established:

- 1.00–1.80 = very low WTS
- 1.81–2.60 = low WTS
- 2.61–3.40 = moderate WTS
- 3.41–4.20 = high WTS
- 4.21–5.00 = very high WTS

Semi-Structured Interview Protocol

To complement the quantitative data, semi-structured interviews were conducted in Week 16 with six participants purposively selected to represent high-, mid-, and low-performing groups based on their speaking scores. The interview protocol was designed to explore how participants perceived the effects of the LT-IPA Model on their speaking performance, as well as their perceptions related to WTS, thereby providing richer qualitative insights into their learning experiences with the LT-IPA Model.

Data Collection

The intervention lasted 16 weeks during the second semester of the 2024 academic year, with 90 minutes of instruction per week. In Week 1, participants were introduced to the LT-IPA Model and learning procedures, followed by a demonstration of an IPA speaking test. In Week 2, an IPA speaking pretest was administered and recorded via Tencent Meeting to establish baseline performance.

From Weeks 3 to 14, the LT-IPA Model was implemented through thematic units organized in two-week cycles. In the first week, participants engaged in interpretive and interpersonal tasks. During this stage, they developed comprehension and interactional skills, supported by peer feedback, self-assessment, and teacher guidance. In the second week, participants completed a presentational task, synthesizing and delivering information to a specific audience. This sequence ensured that presentational speaking was supported by prior input and interaction. Feedback was again

provided after the task, emphasizing co-constructive dialogue between teacher and students to support reflection and improvement.

Localized tasks were adapted from or developed using the prescribed textbook. For example, textbook-based content on food was adapted into a campus dining context, where participants interpreted canteen menus they collected, discussed their preferences, and prepared recommendations for a specific audience. Similarly, textbook themes related to cities and traveling were extended into localized tasks in which participants watched videos introducing Chinese cities, engaged in discussions, and presented recommendations for traveling or living.

In Week 15, a posttest was administered and recorded using the same procedures as the pretest. In Week 16, participants completed the WTS questionnaire, followed by semi-structured interviews with six selected participants.

Data Analysis

All quantitative data were analyzed using SPSS after preliminary organization in Microsoft Excel. Pretest and posttest speaking scores were first entered into Excel and subsequently imported into SPSS for statistical analysis. Inter-rater reliability and data normality were examined before hypothesis testing. Paired-samples t-tests were then conducted to compare pretest and posttest results, with both descriptive and inferential statistics reported to assess the significance and magnitude of learning gains.

Questionnaire data were collected via *Wenjuan Xing* (Questionnaire Star), a widely used online survey platform in China. The researchers then entered all questionnaire item responses into an Excel dataset and imported the file into SPSS to compute descriptive statistics and examine item reliability (Cronbach's α). Only complete pretest-posttest pairs and corresponding questionnaire responses were retained for analysis to maintain the validity and reliability of the quantitative findings.

Qualitative data from semi-structured interviews were managed and coded using Microsoft Word, and then were analyzed thematically. All interviews were conducted in Chinese to facilitate natural expression and prevent language barriers. Each interview was audio-recorded via Tencent Meeting and transcribed into written Chinese. Transcripts were reviewed, coded, and analyzed inductively to allow patterns and themes to emerge from the data. The coding process was repeated three times to enhance consistency and ensure comprehensive identification of themes. Following the analysis, key excerpts were translated into English. To ensure translation accuracy and conceptual equivalence, three bilingual experts reviewed the translated texts.

Findings

The Effects of the LT-IPA Model on Chinese Undergraduates' English Speaking Proficiency

To address research question 1, which investigated the effects of the LT-IPA Model on Chinese undergraduates' English speaking proficiency, a pretest-posttest design was employed, with both descriptive and inferential statistical analyses reported.

Results from the Pretest and Posttest

Inter-rater reliability for the speaking assessments was examined using the intraclass correlation coefficient (ICC) for each speaking mode at the pretest and posttest. A two-way mixed model with absolute agreement was employed to estimate both single-measure and average-measure reliability.

For interpersonal speaking, ICC values indicated moderate to good agreement in the pretest (single measures ICC = .691; average measures ICC = .817) and good agreement in the posttest (single measures ICC = .799; average measures ICC = .888).

Similarly, presentational speaking demonstrated moderate to good reliability in the pretest (single measures ICC = .738; average measures ICC = .849) and moderate to good reliability in the posttest (single measures ICC = .702; average measures ICC = .825).

According to commonly cited guidelines (Koo & Li, 2016), ICC values between .75 and .90 indicate good reliability, whereas values between .50 and .75 reflect moderate reliability. Because the mean of the two raters' scores was used in subsequent analyses, the average measures ICC is reported as the primary reliability index. The results are shown in Table 3.

Table 3

Inter-Rater Reliability Results Based on Intraclass Correlation Coefficient (ICC)

IPA Speaking Mode	Pretest (N = 29)		Posttest (N = 29)	
	Single Measures	Average Measures	Single Measures	Average Measures
Interpersonal Speaking	.691	.817	.799	.888
Presentational Speaking	.738	.849	.702	.825

The normality of the mean difference (posttest minus pretest) for each speaking mode was assessed using the Shapiro-Wilk test before

conducting the paired-samples t-tests. For interpersonal speaking, the mean difference was 2.24 ($SD = 1.35$), with a range of 0 to 5. The distribution showed a skewness of 0.27 and kurtosis of -0.26. The Shapiro-Wilk test was not significant ($p = .083$), indicating that the assumption of normality was met. For presentational speaking, the mean difference was 2.21 ($SD = 1.29$), also ranging from 0 to 5, with a skewness of 0.44 and kurtosis of -0.06. The Shapiro-Wilk test was non-significant ($p = .056$), suggesting that the mean difference for presentational speaking was normally distributed (see Table 4).

Table 4

Results of the Normality Test for the Posttest-Pretest Mean Difference

IPA Speaking Mode	Mean Difference (n = 29)						<i>p</i> (Shapiro-Wilk)
	<i>M</i>	<i>SD</i>	Min	Max	<i>SK</i>	<i>KU</i>	
Interpersonal Speaking	2.24	1.35	0.00	5.00	0.27	-0.26	.083
Presentational Speaking	2.21	1.29	0.00	5.00	0.44	-0.06	.056

As shown in Table 5, both interpersonal and presentational speaking mean scores increased from the pretest to the posttest. Paired-samples t-tests showed statistically significant differences for interpersonal speaking, $t(28) = 8.92, p < .001$, and presentational speaking, $t(28) = 9.20, p < .001$. Large effect sizes were observed for both modes (see Table 5 for details).

Table 5

Results of the Pretest and Posttest

IPA Speaking Mode	Pretest (n = 29)		Posttest (n = 29)		95% CI (mean difference)	$t(28)$	<i>p</i> (two-sided)	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Interpersonal Speaking	14.76	1.60	17.00	1.87	[1.73, 2.76]	8.92	< .001	1.66
Presentational Speaking	14.69	1.56	16.90	1.35	[1.72, 2.70]	9.20	< .001	1.71

As shown in Table 6, paired-samples t-tests showed statistically significant differences for Language Function, $t(28) = 5.11, p < .001$; Text Type, $t(28) = 3.27, p = .003$; Communication Strategies, $t(28) = 6.33, p < .001$; Comprehensibility, $t(28) = 2.53, p = .017$; and Language Control, $t(28) = 4.14, p < .001$. Effect sizes ranged from small to large across the five criteria (see Table 6 for details).

Table 6*Results of Interpersonal Speaking in the Pretest and Posttest*

Criterion	Pretest (n = 29)		Posttest (n = 29)		95% CI (mean difference)	<i>t</i> (28)	<i>p</i> (two- sided)	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Language Function	3.10	0.41	3.59	0.50	[0.29,0.68]	5.11	<.001	0.95
Text Type	2.90	0.31	3.17	0.47	[0.10,0.45]	3.27	.003	0.61
Communication Strategies	2.66	0.61	3.45	0.74	[0.54,1.05]	6.33	<.001	1.18
Comprehensibility	3.34	0.48	3.66	0.48	[0.06,0.56]	2.53	.017	0.47
Language Control	2.76	0.44	3.14	0.44	[0.19,0.57]	4.14	<.001	0.77

As shown in Table 7, paired-samples *t*-tests showed statistically significant differences for Language Function, $t(28) = 5.87, p < .001$; Text Type, $t(28) = 5.95, p < .001$; Impact, $t(28) = 6.77, p < .001$; and Language Control, $t(28) = 3.84, p < .001$. However, Comprehensibility did not show a statistically significant difference, $t(28) = 1.00, p = .326$. Effect sizes ranged from negligible to large across the five criteria (see Table 7 for details).

Table 7*Results of Presentational Speaking in the Pretest and Posttest*

Criterion	Pretest (n = 29)		Posttest (n = 29)		95% CI (mean difference)	<i>t</i> (28)	<i>p</i> (two- sided)	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Language Function	3.28	0.45	3.83	0.38	[0.36,0.74]	5.87	<.001	1.09
Text Type	3.00	0.38	3.62	0.49	[0.41,0.83]	5.95	<.001	1.11
Impact	2.48	0.51	3.10	0.41	[0.43,0.81]	6.77	<.001	1.26
Comprehensibility	3.21	0.49	3.28	0.59	[-0.07,0.21]	1.00	.326	0.19
Language Control	2.72	0.53	3.07	0.26	[0.16,0.53]	3.84	<.001	0.71

Findings from Semi-Structured Interviews Regarding Speaking Proficiency

To complement the quantitative results with richer insights, the researchers conducted semi-structured interviews to explore how participants perceived the effects of the LT-IPA Model on their English speaking proficiency. Six participants were selected based on their speaking test scores to represent high-, mid-, and low-performing groups. The guiding question “How would you describe the overall effects of the LT-IPA Model on your English speaking proficiency?” was specifically designed to elicit detailed responses addressing research question 1. Analysis of the interview data

revealed five key themes regarding the overall effects of the LT-IPA Model on participants' English speaking proficiency: (1) improved language control, (2) enhanced communication strategies, (3) increased presentational confidence, (4) valuable speaking opportunity, and (5) overall positive impact.

Improved language control emerged as a central benefit of the LT-IPA Model, with participants noting gains in vocabulary, grammar, and accuracy. High- and mid-performing participants reported that the model not only corrected their linguistic errors but also provided clear guidance on how to improve. As Kevin (high-performing) explained, *"It enhanced my vocabulary and grammar."* Similarly, Li (mid-performing) remarked, *"It corrected my language errors more systematically and accurately, and helped me know how to improve and where to make improvements."*

Enhanced communication strategies were also mentioned, particularly among high-performing participants who valued the interpersonal speaking practice for fostering authentic interaction. Kevin (high-performing) stated, *"The interpersonal speaking improved my everyday communication strategies, enabling more authentic exchanges in English."*

A third theme, increased presentational confidence, reflected participants' recognition that repeated exposure to presentational speaking within the model reduced anxiety and improved self-confidence. Jelly (high-performing) commented, *"Through continuous practice and presentations, I have overcome my nervousness."* Kevin (high-performing) echoed, *"The practice opportunities provided by presentational speaking helped me overcome my stage fright, and I became more confident when speaking in front of others."*

For low-performing participants, the LT-IPA Model was perceived as a valuable speaking opportunity. The participants stressed that it was one of the few available opportunities to use English orally, given the lack of real-life communication settings, particularly with native speakers outside class. Yara (low-performing) reflected, *"We don't have other conditions, such as being able to communicate with foreigners frequently. This model is already very good for us."* Cynthia (low-performing) further observed that English-speaking opportunities were rare:

We don't have any other way to practice speaking except for watching videos online ourselves. The opportunities to practice spoken English are really very few in everyday life. I feel that this model is already a great opportunity for us.

Finally, participants across performing levels expressed an overall positive impact of the LT-IPA Model on their English speaking proficiency. Kevin (high-performing) affirmed that the model *"enhanced my overall English ability,"* while Jelly (high-performing) agreed that the experience had brought *"obvious improvement"* in his speaking performance. In the mid-performing

group, Li highlighted that the LT-IPA Model “*helped me see clear progress in English.*” Yue summarized this shared perception, stating that “*the overall impact on my English speaking proficiency is quite positive, very helpful for improvement.*” Low-performing participants echoed this view from a more contextualized perspective. As Yara noted, “*for our university, this model is already very good,*” and Cynthia added that it “*had a great and positive impact on our speaking,*” since it provided “*a rare and valuable opportunity to practice.*”

The Levels of Chinese Undergraduates’ Willingness to Speak (WTS) English in Class after Implementing the LT-IPA Model

For the second research question, this study examined Chinese undergraduates’ WTS in class following the implementation of the LT-IPA Model. After the semester-long intervention, participants ($n = 29$) completed a 20-item WTS questionnaire. The internal consistency of the questionnaire was assessed using SPSS, yielding a Cronbach’s alpha of .96, which indicates excellent reliability. The following parts present the quantitative and qualitative findings of the WTS questionnaire.

Results from the WTS Questionnaire

Table 8 presents the descriptive statistics summarizing participants’ responses to the WTS questionnaire. Across the 20 items, the mean scores ranged from 4.24 to 4.83 on the five-point Likert scale. The overall mean score was 4.50, which fell within the predefined descriptive interval corresponding to the “very high” WTS level. The standard deviations across the items ranged from 0.38 to 0.81. The highest mean score was observed for Item 3, “Speaking about the topic I was given time to prepare” ($M = 4.83$, $SD = 0.38$), whereas the lowest mean score was found for Item 13, “Performing interpersonal tasks” ($M = 4.24$, $SD = 0.79$). Despite this variation, both items still fell within the “very high” WTS range.

Among the seven dimensions, Topic of Discussion received the highest mean score ($M = 4.69$), followed by Interlocutor Effect ($M = 4.64$). Localized Tasks also received a relatively high mean rating ($M = 4.56$), while IPA Feedback Loop showed a comparable level ($M = 4.51$). Slightly lower mean scores were observed for IPA Backward Design ($M = 4.45$) and Task Type ($M = 4.37$). The lowest mean score was found for IPA Task Modes ($M = 4.32$), although this value still fell within the “very high” range.

Table 8*Summary of the Results of the WTS Questionnaire*

Dimension	Item	<i>M</i>	<i>SD</i>	Interpretation of WTS Level
Topic of Discussion	1. Speaking about the topic I am interested in	4.62	0.62	very high
	2. Speaking about the topic I am familiar with	4.62	0.56	very high
	3. Speaking about the topic I was given time to prepare	4.83	0.38	very high
	Total	4.69		very high
Localized Tasks	4. Working on speaking tasks aligned with my language level	4.62	0.56	very high
	5. Working on speaking tasks appropriate to my cultural background	4.52	0.57	very high
	6. Working on speaking tasks reflecting my real-life situations	4.55	0.57	very high
	Total	4.56		very high
Task Type	7. Working on speaking tasks individually	4.28	0.80	very high
	8. Working on speaking tasks in pairs	4.41	0.78	very high
	9. Working on speaking tasks in groups	4.41	0.78	very high
	Total	4.37		very high
Interlocutor Effect	10. Speaking with classmates I am familiar with	4.66	0.55	very high
	11. Speaking with classmates who participate actively	4.62	0.62	very high
	Total	4.64		very high
IPA Task Modes	12. Performing interpretive tasks	4.41	0.63	very high
	13. Performing interpersonal tasks	4.24	0.79	very high
	14. Performing presentational tasks	4.31	0.81	very high
	Total	4.32		very high
IPA Backward Design	15. Understanding the assessment criteria	4.38	0.68	very high
	16. Modeling speaking tasks before practice	4.52	0.57	very high
	Total	4.45		very high
IPA Feedback Loop	17. Receiving corrective feedback	4.66	0.55	very high
	18. Co-constructing feedback with the teacher and peers	4.55	0.57	very high
	19. The formative assessment covering five topics	4.41	0.68	very high
	20. Overall, the LT-IPA Model enhanced my willingness to speak English in class.	4.41	0.68	very high
	Total	4.51		very high
Overall Mean Score		4.50		very high

Findings from Semi-Structured Interviews Regarding WTS

Within the semi-structured interview protocol, one guiding question was designed to capture participants' perceptions of their willingness to speak English in class after using the LT-IPA Model: "How do you perceive your WTS after taking this course for a semester? For example, are there any changes in the WTS level?" Responses were collected from six participants across three performing groups (high, mid, and low, with two participants in each group). Thematic analysis of interview data revealed two major themes regarding participants' WTS when engaging with the LT-IPA Model: (1) greater willingness to speak and (2) perceived improvement in speaking proficiency.

Most participants described a greater willingness to speak English. They attributed this to three key factors: awareness of strengths and weaknesses, opportunities for preparation and practice, and a supportive classroom environment. High-performing participants emphasized that the LT-IPA Model provided opportunities to identify both shortcomings and strengths in speaking, which encouraged them to set concrete goals for improvement. As Jelly explained: "*I personally feel that my WTS has improved because, in this model, we can see each other's strengths and shortcomings. In such a situation, we can specifically improve ourselves. I think I am quite willing to speak English.*" Kevin echoed this sentiment, noting the role of preparation: "*It [WTS] has also improved for me because every week we need to prepare each topic. Our willingness to speak English has definitely increased.*"

Mid-performing participants also emphasized the role of consistent practice in building both confidence and fluency. Yue reflected on progress over time: "*I was not willing to speak English before, but this semester, after talking about the topics for a while and becoming more fluent, I felt that my willingness has improved.*" Similarly, Li acknowledged that initial reluctance stemmed from linguistic limitations: "*Compared to before, I was not too willing to speak English because I felt my grammar was not good and my vocabulary was insufficient, but it's slowly getting better now. I have more willingness to speak and make improvements.*"

For low-performing participants, one reported a greater WTS, particularly when compared with previous classroom experiences. Yara highlighted the advantages of a more accessible and supportive environment, stating:

I think my willingness to speak English in class has improved. Before, in the foreign teacher's class, I sometimes couldn't understand what he was saying or how to do it, so we would tend to discuss things in Chinese. But in this model, the tasks are well-structured and easy to understand.

At the same time, it is important to acknowledge that this positive perception was not uniform. Cynthia, another participant from the low-performing group, reported little change in her willingness:

I think, for me personally, my willingness to speak English is pretty much the same. Usually, I only use English when the teacher requires it for certain tasks or activities. But in most cases, our discussions are actually conducted in Chinese.

This contrast indicates that while the LT-IPA Model created more opportunities for some participants to engage in spoken English, others continued to rely primarily on their first language, indicating variability even within the same performing group.

In addition to a greater willingness, participants connected their WTS with perceived improvement in speaking proficiency. Many reported that repeated practice under the LT-IPA Model not only fostered WTS but also improved their spoken English. For instance, Kevin (high-performing group) noted, *“Additionally, I feel that my ability in spoken English has also improved.”* Yue (mid-performing group) similarly described a shift from fragmented speech to greater fluency: *“At first it felt like broken sentences, but after practicing for a period of time, I gradually felt that I could speak with greater fluency.”*

Overall, the thematic analysis demonstrated that the LT-IPA Model positively influenced participants’ WTS by fostering self-awareness of strengths and weaknesses, increasing opportunities for structured preparation and practice, and providing a supportive classroom environment. In turn, greater WTS was associated with their perceived improvement in speaking proficiency.

Discussion

Improvement in Speaking Proficiency

In interpersonal speaking, participants demonstrated steady progress across all five criteria. Performance levels were interpreted based on the IPA analytic rubric descriptors (Adair-Hauck et al., 2013), which define expectations in terms of how effectively learners meet communicative goals. Within this framework, participants’ performance moved toward more consistent fulfillment of task requirements, progressing from minimally meeting expectations to more consistently strong performance. The most notable improvement was observed in Communication Strategies, suggesting increased ability to initiate and sustain interaction and manage communication breakdowns. Improvements in Language Function, Text Type, and Language Control reflected greater capacity to produce extended

discourse with improved fluency, lexical range, and grammatical accuracy. Comprehensibility showed only a small increase.

A similar pattern was observed in presentational speaking. Significant gains were found in four of the five criteria, with the largest increase in Impact, indicating enhanced ability to engage an audience and deliver structured messages. Improvements in Language Function and Text Type suggested stronger organization and coherence in extended discourse, while gains in Language Control reflected improved accuracy and fluency. In contrast, Comprehensibility did not show a statistically significant change, which may be attributed to a ceiling effect, as many participants had already reached mid-level scores at the pretest.

The observed gains can be explained by two key features of the LT-IPA Model: the structured implementation of the IPA framework and the incorporation of localized tasks. The IPA cycle, integrating interpretive, interpersonal, and presentational modes, provided repeated opportunities for practice and performance supported by ongoing feedback. The two-week sequencing of tasks allowed students to build from input and interaction to prepared output, which may have contributed to improvements across multiple criteria.

In addition, localized tasks adapted from textbook content enabled students to engage with familiar and meaningful topics. Prior research has shown that topic familiarity and relevance increase willingness to speak (Cao & Philp, 2006; Riasati, 2012), which in turn promotes language development through increased opportunities for use (Riasati, 2018).

In contrast to the findings of Glisan et al. (2007), where learners performed best in the presentational mode, participants in the current study showed slightly higher scores in interpersonal speaking. One possible explanation lies in the nature of the assessment criteria associated with each mode. The impact dimension in presentational speaking, covering aspects such as audience engagement, intonation, and pacing, was less emphasized in the previous English speaking courses. As a result, participants may have had limited prior exposure and training in these areas, leading to lower initial scores. By comparison, communication strategies in interpersonal speaking draw more heavily on initiating, sustaining, and repairing conversations. These skills can often be supported by participants' accumulated formulaic expressions (e.g., routine phrases and interactional chunks), making them more accessible even at earlier stages. This may explain the relatively higher baseline performance in interpersonal speaking.

However, the larger effect size observed for the impact dimension suggested that participants made more substantial gains in this area over time. This indicated that although participants started with weaker performance in

presentational delivery features, they were able to develop these skills gradually through targeted practice during the intervention.

Insights from the semi-structured interviews further substantiated the quantitative findings. Participants frequently emphasized improvements in language control. This resonates with Zapata (2016), who highlighted IPA's strength in promoting active language processing rather than rote memorization. Enhanced communication strategies also emerged as a recurrent theme among high-performing participants, aligning with Zapata's (2016) argument regarding the authenticity of IPA tasks that mirror real-life communicative demands. Finally, participants expressed generally positive perceptions of the intervention, noting increased speaking opportunities and clearer connections between classroom learning and assessment, echoing findings reported in previous university-level IPA research (Altstaedter & Krosl, 2018; Glisan et al., 2007; Martel, 2018; Zapata, 2016).

Collectively, the quantitative and qualitative findings indicated that the LT-IPA Model effectively enhanced speaking proficiency across communicative modes. Consistent with Adair-Hauck et al. (2013), the model not only developed linguistic ability, interactional competence, and presentational skills but also provided more authentic, pedagogically aligned opportunities for meaningful oral communication.

Levels of Willingness to Speak (WTS) English

The questionnaire and interview findings revealed that participants' WTS within the LT-IPA context is collectively shaped by pedagogical and situational factors. Across the seven dimensions investigated, the topic of discussion emerged as the primary driver of WTS. Questionnaire responses indicated that topic interests, familiarity, and preparation fostered willingness to contribute, corroborating earlier research suggesting that topic familiarity, interest, and preparation are critical determinants of WTS (Riasati, 2012; Riasati & Rahimi, 2018).

Localized task design constituted another key facilitator. When tasks were aligned with participants' language levels, they tended to report higher perceived linguistic competence, a key antecedent of WTC in MacIntyre et al.'s (1998) heuristic model. Equally important, tasks grounded in participants' real-life situations and cultural background fostered authenticity and meaningfulness of classroom communication. This supports the principles of task-based language teaching and the IPA manual (Adair-Hauck et al., 2013; Ellis et al., 2020), which emphasize the design of tasks that mirror real-world communicative needs and reflect learners' cultural identities. These findings indicated that localization functions not merely as a pedagogical enhancement

but as a necessary condition for fostering WTS by ensuring linguistic relevance, contextual authenticity, and cultural appropriateness.

Task type further shaped participants' WTS. All three task types, including individual, pair, and group work, were perceived as beneficial, but collaborative contexts were associated with higher levels of WTS, consistent with previous findings that learners tend to speak more readily in interactional settings (Riasati, 2012, 2018). Interview insights emphasized the value of rotating partners and maintaining small group sizes, supporting the argument that dynamic grouping strategies expand learners' communicative repertoire and create more equitable participation opportunities (Cao & Philp, 2006).

The interlocutor also appeared to play an important role in shaping participants' WTS. Familiarity with classmates and active interlocutor participation were both associated with higher levels of WTS, consistent with previous research suggesting that rapport is linked to reduced apprehension and that responsive partners tend to facilitate more productive interaction (Cao & Philp, 2006; Riasati, 2012; Riasati & Rahimi, 2018). These findings suggested that WTS is not solely an individual disposition but is co-constructed through social relationships and interactional dynamics.

Within the IPA framework, all three task modes, namely interpretive, interpersonal, and presentational, were perceived as supportive of WTS, with interpretive tasks receiving the highest ratings. This is in line with the cyclical nature of IPA, where comprehension-based activities provide a foundation for subsequent interpersonal and presentational output (Adair-Hauck et al., 2013). Presentational tasks were valued for their structured nature, consistent with previous research suggesting that planned production is associated with higher performance (Glisan et al., 2007).

The backward design elements of the IPA, particularly explicit assessment criteria and performance modeling, were also associated with participants' WTS. Transparent criteria appeared to reduce uncertainty, while model performances helped clarify expectations and set goals, consistent with principles of Understanding by Design (Wiggins & McTighe, 2005) and the IPA pedagogy (Adair-Hauck et al., 2013). Participants' positive perceptions of rubric clarity also align with Sedor's (2022) argument that criterion-specific guidance supports accurate self-assessment and targeted improvement.

Finally, the IPA feedback loop proved instrumental in supporting WTS. Corrective feedback, especially on pronunciation and accuracy, was viewed as encouraging rather than discouraging, aligning with findings that well-delivered corrective feedback can enhance performance (Sedor, 2022). At the same time, descriptive and co-constructive feedback were particularly valued for their specificity, actionability, and dialogic nature, resonating with Adair-Hauck and Troyan's (2013) emphasis on co-constructed assessment as central to learner development.

Although this study did not empirically test the relationship between WTS and speaking proficiency, as the two were treated as separate dependent variables, the qualitative findings offered valuable insights into their potential connection. Interview participants often associated higher levels of WTS with perceived gains in speaking proficiency. Several participants reported that repeated practice within the LT-IPA Model encouraged them to speak more and was accompanied by improvements in the quality of their spoken English. These findings align with prior research that has explored the relationship between WTS and speaking proficiency. For example, Riasati (2012) qualitatively observed that learners who are more willing to speak tend to engage more frequently in target-language practice, which, in turn, supports language development. Similarly, Riasati's (2018) quantitative study demonstrated that learners with higher levels of WTS achieved higher speaking test scores, attributing this improvement to increased practice and language use during class. These studies suggested that fostering WTS is not only a motivational objective but also a pedagogical pathway to achieving measurable gains in speaking proficiency.

Conclusion

The findings of this study yielded several pedagogical implications for language educators, curriculum designers, and assessment practitioners working within EFL contexts. First, the results indicated that localizing IPA can better address students' linguistic needs, contextual realities, and cultural backgrounds. Localization involves not only modifying task topics but also providing appropriate linguistic and strategic scaffolding to support students across performing levels. Second, the study highlighted that the pedagogical value of the LT-IPA Model emerges from complete and sustained implementation. Semester-long, multi-theme IPA cycles, supported by backward design principles and an integrated feedback loop, ensure that students benefit from iterative practice, performance, and co-constructed feedback, which short-term or single-theme interventions may fail to capture (Jiang et al., 2021). Third, the study demonstrated the value of engaging students as active stakeholders throughout the whole learning process. Beyond participating in the feedback session, students can contribute to topic selection, task localization, and refinement of assessment rubrics, thereby enhancing engagement, motivation, and alignment between performance expectations and students' perceptions of communicative competence.

The present study made three key contributions to the field of performance-based language assessment. First, it addressed an empirical gap by providing a systematic account of IPA implementation in Chinese tertiary EFL education, a context where speaking proficiency is often

underdeveloped and empirical evidence remains scarce. By applying and examining the LT-IPA Model in a college English course, the study demonstrated how the IPA framework can be effectively adapted to meet the linguistic and pedagogical needs of Chinese undergraduates. Second, the study advanced research on IPA localization by developing a semester-long, thematically coherent IPA design grounded in students' real-life experiences. This localized approach extends prior calls for contextual adaptation (Martel & Bailey, 2016), offering empirical validation for integrating ACTFL's IPA framework with localized, context-sensitive tasks. Third, the study expanded the theoretical scope of IPA research by linking IPA implementation to students' willingness to speak (WTS) English, a critical yet underexplored construct in the Chinese EFL context. The findings showed that localized IPA can reduce reticence and enhance communicative willingness, positioning IPA not only as an assessment tool but also as an affective catalyst for active participation.

Despite its contributions, the study's findings must be interpreted within their methodological boundaries. The small, intact sample and single-institution context limited generalizability, while the 16-week duration prevented claims about long-term developmental trajectories. Methodological refinements, such as incorporating pre-intervention questionnaires, delayed posttests, and multi-site sampling, would strengthen causal inferences and external validity. Additionally, although the study treated speaking proficiency and WTS as separate dependent variables, interview evidence suggested a potential connection between the two constructs, warranting future investigations employing correlation or regression analyses.

Building on these insights, several avenues for future research are proposed. Large-scale curricular reform initiatives could explore the feasibility of implementing a "pure IPA curriculum" (Zapata, 2016) with multi-semester sequencing, enabling longitudinal analyses of linguistic, communicative, and affective growth. The development of IPA-based teaching materials and assessment resources also represents a critical next step for institutionalizing IPA within Chinese higher education. Moreover, integrating IPA with complementary pedagogical frameworks, such as project-based learning (Madel, 2017) or content-based learning (Darhower & Smith-Sherwood, 2021), may further broaden its applicability, allowing for richer cognitive, cultural, and affective outcomes.

In conclusion, this study provided evidence that localized IPA implementation had a positive effect on students' speaking proficiency and examined the situational and pedagogical factors within the LT-IPA Model that are associated with students' willingness to speak English. By bridging empirical gaps, extending theoretical perspectives, and offering actionable

pedagogical insights, the LT-IPA Model provided a foundation for future research and innovation in performance-based language instruction and assessment within China and beyond.

Acknowledgements

The research grant funds have been provided by the 72nd Anniversary of His Majesty King Bhumibol Adulyadej Scholarship and the 90th Anniversary Chulalongkorn University Fund (Ratchadapiseksomphot Endowment Fund).

About the Authors

Lingyi Liu: A Ph.D. candidate in the Language Assessment and Evaluation track of the English as an International Language (EIL) Program, College of Interdisciplinary and Integrative Studies, Chulalongkorn University, Bangkok, Thailand. Her research interests include English language teaching and assessment.

Punchalee Wasanasomsithi: An associate professor at Chulalongkorn University Language Institute. She holds a Ph.D. in Language Education from Indiana University, Indiana, USA. Her research interests include writing instruction, assessment literacy, and learning-oriented assessment.

References

- Adair-Hauck, B., Glisan, E., & Troyan, F. (2013). *Implementing integrated performance assessment*. American Council on the Teaching of Foreign Languages.
- Adair-Hauck, B., Glisan, E. W., Koda, K., Swender, E. B., & Sandrock, P. (2006). The integrated performance assessment (IPA): Connecting assessment to instruction and learning. *Foreign Language Annals*, 39(3), 359–382.
- Adair-Hauck, B., & Troyan, F. J. (2013). A descriptive and co-constructive approach to integrated performance assessment feedback. *Foreign Language Annals*, 46(1), 23–44. <https://doi.org/10.1111/flan.12017>
- Altstaedter, L. L., & Kros, M. (2018). Perceptions of integrated performance assessments among beginning Spanish college students: A preliminary study. *Journal of Foreign Language Teaching and Applied Linguistics*, 5(2), 99–114.

- Amoah, S., & Yeboah, J. (2021). The speaking difficulties of Chinese EFL learners and their motivation towards speaking the English language. *Journal of Language and Linguistic Studies*, 17(1), 56–69.
- Butler, Y. G. (2011). The implementation of communicative and task-based language teaching in the Asia-Pacific region. *Annual Review of Applied Linguistics*, 31, 36–57. <https://doi.org/10.1017/s0267190511000122>
- Cao, Y., & Philp, J. (2006). Interactional context and willingness to communicate: A comparison of behavior in whole class, group and dyadic interaction. *System*, 34(4), 480–493. <https://doi.org/10.1016/j.system.2006.05.002>
- Cheng, X. (2000). Asian students’ reticence revisited. *System*, 28(3), 435–446. [https://doi.org/https://doi.org/10.1016/S0346-251X\(00\)00015-4](https://doi.org/https://doi.org/10.1016/S0346-251X(00)00015-4)
- Creswell, J. W. (2024). My 35 years in mixed methods research. *Journal of Mixed Methods Research*, 18(3), 203–215.
- Darhower, M., & Smith-Sherwood, D. (2021). Bridging the language-literature divide via integrated performance assessments in an introductory Hispanic literature course. *Hispania*, 104(3), 395–413.
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.
- ETS.org. (2023). *TOEFL iBT Test and Score Data Summary* <https://www.ets.org/pdfs/toefl/toefl-ibt-test-score-data-summary-2023.pdf>
- Fang, F. (2010). A discussion on developing students’ communicative competence in college English teaching in China. *Journal of language teaching and research*, 1(2), 111–116.
- Gan, Z. (2013). Understanding English speaking difficulties: an investigation of two Chinese populations. *Journal of Multilingual and Multicultural Development*, 34(3), 231–248. <https://doi.org/10.1080/01434632.2013.768622>
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language teaching*, 40(2), 97–118.
- Glisan, E. W., Uribe, D., & Adair-Hauck, B. (2007). Research on integrated performance assessment at the post-secondary level: Student performance across the modes of communication. *Canadian Modern Language Review*, 64(1), 39–67.
- Hughes, R., & Reed, B. S. (2016). *Teaching and researching speaking*. Routledge.
- IELTS.org. (2022). *Test Statistics* https://ielts.org/researchers/our-research/test-statistics#Test_performance
- Jiang, S., Wu, J., Xu, J., & Tian, J. (2021). Connecting integrated performance assessment to Chinese classroom teaching and learning. In Y. Zhang & X. Gao (Eds.), *Frontiers of L2 Chinese Language Education* (pp. 68–89). Routledge.

- Kim, N. (2019). Challenges and trials: Implementing localized TBLT for novice L2 learners throughout three semesters. *English Teaching*, 74(3), 113–139. <https://doi.org/10.15858/engtea.74.3.201909.113>
- Kim, Y., Jung, Y., & Tracy-Ventura, N. (2017). Implementation of a localized task-based course in an EFL Context: A study of students' evolving perceptions. *TESOL Quarterly*, 51(3), 632–660. <https://doi.org/10.1002/tesq.381>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
- Kramsch, C., & Hua, Z. (2016). Language and culture in ELT. In G. Hall (Ed.), *The Routledge handbook of English language teaching* (pp. 38–50). Routledge.
- Littlewood, W. (2011). Communicative language teaching: An expanding concept for a changing world. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 541–557). Routledge.
- Liu, C., & Guo, R. (2020). A study of localization of task-based language teaching in China. *Chinese Journal of Applied Linguistics*, 43(2), 205–218. <https://doi.org/10.1515/cjal-2020-0013>
- Liu, Y., Mishan, F., & Chambers, A. (2021). Investigating EFL teachers' perceptions of task-based language teaching in higher education in China. *The Language Learning Journal*, 49(2), 131–146.
- 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.
- Madel, R. (2017). The Intersection of the IPA and PBL: Exploring a unit that demonstrates how the integrated performance assessment pairs naturally with project based learning. *Pennsylvania Language Forum*, 87(1), 18–20.
- Malone, M. E., Marshall, C., Vyn, R., & Favero, C. (2024). *ACTFL Proficiency Guidelines 2024*. The American Council on the Teaching of Foreign Languages.
- Martel, J. (2018). Postsecondary students' and instructors' evaluative comments about ACTFL's integrated performance assessment. *Applied Language Learning*, 28(1), 1–18.
- Martel, J., & Bailey, K. M. (2016). Exploring the trajectory of an educational innovation: Instructors' attitudes toward IPA implementation in a postsecondary intensive summer language program. *Foreign Language Annals*, 49(3), 530–543. <https://doi.org/10.1111/flan.12210>
- McCroskey, J. C. (1992). Reliability and validity of the willingness to communicate scale. *Communication Quarterly*, 40(1), 16–25.

- McCroskey, J. C., & Baer, J. E. (1985). *Willingness to communicate: The construct and its measurement*. [Paper presentation]. The Annual Convention of the Speech Communication Association. Denver, Colorado, USA.
<https://doi.org/10.1080/01463379209369817>
- MOE. (2020). *Guidelines on college English teaching*. Ministry of Education.
- Plengkham, B., & Wasanasomsithi, P. (2023). Effects of integrated performance assessment modules on English speaking ability of Thai EFL undergraduate students. *LEARN Journal: Language Education and Acquisition Research Network*, 16(1), 448–472.
- Ren, W., & Han, Z. (2016). The representation of pragmatic knowledge in recent ELT textbooks. *ELT Journal*, 70(4), 424–434.
- Riasati, M. J. (2012). EFL learners' perception of factors influencing willingness to speak English in language classrooms: A qualitative study. *World Applied Sciences Journal*, 17(10), 1287–1297.
- Riasati, M. J. (2018). Willingness to speak English among foreign language learners: A causal model. *Cogent Education*, 5(1), 1455332.
- Riasati, M. J., & Rahimi, F. (2018). Situational and individual factors engendering willingness to speak English in foreign language classrooms. *Cogent Education*, 5(1), 1513313.
- Sandrock, P., & Swender, E. (2012). *ACTFL performance descriptors for language learners*. The American Council on the Teaching of Foreign Languages.
- Sedor, N. (2022). Integrated performance assessment (IPA): Implementation, task types, and feedback. *Foreign Language Annals*, 56(1), 170–190. <https://doi.org/10.1111/flan.12661>
- Shrum, J., & Glisan, E. (2010). *Teacher's handbook: Contextualized language instruction*. Heinle & Heinle.
- Wei, R., & Su, J. (2012). The statistics of English in China: An analysis of the best available data from government sources. *English Today*, 28(3), 10–14.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed). Association for Supervision and Curriculum Development (ASCD).
- Xu, Z. (2018). Exploring English as an international language - curriculum, materials and pedagogical Strategies. *RELC Journal*, 49(1), 102–118.
<https://doi.org/10.1177/0033688217753848>
- Zapata, G. C. (2016). University students' perceptions of integrated performance assessment and the connection between classroom learning and assessment. *Foreign Language Annals*, 49(1), 93–104.
<https://doi.org/10.1111/flan.12176>

Appendix A

Pretest and Posttest

IPA Speaking Pretest

Interpersonal Task

Instruction:

Work in pairs to plan a trip to Dali using the information provided in the travel guide. Discuss the following questions with your partner and decide together on your travel plan.

Questions to discuss:

- 1) When are you planning to travel, and why?
- 2) What mode of transportation will you use to get to Dali, and how will it suit your travel preferences?
- 3) What attractions, activities, or experiences are you going to explore in Dali?
- 4) How many days will your trip last, and how will you structure your itinerary?

You will have 2 minutes to prepare your ideas individually.

Then, you will have 4 minutes to discuss your travel plan with your partner.

(If there are three participants in a group due to an odd number of students, allocate 5 minutes for the discussion.)

Presentational Task

Instruction:

Present your planned trip to Dali independently based on your discussion with your partner. Your presentation should address the four questions discussed during the interpersonal task, providing detailed and well-organized information about your travel plan.

You will have 1 minute to prepare your presentation.

Then, you will have 3 minutes to deliver your presentation individually.

IPA Speaking Posttest

Interpersonal Task

Instruction:

Work in pairs to recommend a Chinese city to move to for a common friend of yours using the information provided in the “Chinese Cities of Opportunity 2024” report.

Discuss the following questions with your partner and decide together which city would be the best option.

Questions to discuss:

- 1) Which city do you recommend for your friend to move to, and why?
- 2) What are the main opportunities and features offered by the city you recommend?
- 3) What challenges might your friend face when relocating to that city?
- 4) Evaluate and decide on the best city for your friend to move to and explain why you agree.

You will have 2 minutes to prepare your ideas individually.

Then, you will have 4 minutes to discuss your recommendation with your partner.

(If there are three participants in a group due to an odd number of students, allocate 5 minutes for the discussion.)

3. Presentational Task

Instruction:

Present the recommended city for your friend independently based on your discussion with your partner. Your presentation should address the four questions discussed during the interpersonal task, providing detailed and well-organized information about your recommendation.

You will have 1 minute to prepare your presentation.

Then, you will have 3 minutes to deliver your presentation individually.

Appendix B

Five Thematic Units of the LT-IPA Model

Topic 1: Food in the University Canteen

Interpersonal Speaking Task

Instructions:

Today you and your friend are having lunch/dinner at the university canteen. Look at the canteen menu and discuss what each of you would like to eat.

Your conversation should include:

1. Meal Recommendation & Location
2. Reasons for Recommendation (consider factors such as taste, portion size, price, and nutrition)
3. Deciding which one you would like to go for lunch/dinner

Since you may have different preferences, share your opinions and reasons, then work together to choose a meal that satisfies both of you. Be ready to explain your final decision.

Presentational Speaking Task

Instructions:

Christine is new to the university and needs your help choosing what to eat at the canteen. Based on your group discussion, prepare a short presentation recommending a specific meal.

Your presentation should include:

1. Meal Recommendation & Location – Introduce the meal you recommend and specify where Christine can find it in the canteen.
2. Reasons for Recommendation – Explain why you chose this meal. Consider factors such as taste, portion size, price, and nutrition.
3. Rating & Personal Opinion – Give a rating (e.g., out of 5 stars) and share your personal thoughts on the meal.

At the end of the presentations, Christine might ask follow-up questions, so be prepared to justify your recommendation.

Topic 2: City for Traveling or Living in

Interpersonal Speaking Task

Instructions:

You and your partner will have a discussion about a Chinese city you want to visit or live in. Each of you will share information about your chosen city and ask each other follow-up questions. Try to learn about your partner's city and express your opinions.

Your conversation should include:

1. Introduce the Chinese city you want to visit/live in.
2. Explain why you chose this city (personal experiences, attractions, lifestyle, etc.).
3. Ask at least two follow-up questions to your partner.
4. React to your partner's responses by expressing interest or giving your opinion.
5. Conclude by summarizing what you learned about your partner's city and stating if you would like to visit it.

Presentational Speaking Task

Instructions:

Now that you have discussed your chosen city with your partner, it's time to present! You will give a short presentation about the Chinese city you want to visit or live in.

Your presentation should include:

1. Introduction: Introduce the city and why you chose it.
2. Personal Connection: If you have visited or lived there before, share your experience.
3. Main Attractions: Highlight what makes the city special (culture, lifestyle, attractions, etc.).
4. Conclusion: Summarize why this city is your top choice and encourage your audience to visit.

Topic 3: Talking about People

Interpersonal Speaking Task

Instructions:

Work with your partner to discuss someone who has influenced you the most.

Your conversation should include:

1. Who is this person? (family member, friend, teacher, celebrity, etc.)
 2. What is this person like? (e.g., appearance, personality traits)
 3. Why did this person influence you the most? (Tell a story or give a specific example.)
 4. Ask your partner questions and respond to their story. Try to find something you have in common or a difference in your experiences.
- Note: This is a conversation, not a speech. Try to keep the conversation going by showing interest and asking questions.

Presentational Speaking Task

Instructions:

Now that you've talked with your partner about a person who influenced you the most, you will give a short individual presentation to your classmates.

Your presentation should include:

1. Introduce the person.
2. Describe the person's appearance or personality.
3. Explain how this person influenced you.
4. Reflect on what you learned from this person.

Note: Your audience does not know this person, so make your message interesting and meaningful.

Topic 4: Recommending Movies/ TV series

Interpersonal Speaking Task

Instructions:

In pairs or small groups, discuss your favorite movie or TV series. You will each take turns asking and answering questions. Use *Into the Wild* as a reference example if you need inspiration.

Your conversation should include:

1. What is your favorite movie or TV series? Why do you like it?
2. Who is the main character? What kind of person is she or he?
3. What happens in the story? (Tell about one or two main events.)
4. Is there a message or theme in it? What do you think it is?
5. Would you recommend this movie/TV series to a friend? Why or why not?

Presentational Speaking Task

Instructions:

Prepare a short presentation about a movie or TV series that you would like to recommend. You may use *Into the Wild* as a reference for structure and vocabulary.

Your presentation should include:

1. The title and genre of the movie/TV series.
2. A summary of the plot.
3. A description of the main character.
4. A theme or message of the story.
5. Why is it meaningful or inspiring to you?

Topic 5: Sharing Chinese Culture

Interpersonal Speaking Task

Instructions:

A group of foreign exchange students will be visiting your school next week. Your class is responsible for introducing them to an aspect of Chinese culture. With your partner, discuss different elements of Chinese culture (e.g., Chinese calligraphy 书法, paper-cutting 剪纸, dumpling making 包饺子, traditional festivals, tea ceremony 茶道, Peking Opera 京剧, martial arts 武术, etc.). Choose one to introduce to foreign exchange students. You must explain your reasons and ask your partner for their opinion. Your final choice must be one you both agree on.

Your conversation should include:

1. What is the cultural item you are thinking about? (Describe what it is and what people do with it.)
2. Why do you want to teach this cultural item?

(Is it fun? meaningful? important to Chinese culture?)

3. How will you teach it or help the exchange students experience it?

(Will you do an activity, show something, let them try something?)

4. What challenges might the students have when learning this?

(Is it hard to understand, pronounce, make, or do?)

5. How can you help them overcome these challenges?

(What support, explanation, or tools can you give?)

Presentational Speaking Task

Instructions:

Now that you have chosen the cultural item, prepare a presentation for the foreign exchange students. Use the same five questions to organize your speech.

Your presentation should include:

1. What is the item?

(Name it in English and Chinese; briefly explain what it is.)

2. Why do you want to teach this item to foreign students?

(Why is it special or important in Chinese culture?)

3. How will you teach it or let them experience it?

(Describe the activity or demonstration.)

4. What could be difficult for them to understand or do?

(Identify possible challenges.)

5. How will you help them with these challenges?

(Explain how you will make it easier or more fun.)