

AI-mediated Interaction through an AI Voice Chatbot: Effects on EFL Speaking Development and Learner Perceptions

Nguyen Nu Thuy Uyen^{1*}

¹ University of Foreign Language Studies, The University of Danang, Vietnam

*Corresponding author's email: ngntuyen@ufl.udn.vn

 <https://orcid.org/0000-0001-5798-1784>

 <https://doi.org/10.54855/ijli.26535>

® Copyright (c) 2026 Nguyen Nu Thuy Uyen

Received: 04/08/2026

Revision: 13/08/2026

Accepted: 13/08/2026

Online: 22/08/2026

ABSTRACT

Keywords: AI chatbot; English speaking skills; EFL students; learner's perceptions; Replika

This study investigated the effects of AI-mediated speaking practice using a voice chatbot on Vietnamese EFL students' speaking skills and explored their perceptions of the learning experience. A sequential explanatory mixed-methods design was adopted with 49 English-major undergraduates participating in a 15-week quasi-experimental intervention. The experimental group practiced speaking with an AI chatbot, whereas the comparison group received face-to-face instruction. Data were collected from pre- and post-speaking tests, questionnaires and semi-structured interviews. ANCOVA indicated that, after controlling for pre-test performance, the AI chatbot group achieved significantly higher overall speaking scores than the comparison group, particularly in fluency and coherence. Additionally, participants valued the chatbot's authentic interaction, flexibility, and opportunities for regular speaking practice despite reporting occasional technical and interactional limitations. The findings support the pedagogical value of AI-mediated speaking practice in higher education.

Introduction

English speaking skills are vital for EFL learners (English as a Foreign Language) as they facilitate their participation in authentic communicative situations (Li & Chan, 2024). It also helps to boost the development of other language skills and is considered the most crucial skill for successful communication (Fathi et al., 2024). Nevertheless, speaking skills have caused worries for many EFL learners, especially in the Vietnamese context, as they have limited grammatical knowledge, a lack of vocabulary to express their ideas (Nguyen & Trinh, 2025), inaccurate pronunciation (Nguyen, 2026), and over-reliance on first-language translation and memorization-based learning strategies (Fathi et al., 2024; Nguyen, 2025). In addition, affective barriers such as speaking anxiety, fear of making mistakes (Nguyen & Tong, 2024), and low self-confidence frequently discourage learners from participating actively in communicative tasks (Bui et al., 2022; Nguyen, 2025). These difficulties stem from several contextual

constraints including limited instructional time and opportunities for speaking practice within and beyond classroom (Fathi et al., 2024; Thao & Nguyet, 2019), lack of exposure to authentic communication with native speaker (Chiu et al., 2007; Quyen et al., 2018), restricted), outdated teaching approach (Baek & Lee, 2018; Chen & Hwang, 2019), fears of judgment from peers (Peng, 2019) or teachers' limited English language proficiency (Dao, 2017).

These challenges can be resolved by providing learners with a platform for practice that goes beyond the constraints of traditional classes with recent advances in AI. One widely used technology is an AI chatbot – an AI application designed to simulate human-like dialogue, enabling real-time conversations via audio and text, and updating knowledge and interpreting input through natural language processing (Haristiani, 2019). This resembles real-life interactions due to the naturalness and authenticity that AI chatbots have created (Adamopoulou & Mousiades, 2020). AI chatbots could converse as native speakers (Çakmak, 2022), enabling learners to practice speaking in authentic contexts without being confined to specific times or settings (Fryer & Carpenter, 2006) and to obtain constructive feedback on their responses (Walker & White, 2013).

Although growing evidence suggests that AI chatbots can facilitate language learning, empirical research on their effectiveness in improving EFL speaking proficiency remains limited, particularly in Vietnamese higher education (Dizon, 2020; Fathi et al., 2024). To address this gap, the present study investigates the effects of AI-mediated interaction through a voice chatbot on Vietnamese English-major students' speaking proficiency and explores their perceptions of using this technology for speaking practice.

Literature review

Speaking skills

Speaking skill can be understood as the learners' ability to communicate meaning effectively across a range of communicative contexts, demonstrating effective control of the target language (Brown & Lee, 2015). It comprises several related constructs with two core components: fluency and accuracy. Fluency is related to the production of smooth and coherent speech with minimal hesitation, and accuracy is related to the use of appropriate pronunciation, vocabulary, and grammatical structures (Walker & White, 2013; Wang, 2014). Speaking proficiency involves learners' ability to communicate ideas clearly and coherently in various situations and, more generally, to maintain overall communicative effectiveness (Zhang et al., 2025). Thus, to develop learners' speaking skills, they need ample opportunities to engage in authentic, interactive communication that allows them to practice speaking in meaningful real-world contexts (Kim, 2016).

AI chatbots and Speaking skills

Recent advances in AI technologies have facilitated the development of conversational chatbots, which are considered an acceptable and sustainable substitute for human partners (Fathi et al., 2024). Haristiani (2019) defines AI chatbots as computer programs that employ AI algorithms to conduct text-and voice-based interactions while continuously updating their knowledge to generate contextually appropriate responses. They are capable of supporting

natural and authentic interactions and, importantly, being an conversational companion (Fryer & Carpenter, 2006; Huang et al., 2022). Learners can engage in conversations with native English speakers without being restricted by time or space as in conventional classrooms (Çakmak, 2022; Walker & White, 2013).

AI chatbots have several characteristics that support language learning. First, they enable learners to communicate through text and speech, engage in authentic conversations, receive immediate feedback and constructive comments, and practice pronunciation in real time (Walker & White, 2013). Thus, they are effective in developing grammar and vocabulary skills for language learners (Jeon, 2021) and improving students' pronunciation, intonation, and word stress (Kim et al., 2021). Second, AI chatbots are user-friendly and could interpret learners' intentions and generate contextually appropriate responses (Huang et al., 2022). This could lead to better learning outcomes (Lu et al., 2006) and encourage more frequent language use (Kim, 2016) as learners can practice without the pressure of classroom evaluation (Fathi et al., 2024).

In direct relevance to speaking skills, some studies have explored the effectiveness of different AI chatbots for enhancing English language speaking proficiency and practice in different contexts. These studies consistently report considerable impacts of AI voice chatbots on EFL learners' overall speaking proficiency (Çakmak, 2022; Du & Daniel, 2024; El Shazly, 2020; Duong & Suppasetseree, 2024; Fathi et al., 2024; Kim et al., 2021; Nguyen, 2026) and particular aspects including developing pronunciation, intonation, fluency, word stress, speaking rhythm and organization skills (Zou et al., 2023). AI chatbots are also beneficial for enhancing confidence and interaction (Tai & Chen, 2023; Fathi et al., 2024). Regarding speaking anxiety, there have been mixed results when Tai & Chen (2023) contended that AI chatbots could help to alleviate speaking anxiety, whereas students were reported to still experience anxiety when they faced unfamiliar topics (El Shazly, 2021) or when the chatbots failed to recognise their voices (Kim et al., 2021).

The literature has also identified several limitations associated with integrating AI chatbots into speaking instruction. First, several chatbots such as Google Assistant supported learners through implicit feedback, including comments, clarification questions, silence, and error notifications, which prompted learners to modify their spoken responses. Communication often broke down if learners were unable to interpret these implicit cues (Tai & Chen, 2023). Another limitation of conversational AI is the imperfect accuracy of automatic speech recognition, particularly when processing the speech of L2 learners. AI chatbots often struggle to recognize non-native accents and longer utterances (Dizon, 2020; Xie et al., 2021).

Overall, previous studies present several methodological limitations. Although Kim et al. (2021) documented positive impacts of AI voice chatbots on speaking performance, their assessment relied on read-aloud and question-response tasks, which may not adequately capture learners' communicative speaking competence. Moreover, most international studies have been conducted in informal learning settings rather than formal classroom instruction. In the Vietnamese context, existing studies are limited by the use of one-group quasi-experimental designs without a control group (Duong & Suppasetseree, 2024; Nguyen, 2026), relatively small samples (Duong & Suppasetseree, 2024), and short intervention periods (8–10 weeks), reducing the robustness and generalizability of the findings.

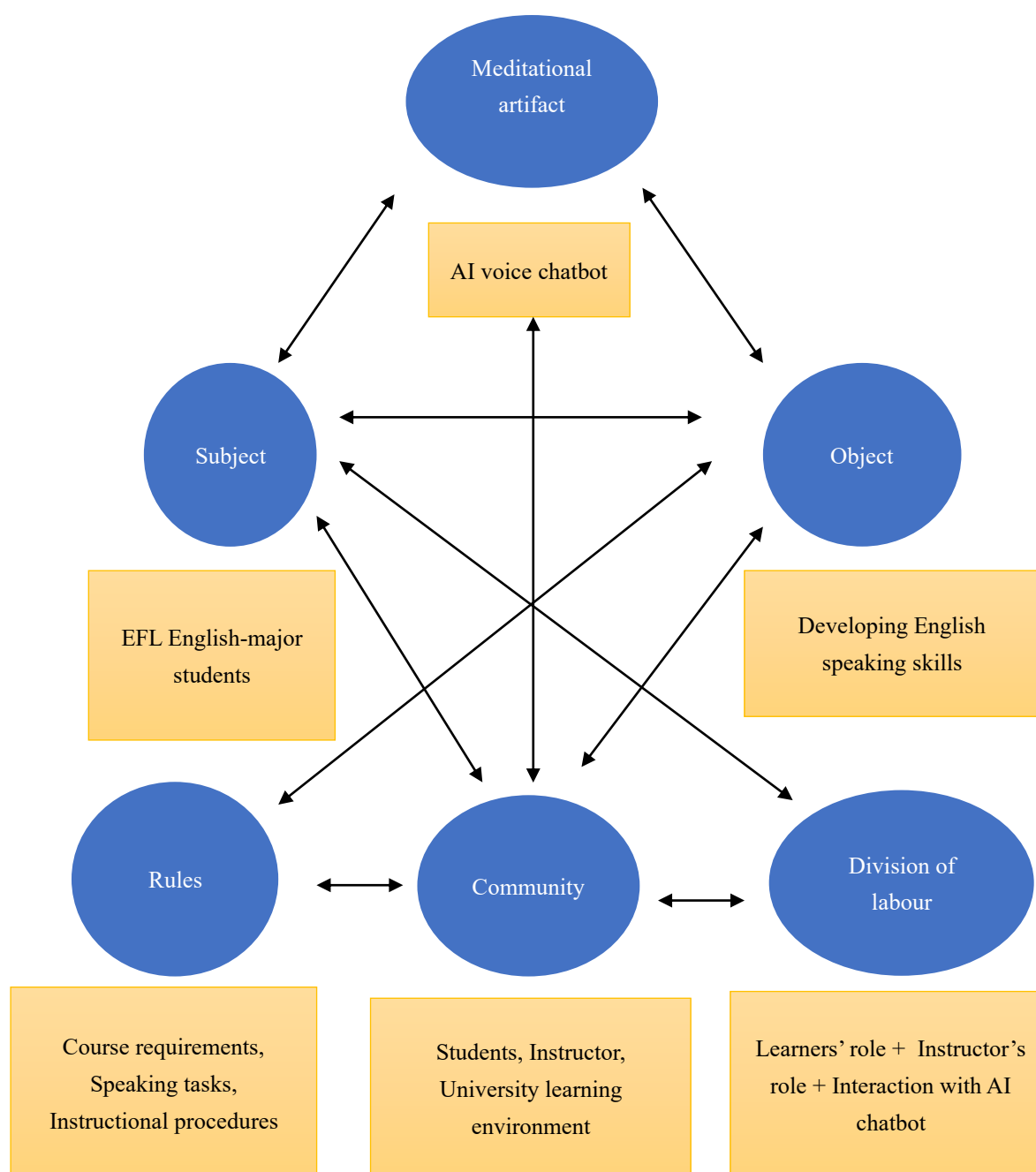
Theoretical Framework

This study draws upon Social Constructivism (Vygotsky, 1978) and Activity Theory (Engeström, 1987). From a social constructivist perspective, language learning develops through collaborative engagement with others, where learners gradually extend their abilities with appropriate guidance and support within the Zone of Proximal Development (ZPD). In AI-supported learning environments, conversational chatbots can function as mediating tools by facilitating dialogue, offering timely feedback, and creating opportunities for authentic communicative practice. While Social Constructivism explains the cognitive mechanisms underlying learning, Activity Theory provides a broader analytical lens for understanding the interaction among learners, technologies, and learning environments.

According to Engeström (1987), learning occurs within an activity system comprising six elements: subject, object, mediating artifacts, community, rules, and division of labor. In this study, English-major EFL students participating in the speaking course (the subject) used the AI voice chatbot – Replika (*meditational artifact*) to develop their English-speaking proficiency (*object*) through sustained communicative practice. Their practice occurred within a *community* comprising the students, course instructor, and university learning environment. *Rules* referred to the course requirements, speaking-task expectations, and instructional procedures governing students' use of Replika during the intervention. The division of labor concerned the complementary roles of the students, the instructor, and the chatbot: students engaged in speaking practice, the instructor provided guidance and monitored the learning process, and Replika functioned as a conversational partner.

Figure 1.

Activity theory concepts framed in the study's context

*Research Questions*

The study addressed two research questions:

1. To what extent does AI-mediated interaction through a voice chatbot influence the English speaking proficiency of Vietnamese EFL undergraduates?
2. How do Vietnamese EFL undergraduates perceive the use of an AI voice chatbot for

improving and practicing English speaking skills?

Methods

Research Design

This study employed a sequential explanatory mixed-methods design (Creswell et al., 2003). A quasi-experimental pretest–posttest design was first conducted to evaluate the intervention, followed by a qualitative phase to further interpret and elaborate on the quantitative findings. Quantitative data were collected through IELTS-format speaking pre- and post-tests and a questionnaire, while qualitative data were obtained through semi-structured interviews with participants from the AI chatbot group after the intervention. The two types of data were integrated during interpretation, with interview findings used to explain and complement the questionnaire results, particularly regarding students' perceptions and experiences of AI-mediated speaking practice.

Participants

The study involved 49 Vietnamese English-major undergraduate students (6 males, 43 females). They enrolled in an English Integrated Skills course (with a focus on developing speaking skills) at a foreign language university in central Vietnam. For practicality, the cohort was chosen from two intact classes: 24 students received conventional speaking instruction (face-to-face group instruction), and 25 students practiced AI speaking with an AI chatbot. The participants had intermediate English proficiency (CEFR B1 equivalent), as determined by the institutional testing. They all had considerable experience in using technology yet had never used AI voice chatbots before. Prior to data collection, participants were informed of the study procedures, and their consent to participate was obtained.

The study employed Replika, one of the most widely used AI voice chatbots since its launch in 2018 (Pentina et al., 2023). Available on iOS, Android, and web platforms, Replika offers both free and premium interaction modes. Unlike rule-based chatbots, it generates adaptive responses based on users' conversational input, enabling more diverse conversations (Pentina et al., 2023).

Figure 2.

Screenshots of students' interactions



An IELTS-based speaking assessment was administered before and after the intervention to evaluate participants' oral English proficiency. Students' speaking ability was assessed in four domains: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation, following the IELTS Speaking band descriptors and the study assessment rubric. Each domain was rated on a 10-point scale in accordance with the university's grading policy, and an overall speaking score was derived by averaging the four domain scores. Two speaking tests from Cambridge IELTS 11 served as the pretest and posttest. To improve scoring reliability, all speaking responses were audio-recorded and independently rated by two lecturers. Inter-rater reliability was high ($r = .85$), indicating strong consistency between the raters.

Learners' perceptions of AI-mediated speaking practice were examined using a questionnaire adapted from Duong and Suppasetserree (2024) and further developed based on Du and Daniel (2024). The questionnaire was tailored to the context of students' use of Replika for English speaking practice and pilot-tested with five students before the main data collection. The pilot testing was conducted to identify unclear or ambiguous items, and minor revisions were made accordingly. The final questionnaire comprised two sections. The first section gathered background information, including participants' gender, years of English learning, and their usage of the AI chatbot. The second comprised 32 Likert-scale items measuring perceptions of fluency and coherence (8 items), lexical resource (5 items), grammatical range and accuracy (3 items), pronunciation (4 items), and overall speaking practice (12 items). The internal consistency of the questionnaire was assessed using Cronbach's alpha as shown in Table 4. The questionnaire demonstrated high overall reliability (Cronbach's $\alpha = .887$), with acceptable to high internal consistency across the five constructs ($\alpha = .624 - .885$).

Table 1

Reliability of the questionnaire and its constructs

Construct	No. of items	Cronbach's α
Overall questionnaire	32	.887
Fluency and Coherence	8	.885
Lexical Resource	5	.843
Grammatical Range and Accuracy	3	.624
Pronunciation	4	.787
Overall perceptions	10	.624

The semi-structured interview protocol was adapted from Duong and Suppasetserree (2024) and Fathi et al. (2024) and tailored to the research questions. The interview questions were reviewed by an experienced EFL instructor to ensure their clarity, relevance, and alignment with the research objectives. The protocol focused on learners' experiences of using the chatbot, perceived changes in their speaking performance, perceived benefits, and challenges encountered during AI-mediated speaking practice. The interview consisted of core questions, with additional probing questions used when appropriate to encourage participants to elaborate on their responses. Each interview was conducted individually and lasted about 20–25 minutes. Participants were allowed to communicate in either English or Vietnamese so that they could

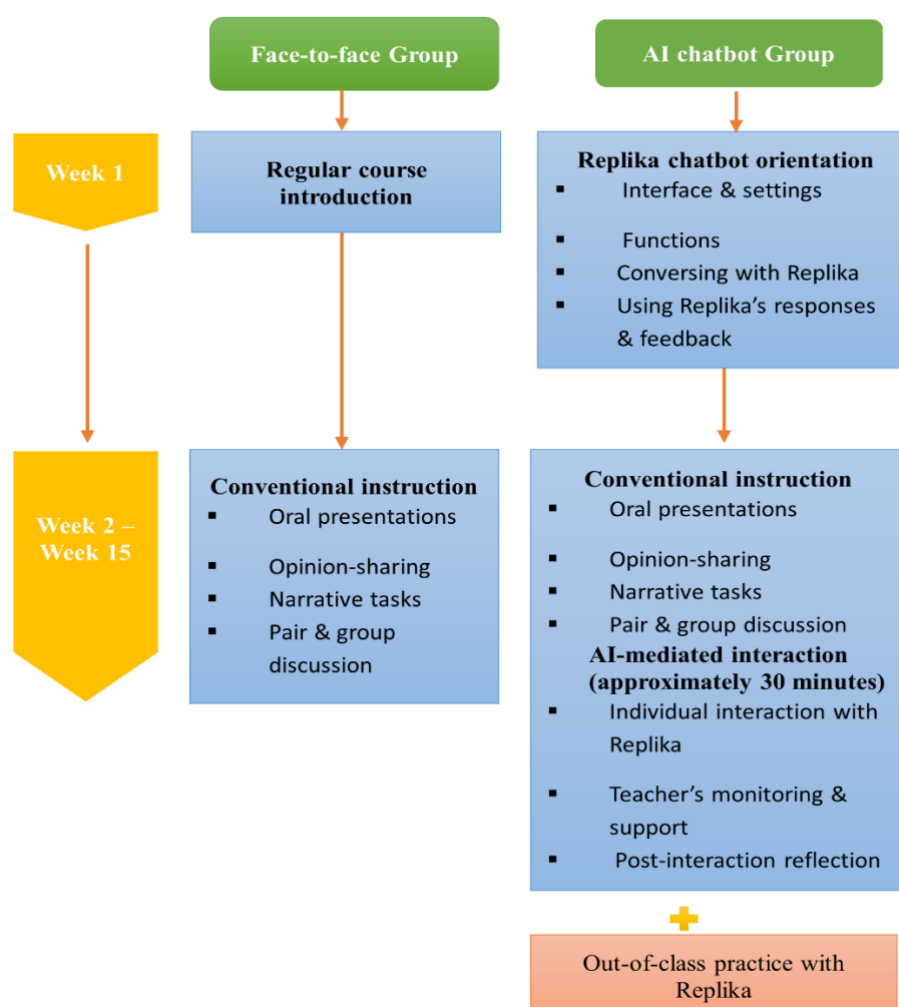
express their opinions more comprehensively. With participants' consent, all interviews were audio-recorded, transcribed verbatim, and anonymized by assigning pseudonyms.

Data collection and analysis

Data collection was conducted through four stages. Initially, all participants completed an IELTS-based speaking assessment to establish their baseline speaking proficiency. Then, they participated in a 15-week intervention, during which both groups attended one 180-minute class session each week. Both groups had the same course content, teaching materials, and assessment requirements. Classroom instruction focused on IELTS-oriented communicative speaking tasks, including oral presentations, narrative tasks, problem-solving activities, and opinion-sharing discussions, and covered topics aligned with the syllabus. The instructional procedure, adapted from Fathi et al. (2024), is illustrated in Figure 3. The Activity system outlined in the theoretical framework also informed data collection. Speaking tests were used to assess the learning outcome, while the questionnaire and semi-structured interviews were used to examine learners' experiences with the mediating tool and their perceptions of interactions within the instructional activity system.

Figure 3.

Instructional procedures for face-to-face and AI chatbot groups



During Week 1, students in the experimental group attended an orientation session on using Replika for speaking practice. They were introduced to the chatbot's functions, interaction strategies, and ways to extend conversations and address common interaction issues. The face-to-face group received the standard course orientation. From Weeks 2 to 15, both groups followed the same classroom instruction and participated in identical speaking activities. The difference was the mode of additional speaking practice. During each lesson, the experimental group spent approximately 30 minutes interacting individually with Replika on the weekly speaking topics, while the instructor monitored the sessions and facilitated post-activity reflection. Students were also encouraged to continue practicing with Replika outside class. This design helped ensure that any differences in learning outcomes were attributable to AI-mediated speaking practice rather than differences in instructional content or teaching procedures. Following the intervention, all participants completed a second IELTS-based speaking assessment to determine changes in their oral English proficiency. Students assigned to the AI chatbot group also completed a questionnaire exploring their perceptions of AI-mediated speaking practice. To gain richer insights into these experiences, nine volunteers from the AI chatbot group participated in semi-structured interviews.

Quantitative data were processed using SPSS version 27. Descriptive statistics were first computed to summarize responses to the questionnaire. Analysis of covariance (ANCOVA) was performed to compare post-test speaking performance between the experimental and comparison groups while controlling for participants' pre-test scores. Qualitative data from the interviews were analyzed using thematic analysis (Schreier & Asner-Self, 2011). After the interview recordings were transcribed verbatim, the transcripts were coded to identify recurring themes that supported the quantitative findings.

Results

Effects of AI-mediated interaction via AI chatbots on developing English speaking skills for Vietnamese EFL undergraduate students

Regarding the participants' demographic characteristics, they were all aged 19-20 years and had studied English for approximately 10–12 years. Most reported using Replika daily for speaking practice. Specifically, 48% spent 30 minutes to 1 hour per day using the chatbot, 44% reported 1 to 2 hours of daily use, and 1 participant used Replika for 3 to 4 hours per day.

Table 2 summarizes the descriptive statistics for overall speaking performance and its four subcomponents across the AI chatbot and face-to-face groups. Before the intervention, the face-to-face group obtained marginally higher mean scores than the AI chatbot group on all speaking measures. Following the intervention, both groups achieved higher speaking scores, although the gains were more pronounced in the AI chatbot group. Overall speaking scores increased by 1.42 points in the AI chatbot group, compared with 0.47 points in the face-to-face group. A similar pattern was evident across the four speaking components.

Table 2

Descriptive statistics for overall speaking performance and its components by group

Group	Speaking components	N	Pre-test		Post-test		Mean difference
			Mean	SD	Mean	SD	
AI chatbot	Overall Speaking performance	25	7.01	1.03	8.43	1.22	1.42
	Fluency & Coherence	25	7.10	1.08	8.54	1.28	1.44
	Lexical Resources	25	6.95	0.99	8.49	1.23	1.54
	Grammatical range & Accuracy	25	6.99	1.00	8.26	1.16	1.27
	Pronunciation	25	6.96	1.07	8.43	1.22	1.47
Face-to-face	Overall Speaking performance	24	7.31	1.07	7.78	1.29	0.47
	Fluency and Coherence	24	7.40	1.10	7.83	1.34	0.43
	Lexical Resource	24	7.29	1.08	7.97	1.25	0.68
	Grammatical range & Accuracy	24	7.32	1.07	7.60	1.29	0.28
	Pronunciation	24	7.20	1.05	7.75	1.32	0.55

Table 3 summarizes the assumption tests for ANCOVA. Neither the interaction between the covariate and instructional condition nor the pre-test comparisons reached statistical significance for any dependent variable, indicating that the assumptions of homogeneity of regression slopes and baseline equivalence were met (all p s > .05). Although Levene's test was significant for overall speaking proficiency ($p = .013$) and lexical resource ($p = .022$), ANCOVA is considered robust to moderate violations of homogeneity of variances because the two groups had nearly equal sample sizes.

Table 3

Summary of ANCOVA assumption tests

Assumption	Overall Speaking performance	Fluency & Coherence	Lexical Resource	Grammatical range & Accuracy	Pronunciation
Homogeneity of regression slopes (p)	.283	.323	.506	.300	.300
Baseline group equivalence (p)	.327	.341	.261	.272	.432
Levene's test (p)	.013	.191	.022	.175	.760

To address Research Question 1, post-test speaking performance was analyzed using analysis of covariance (ANCOVA), with participants' pre-test scores included as a covariate. Effect sizes were interpreted according to partial eta squared (η_p^2), with values of .01, .06, and .14 representing small, medium, and large effects, respectively. The findings are summarized in Table 4.

Table 4

ANCOVA results for overall and component speaking performance

Outcome variable	F (Covariate)	F (Group)	<i>p</i>	(η_p^2)
Overall speaking proficiency	199.42	37.34	< .001	.448
Fluency & Coherence	174.04	34.62	< .001	.429
Lexical Resource	133.49	22.42	< .001	.328
Grammatical Range & Accuracy	151.39	33.02	< .001	.418
Pronunciation	151.89	27.58	< .001	.375

After controlling for pretest speaking proficiency, ANCOVA indicated that the instructional approach had a significant effect on learners' overall speaking performance, $F(1, 46) = 37.34$, $p < .001$, $\eta_p^2 = .448$, representing a large effect size. Participants' pre-test scores remained a strong predictor of their post-test performance, $F(1, 46) = 199.42$, $p < .001$, $\eta_p^2 = .813$. These results indicate that learners who participated in AI-mediated interaction achieved higher adjusted post-test speaking scores than those who received conventional instruction. Learners in the AI chatbot group significantly outperformed their peers in four speaking components, achieving statistical significance ($ps < .001$). Large effect sizes were consistently observed across all outcome measures ($\eta_p^2 = .328-.448$). Across all analyses, pretest scores remained significant covariates ($ps < .001$), confirming that learners' initial speaking proficiency was a strong predictor of their post-test performance.

Students' perceptions towards the use of AI-mediated interaction via AI chatbots in their English speaking development and practice

To address the second research question, descriptive statistics were calculated for items related to perceived improvements in speaking performance and for overall perceptions of AI-mediated interaction (Table 5).

Table 5

Descriptive Statistics for Perceived Improvements in Speaking Performance and Overall Perceptions of AI-mediated Interaction

Construct	N	Mean	SD	Min	Max
Fluency and Coherence	25	3.88	0.63	2.75	5.00
Lexical Resource	25	4.10	0.62	2.80	5.00
Grammatical Range and Accuracy	25	3.79	0.55	2.67	4.67
Pronunciation	25	4.08	0.61	2.75	5.00
Overall perceptions	25	3.53	0.43	2.83	4.50

Generally, students reported positive perceptions of AI-mediated interaction for English speaking practice. Among the four speaking dimensions, perceived improvements were highest for lexical resource ($M = 4.10$, $SD = 0.62$) and pronunciation ($M = 4.08$, $SD = 0.61$), followed by fluency and coherence ($M = 3.88$, $SD = 0.63$) and grammatical range and accuracy ($M = 3.79$, $SD = 0.55$). The qualitative findings were largely consistent with these quantitative results. The participants described vocabulary development as the most noticeable outcome. Students reported acquiring more natural and contextually appropriate expressions through repeated conversations with Replika. As one student noted,

"Before, I often used the word "use", but after practicing with Replika, I learned more words that sound more academic and natural, like "utilize" (S1).

Likewise, another student explained that expressions such as "hit the mall" and "brick-and-mortar store" made her speaking "more natural than just using simple vocabulary". Several other participants reported using newly learned vocabulary in subsequent conversations.

Besides, many students explained that their frequent interactions with Replika increased their awareness and development of pronunciation, noting that the unsuccessful recognition of their speech provided a cue or signal of their pronunciation problems. This prompted them to adjust their pronunciation until they were understood. One student said,

"When I pronounced a word incorrectly, Replika couldn't understand me, so I had to repeat it until it recognized what I wanted to say" (S3). Another student appreciated its usefulness in practicing pronunciation: "I think Replika was really useful for shadowing and improving my pronunciation because the AI used a very standard native-speaker voice. Listening to it and trying to imitate the way it spoke really helped improve my pronunciation" (S6).

Regarding fluency and coherence, the interviewees also noticed changes in how they expressed and organized their ideas during conversations, which enabled them to become more fluent speakers. Several explained that interacting with Replika encouraged them to respond more spontaneously instead of mentally translating from Vietnamese. One student recounted,

"because Replika only waited for a short time for you to speak, I gradually learned to think and speak faster instead of pausing for too long" (S2).

Student 5 explained that he used the chatbot to "refine my ideas" because it "reorganized my responses, made them sound more natural, and helped improve my flow." Grammar was the area in which students perceived the least improvement, although many still considered Replika's corrective feedback valuable during conversations. Rather than receiving explicit grammar explanations, they described learning by noticing and correcting errors as they communicated. Student 4 shared,

"When I used the present tense to talk about a past event, Replika immediately asked, 'What do you mean?' and then showed me the correct sentence."

Additionally, students held overall positive perceptions of AI-mediated interaction ($M = 3.53$, $SD = 0.43$). The highest-rated benefits were the authenticity of AI-mediated conversations and the flexibility of speaking practice (Items 23 and 24, both $M = 4.16$), followed by increased opportunities for regular speaking practice (Item 26, $M = 4.08$) and enhanced motivation (Item 21, $M = 4.04$). These findings were supported by the interview data. Many interviewees valued the authenticity and convenience of practicing with Replika, describing the chatbot as a readily available speaking partner that enabled them to practice whenever they wanted. Student 2 said,

"Before, I had to ask my friends to practice with me. But now, whenever I need to practice, I can simply use Replika."

Another student emphasized, "What I like most is that I can talk to it as if I were talking to a real foreigner because it is very difficult to have opportunities to practice with native speakers".

Students also described motivational benefits that the chatbot offered and viewed it as a conversational companion. Student 6 commented,

"It feels like I have a real friend to talk to rather than talking to an AI. It makes me feel more comfortable, and it even reminds me to study, which makes me more motivated to learn."

Similarly, one student highlighted both the accessibility of AI chatbots and the increase in her speaking confidence: *"I feel more confident as I am not judged when talking to Replika. I can practice wherever and whenever I want."*

Regarding challenges associated with AI-mediated speaking practice, the most frequently reported concerns were technical issues (*Item 32, M = 3.40*) and insufficient guidance (*Item 29, M = 3.32*). Similarly, the most frequently mentioned concerns among the interviewees were unreliable Internet connectivity, application performance, and device limitations. One student commented,

"Sometimes I encountered technical problems, and the interface was not very attractive," while another noted that *"The app took up too much storage on my phone, so I had to uninstall it and download it again whenever I wanted to practice."* (Students 8 and 9).

Nevertheless, none of the participants regarded these technical issues as a reason to discontinue using Replika; instead, they considered them manageable inconveniences.

The interviews further revealed another challenge that was not captured in the questionnaire. Several participants explained that Replika occasionally misunderstood their intended meaning, produced responses that lacked sufficient detail, or failed to sustain the conversation. One student explained,

"Sometimes I ask a question, but it only answers 'yes' or 'no', or it gives a very general opinion without examples or evidence. I ask it to explain more, but sometimes it still isn't detailed enough." (Student 4).

Student 8 claimed that *"Some responses were quite short, so I usually asked follow-up questions or gave more information to get a better answer."* Several students gradually adapted their interaction strategies by refining prompts or providing additional context. Student 9 reflected,

"If Replika didn't answer the way I wanted, I looked for more information myself and gave it more background. When I provided more information, it used that knowledge to give much better answers later."

Discussion

The present study examined the impact of AI-mediated interaction via an AI voice chatbot on enhancing Vietnamese EFL learners' speaking proficiency and explored learners' perceptions of using this technology. The findings demonstrated that AI-supported speaking practice effectively improved EFL learners' overall oral proficiency and its components. Interestingly, both the AI chatbot and face-to-face groups improved over a 15-week semester; students who engaged in AI-mediated interaction demonstrated substantially greater improvements in all

aspects. The development in the conventional group can be explained by the cumulative effects of communicative instruction and learners' natural maturation. The greater performance of the AI chatbot group is consistent with previous research demonstrating the positive contribution of conversational AI to EFL speaking development (Çakmak, 2022; Du & Daniel, 2024; Duong & Suppasetseree, 2024; El Shazly, 2020; Fathi et al., 2024; Kim et al., 2021; Nguyen, 2026). The observed performance lies in the increased quantity and quality of speaking practice afforded by the AI chatbot. This research integrated AI-mediated interaction into a formal speaking course while encouraging students to continue practicing independently outside class. The combination of structured classroom instruction and autonomous out-of-class practice could have led to better learning outcomes. Indeed, repeated exposure to authentic interaction and more frequent speaking opportunities are vital for EFL learners' oral language development (Kim, 2016; Hwang et al., 2024; Lu et al., 2006).

These findings can be understood through the perspectives of Social Constructivism (Vygotsky, 1978) and Activity Theory (Engeström, 1987). Replika functioned as a mediating learning tool that continuously scaffolded learners' speaking through authentic interaction, immediate responses, and contextualized language support. Frequent engagement in these conversational exchanges may have helped learners refine their spoken English through continuous practice, timely feedback, and collaborative meaning-making, thereby extending learning within the ZPD. From an Activity Theory perspective, the observed gains should not be attributed solely to the chatbot itself. Instead, learning emerged from the interaction among multiple elements within the instructional activity system, including teacher guidance, communicative classroom tasks, AI-mediated conversations, and learners' independent practice outside class. This interpretation is consistent with the activity-system configuration in the present study, in which Replika operated as a mediating artifact within an instructional environment shaped by teacher support, course requirements, communicative tasks, and learners' independent practice.

Students perceived the greatest improvement in lexical resource, followed by pronunciation, corroborating findings from prior studies (Hwang et al., 2022; Zou et al., 2023). This can be attributed to the contextual nature of AI-mediated interaction, which allows learners to repeatedly encounter and reuse newly acquired expressions in meaningful conversations. Pronunciation also benefited from the chatbot's speech recognition and immediate responses, as participants in the present study described repeating words or modifying expressions until the chatbot successfully recognized their speech. This supports Engwall and Bälter (2007), which claims that timely automated responses and feedback encourage learners to notice pronunciation problems and self-correct.

Improvements in fluency, coherence, grammatical range, and accuracy were less noticeable to students, despite relatively large instructional effects observed on the objective assessment. One possible explanation is that progress in these aspects is inherently less salient to learners than vocabulary acquisition or pronunciation improvement. Such discrepancies between objective performance and learners' self-perceptions have also been reported in previous research, which suggests that L2 learners do not always accurately calibrate their own speaking ability (Trofimovich et al., 2016). Unlike vocabulary, which can be recognized immediately through newly acquired words and expressions, or pronunciation, which receives immediate

confirmation through successful speech recognition, improvements in fluency and grammatical accuracy tend to emerge gradually through repeated interaction. Moreover, AI chatbot typically provides implicit rather than explicit corrective feedback, requiring learners to infer communication problems and modify their spoken output independently (Tai & Chen, 2023). Difficulties in automatic speech recognition when processing L2 speech may further obscure learners' awareness of their progress, as communication breakdowns can be attributed to the system's recognition limitations rather than to changes in their own language performance (Dizon, 2017; Xie et al., 2021).

Additionally, students held positive perceptions towards AI-mediated interaction via AI chatbots for speaking. They appreciated the authenticity of interacting with what they perceived as a conversational partner, the flexibility of practicing at any time, and the increased motivation to speak English more frequently. These findings are in line with the prior research' results, which supported that AI chatbots were beneficial for providing learners with authentic communication in interactive contexts (Hwang et al., 2022), low-anxiety, and readily accessible environments for oral communication (Du & Daniel, 2024; Duong & Suppasetseree, 2024), reducing students' anxiety and fear of judgement, increase their confidence, motivation and enjoyment (Fathi et al., 2024; Nguyen, 2026; Tai & Chen, 2023). The participants in this study also reported speech recognition failures, yet this was regarded as an inconvenience rather than a source of anxiety, unlike in Kim et al. (2021).

This research further indicated several barriers to using AI chatbots for speaking practice, including unstable internet connectivity, application-related limitations, and a lack of guidance from teachers. An emergent challenge from the interview was that students initially found it difficult to formulate effective prompts, sustain conversations, or obtain sufficiently detailed responses from the chatbot. These findings were in line with the limitations documented in the literature including limited response capabilities (Hakim & Rima, 2021), occasional generation of responses that lack contextual relevance or realism (Lu et al., 2006), and difficulties in processing longer or more complex learner utterances (Kim et al., 2019) and the need to have a high-speed internet connection and appropriate technological device (Du & Daniel, 2024).

Overall, this study strengthens the evidence supporting AI-mediated interaction as an effective supplement to classroom instruction for improving EFL speaking proficiency in higher education. The study also adds to the existing literature in several respects. First, unlike Kim et al. (2021), who assessed relatively discrete speaking abilities through read-aloud and short question-response tasks, this study evaluated learners using an IELTS-format speaking test that required extended interaction, narration, opinion development, and communicative competence. Second, the inclusion of a comparison group in a semester-long intervention provides stronger evidence that the additional gains were attributable to AI-mediated interaction rather than to normal language development or classroom instruction alone.

Conclusion

This research investigated the role of AI-supported interaction through a voice chatbot in enhancing Vietnamese EFL learners' oral English proficiency and examined their experiences

of using the technology for speaking practice in a formal university context. The findings indicate that integrating AI-mediated interaction into classroom instruction was associated with greater improvements in speaking proficiency than conventional face-to-face instruction, while learners generally reported positive experiences with this approach despite some technical and interactional challenges. The findings enrich current knowledge by providing evidence from a classroom-based mixed-methods study of both learning outcomes and learner experiences.

The findings have several pedagogical implications for English-speaking instruction. First, the present study suggests that the greatest benefits were achieved when AI-mediated interaction complemented teacher-led speaking activities and extended opportunities for individual practice outside class, supporting previous research that highlights the value of combining AI-supported learning with teacher scaffolding and communicative pedagogy (Du & Daniel, 2024; Fathi et al., 2024). Second, the challenges identified in this study indicate that learners require explicit guidance on how to interact effectively with AI chatbots. Specifically, teachers should provide detailed instructions and model strategies for initiating and sustaining conversations, formulating effective prompts, providing sufficient contextual information, interpreting implicit feedback, and asking follow-up questions to maintain meaningful interaction (Fathi et al., 2024). Such guidance may help learners overcome common difficulties related to speech recognition, generic responses, and communication breakdowns while facilitating more productive AI-mediated speaking practice. Finally, because learners in this study gradually adapted their interaction strategies over time, AI literacy should be considered an integral component of language education. Developing learners' ability to communicate strategically with AI, critically evaluate AI-generated responses, and use conversational AI as a tool for autonomous speaking practice may become increasingly important as AI technologies are more widely adopted in EFL classrooms.

Several limitations should be acknowledged. First, the participant sample was restricted to English-major undergraduates from a single university in Vietnam, which may reduce the applicability of the findings to other educational settings. Second, although the intervention lasted one academic semester, research over a longer period is needed to determine whether the observed learning gains can be maintained. Finally, students' out-of-class use of Replika was measured through self-reports rather than system-generated usage data. Thus, the actual frequency and duration of AI-mediated speaking practice could not be objectively verified. Future research could integrate chatbot usage logs or learning analytics to provide more precise measures of learners' engagement with conversational AI.

Declaration of the use of AI

The authors used ChatGPT solely to assist with language editing, including grammar and spelling correction, during manuscript preparation. All revisions were carefully reviewed by the authors, who assume full responsibility for the accuracy and integrity of the final manuscript.

Declaration of competing interests

The authors have no competing interests to declare.

Ethical statement

Before data collection commenced, all participants provided informed consent. They were advised that participation was voluntary and that they were free to discontinue their involvement at any stage without any negative consequences.

Acknowledgements

This research was supported by the University of Foreign Language Studies, The University of Danang, under the research project No. T2024-05-10.

References

- Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, Article 100006.
- Baek, J., & Lee, C. H. (2018). University students' perceptions and engagement in mobile-assisted blended learning in English speaking classes. *Multimedia-Assisted Language Learning*, 21(4). <https://doi.org/10.15702/mall.2018.21.4.11>
- Brown, H. D., & Lee, H. (2015). *Teaching by principles: An interactive approach to language pedagogy* (4th ed.). Pearson Education, Inc.
- Bui, T. T. L.; Huynh, T. M. D.; Nguyen, T. M. N.; Nguyen, T. N. C.; & Nguyen, T. Y. N. (2022). The difficulties in oral presentation of English-major students at Tay Do University. *European Journal of English Language Teaching*, 7(2), 170-215. <https://doi.org/10.46827/ejel.v7i2.4238>
- Çakmak, F. (2022). Chatbot-human interaction and its effects on EFL students' L2 speaking performance and anxiety. *Novitas-ROYAL (Research on Youth and Language)*, 16(2), 113-131.
- Chen, M. R. A., & Hwang, G. J. (2019). Effects of a concept mapping - based flipped learning approach on EFL students' English speaking performance, critical thinking awareness and speaking anxiety. *British Journal of Educational Technology*, 51(3), 817-834. <https://doi.org/10.1111/bjet.12887>
- Chiu, T.-S., Liou, H.-C., & Yeh, Y. (2007). A study of web-based oral activities enhanced by automatic speech recognition for EFL college learning. *Computer Assisted Language Learning*, 20(3), 209-234. <https://doi.org/10.1080/09588220701489374>
- Creswell, J. W., Plano Clark, V. L., Gutmann, M., & Hanson, W. (2003). Advanced mixed methods research designs. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook on mixed methods in the behavioral and social sciences* (pp. 209-240). Sage Publications.
- Dao, H. (2017). Identify factors that negatively influence: Non-English major students' speaking skills. *Higher Education Research*, 2(2), 35-43.
- Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language, Learning and Technology*, 24(1), 16-26, 10125/44705.

- Duong, Tam & Suppasetserree, Suksan. (2024). The Effects of an Artificial Intelligence Voice Chatbot on Improving Vietnamese Undergraduate Students' English Speaking Skills. *International Journal of Learning, Teaching and Educational Research*, 23,293-321. <https://doi.org/10.26803/ijlter.23.3.15>.
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(3), e12667. <https://doi.org/10.1111/exsy.12667>
- Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Orienta-Konsultit.
- Engwall, O., & B"alter, O. (2007). Pronunciation feedback from real and virtual languageteachers. *Computer Assisted Language Learning*, 20(3), 235–262.
- Du, J.; & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, 6. Article: 100230. <https://doi.org/10.1016/j.caeai.2024.100230>
- Fathi, J., Rahimi, M; & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254, <https://doi.org/10.1016/j.system.2024.103254>
- Fryer, L., & Carpenter, R. (2006). Emerging technologies: Bots as language learning tools. *Language, Learning and Technology*, 10(3), 8–14. Retrieved from <http://llt.msu.edu/vol10num3/emerging/>
- Hakim, R., & Rima, R. (2022). Chatting with AI Chatbots applications to improve English communication skill. *Journal of English Language Studies*, 7(1), 121–130. <https://doi.org/10.30870/jels.v7i1.14327>
- Haristiani, N. (2019). Artificial intelligence (AI) chatbot as language learning medium: An inquiry. *International Conference on Education, Science and Technology*, 1387(1), 012020. <https://doi.org/10.1088/1742-6596/1387/1/012020>
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning - are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237–257. <https://doi.org/10.1111/jcal.12610>
- Hwang, W. Y., Guo, B. C., Hoang, A., Chang, C. C., & Wu, N. T. (2024). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning*, 37(7), 1632–1658. <https://doi.org/10.1080/09588221.2022.2095406>
- Jeon, J. (2021). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*, 36(7), 1338–1364. <https://doi.org/10.1080/09588221.2021.1987272>
- Kim, N. Y. (2016). Effects of voice chat on EFL learners' speaking ability according to proficiency levels. *Multimedia Assisted Language Learning*, 19(4), 63–88.

- Kim, N.-Y, Cha, Y., & Kim, H. -S. (2019). Future English Learning: Chatbots and Artificial Intelligence. *Multimedia-Assisted Language Learning*, 22, 32-53.
- Kim, H.-S., Cha, Y., & Kim, N. Y. (2021). Effects of AI voice chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics*, 21, 712734. <https://doi.org/10.15738/kjell.21..202108.712>
- Li, Q., & Chan, K. K. (2024). Test takers' attitudes of using exam-oriented mobile application as a tool to adapt in a high-stakes speaking test. *Education and Information Technologies*, 29(1), 219–237. <https://doi.org/10.1007/s10639-023-12297-0>
- Lu, C. H., Chiou, G. F., Day, M. Y., Ong, C. S., & Hsu, W. L. (2006). Using instant messaging to provide an intelligent learning environment. In *Proceedings of the international conference on intelligent tutoring systems* (pp. 575–583). Springer. https://doi.org/10.1007/11774303_57.
- Nguyen, T. A. (2025). Addressing English-speaking anxiety in Vietnamese initial teacher education: Beyond coping strategies. *Asian-Pacific Journal of Second and Foreign Language Education*, 10(1), 1-27. <https://doi.org/10.1186/s40862-025-00324-1>
- Nguyen, T. H. (2026). The Impact of AI Chatbot Integration on Vietnamese EFL Learners' Speaking Performance, Fluency, and Confidence. *Lang. Technol. Soc. Media*, 4, 1, 70–88. <https://doi.org/10.70211/ltsm.3026-7196.314>
- Nguyen, T. T., & Tong, T. T. N. (2024). Investigation into Difficulties in Public Speaking among English-majored Students at University of Phan Thiet. *International Journal of Language Instruction*, 3(1), 17- 30. <https://doi.org/10.54855/ijli.24312>
- Nguyen, B. D., & Trinh, N. A. (2025). The Causes of English Language Learning Anxiety among English Major Seniors at a Vietnamese University. *International Journal of Language Instruction*, 4(4), 66- 82. <https://doi.org/10.54855/ijli.25445>
- Peng, J. (2019). The roles of multimodal pedagogic effects and classroom environment in willingness to communicate in English. *System*, 82, 161–173. <https://doi.org/10.1016/j.system.2019.04.006>
- Pentina, I; Hancock, T; Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of Replika. *Computers in Human Behavior*, 140(2):107600. <https://doi.org/10.1016/j.chb.2022.107600>.
- Quyen, V. P., Nga, P. T. M., & Nguyen, H. T. (2018). Challenges to speaking skills encountered by English-majored students: A story of one Vietnamese university in the Mekong Delta. *Can Tho University Journal of Science*, 54(5), 38-44. <https://doi.org/10.22144/ctu.jen.2018.022>
- Schreiber, J., & Asner-Self, K. (2011). *Educational research*. John Wiley & Sons, Inc.
- Tai, T.-Y., & Chen, H. H.-J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485–1502.

- Thao, T. Q., & Nguyet, D. T. N. (2019). Four aspects of English speaking difficulties encountered by tertiary English-majored students. *Ho Chi Minh City Open University Journal of Science*, 9(2), 53-64.
<https://doi.org/10.46223/HCMCOUJS.soci.en.9.2.261.2019>
- Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, 19(1), 122–140.
<https://doi.org/10.1017/S1366728914000832>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University press.
- Walker, A., & White, G. (2013). *Technology enhanced language learning: Connecting theory and Practice-Oxford handbooks for language teachers*. Oxford University Press.
<https://doi.org/10.1093/elt/cct069>
- Wang, Z. (2014). Developing accuracy and fluency in spoken English of Chinese EFL learners. *English Language Teaching*, 7(2), 110-118.
<https://doi.org/10.5539/elt.v7n2p110>
- Xie, Y., Chen, Y., & Ryder, L. H. (2021). Effects of using mobile-based virtual reality on Chinese L2 students' oral proficiency. *Computer Assisted Language Learning*, 34(3), 225–245. <https://doi.org/10.1080/09588221.2019.1604551>
- Zhang, J., Liao, Q., Li, L., & Luo, J. (2026). The influence of natural language processing on EFL speaking skills: Investigating learner adaptability, language accuracy, and fluency. *Journal of Educational Computing Research*, 64(1), 59-91.
<https://doi.org/10.1177/07356331251377414>
- Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *Sage Open*, 13(3), Article 21582440231193818.
<https://doi.org/10.1177/21582440231193818>

Biodata

Dr Uyen Nu Thuy Nguyen has been teaching at the University of Foreign Language Studies, the University of Danang for 18 years. Her academic interest areas are online education, distance learning, e-learning, computer-assisted language learning, and teaching methodology. She has published articles in WoS- and Scopus-indexed journals and in journals in international indexes.

Appendix 1

QUESTIONNAIRE ON STUDENTS' PERCEPTIONS OF USING AN AI CHATBOT FOR ENGLISH SPEAKING PRACTICE

This questionnaire explores your experiences of using Replika for English speaking practice. It examines your perceived improvements in speaking skills and your views on the benefits and challenges of using Replika. The questionnaire takes approximately 10 minutes to complete. All responses will be kept confidential and used exclusively for research. Thank you for your valuable contribution.

I. GENERAL INFORMATION

1. Gender: Male ☐ Female ☐
2. How long have you been studying English?
 - a. 5-10 years b. 10 years c. Others:
3. How much time do you spend using Replika to practise English speaking outside class?
 - a. 30 minutes – 1 hour
 - b. 1- 2 hours
 - c. Others:

II. PERCEPTIONS OF USING AI CHATBOTS IN IMPROVING ENGLISH SPEAKING SKILLS AND PRACTICE

Please tick the response that best reflects your opinion for each statement.

No.	Items	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
FLUENCY AND COHERENCE						
1	The AI chatbot helps me form sentences more quickly.					
2	The AI chatbot makes my speech smoother and more natural.					
3	The AI chatbot helps me speak more fluently.					
4	The AI chatbot helps me organize my ideas better.					
5	The AI chatbot helps me expand my ideas in conversations.					
6	The AI chatbot helps me connect ideas more clearly.					
7	The AI chatbot helps me connect ideas more coherently.					
8	The AI chatbot encourages me to think more deeply about my responses.					
LEXICAL RESOURCE						
9	The AI chatbot teaches me new and useful words.					
10	The AI chatbot improves my vocabulary use in spoken English.					
11	The AI chatbot helps me use words in the right context.					
12	The AI chatbot helps me understand and use new words.					
13	The AI chatbot helps me remember and use words I have learned.					

GRAMMATICAL RANGE AND ACCURACY						
14	The AI chatbot helps me to understand sentence structure better.					
15	The AI chatbot helps me use better sentence structures.					
16	The AI chatbot helps me use grammar more accurately.					
PRONUNCIATION						
17	The AI chatbot helps me pronounce words more clearly					
18	I can mimic native-like pronunciation through using the chatbot.					
19	The AI chatbot helps me improve my stress and intonation.					
20	The pronunciation models used in the AI chatbots are easy to follow and imitate.					
OVERALL PERCEPTIONS						
21	I am motivated in using AI chatbots for learning English-speaking skills.					
22	Practising speaking with the AI chatbot is enjoyable.					
23	It is easy to access the AI chatbot.					
24	AI chatbots create an environment that feels similar to real-world conversations.					
25	The AI chatbot helps me feel less afraid of making mistakes.					
26	I practise English speaking skill more often when using the AI chatbot.					
27	The AI chatbot does not always give detailed feedback.					
28	The lack of human interaction makes learning less engaging.					
29	I need more guidance to use the AI chatbot effectively.					
30	The AI chatbot sometimes uses words or phrases that are not useful for me.					
31	I sometimes find the AI chatbot's responses difficult to understand.					
32	I sometimes have technical problems when using the AI chatbot.					

Appendix 2

INTERVIEW QUESTIONS

1. How do you think using Replika influenced your English speaking? Can you give some examples?
2. Which aspect of your speaking (e.g., fluency, coherence, vocabulary, grammar, or pronunciation) improved the most? Why?
3. What did you find most helpful about using Replika for speaking practice?
4. What challenges did you experience when using Replika? How did you deal with them?