Hue University Journal of Science: Social Sciences and Humanities
ISSN 2588-1213
Vol. 135, No. 6B, 2026, p.p. 185–214, **DOI**: 10.26459/hueunijssh.v135i6B.8206

DIMENSIONS OF LEARNER SATISFACTION WITH CHATBOT-HUMAN INTERACTION IN EFL SPEAKING PRACTICE

Ho Thi Hien[1,2], Nguyen Quang Ngoan[3], Nguyen Van Huy[1]

[1]University of Foreign Languages and International Studies, Hue University

57 Nguyen Khoa Chiem Str, Hue, Vietnam

[2]Industrial University of Ho Chi Minh City, Vietnam

[3]Quy Nhon University, Vietnam

* Correspondence to **Nguyen Van Huy** < nguyenvanhuy@hueuni.edu.vn >

(*Received: December 31,* 2025; *Accepted: March 05, 2026*)

Abstract: AI voice-based chatbots (AVCs) have been increasingly integrated in teaching and learning speaking due to their potential benefits for language learning. Current studies have mainly examined the impacts of AVCs on learners' speaking performance and affective factors. However, research on learner satisfaction with chatbot-integrated learning experiences remains scarce. This study aims to explore Vietnamese university students' satisfaction with chatbot-human interaction across three dimensions, including language experience (LEX), design experience (DEX), and user experience (UEX) as guided by the chatbot-human interaction satisfaction model (CHISM). A quantitative research design was conducted with 80 EFL students who engaged in conversations with ChatGPT as a speaking partner for over 8 weeks in a classroom setting. Data were collected through a questionnaire, and descriptive and inferential statistical analyses were employed using SPSS. The findings revealed that learners were highly satisfied with the chatbot-generated language, chatbot design, and overall learning experience. In addition, Pearson correlation results indicate a strong, positive association among the dimensions. The study contributes empirical evidence to the body of literature on learner satisfaction with chatbot-human interaction in speaking practice.

Keywords. chatbot-human interaction, AVCs, satisfaction, chatbot-integrated language learning, ChatGPT

1. Introduction

Advances in Generative artificial intelligence (GenAI) tools have increasingly created unbelievable opportunities for students to learn languages. One of those is to practice speaking skills with AI voice-based chatbots (hereafter, AVCs). An AVC refers to a computer program designed to simulate dialogues with human users over the Internet, with the implementation of

natural language processing and speech technologies (Adamopoulou & Moussiades, 2020; Yusuf et al., 2025). Research has shown that AVCs such as Google Assistant, ChatGPT, Replika, Andy English Bot, TalkPal AI, CallAnnie, and Mondly can act as speaking partners to help learners of English as a foreign language (EFL) improve speaking skills. Moreover, there has been a growing number of studies focusing on affective aspects in chatbot-human interaction, such as motivation (Huang & Mizumoto, 2025; Wang et al., 2025), anxiety (Azizimajd, 2023; Ding & Yusof, 2025), self-efficacy (Celik et al., 2025), and willingness to communicate (Fathi et al., 2024; Zhang, 2026). However, the studies on learner satisfaction with chatbot-human interaction remain scarce.

Learner satisfaction in education refers to a psychological state in which learners subjectively evaluate whether they get what they desired during the learning experience (Shee & Wang, 2008; Yu, 2022). Understanding learner satisfaction plays an important role in language education because it helps educators adjust instructional design to meet the needs of learners (Asakereh & Dehghannezhad, 2015; Elliott & Shin, 2002). Moreover, having engaged in a satisfying learning experience can motivate students to attend class, improve self- efficacy, and reduce the possibility of dropping out of school (Elliott & Shin, 2002; Gao, 2010). While researchers have studied learner satisfaction in traditional and online classroom environments (Abdullah et al., 2022; Asakereh & Dehghannezhad, 2015; Le, 2025; Qutob, 2018), learner satisfaction with chatbot-human interaction remains underexplored. To date, two studies have examined learners' voices regarding how satisfied they were with interacting with AVCs as conversational partners for practicing speaking skills (Belda-Medina & Calvo-Ferrer, 2022; Belda-Medina & Kokošková, 2023). Although the studies conducted interventions of chatbot-human interaction with Western university students having high levels of speaking proficiency, they showed inconsistent levels of learner satisfaction due to different uses of AVCs.

These studies exposed some contextual and design limitations. First, there has been no research on Vietnamese higher education students' satisfaction with chatbot-integrated speaking practice. In this context, while previous studies have explored the effectiveness of chatbot-assisted speaking activities in improving university students' speaking skills (Duong & Suppasetseree, 2024; H. H. Nguyen, 2025; T. H. Nguyen, 2025), little attention has been paid to learner satisfaction with such experiences. Without listening to learners' voices, it was unclear to teachers how to refine the lessons and address the barriers hindering the effectiveness of chatbot-human interaction. Second, the existing studies on learner satisfaction with chatbot-human interaction incorporate multiple AVCs within a single intervention (Belda-Medina & Calvo-Ferrer, 2022; Belda-Medina & Kokošková, 2023). This diversity of AVCs might confuse learners due to variations in chatbot designs and functionality. Third, none of those studies have used ChatGPT as the AVC in the interventions, although ChatGPT is widely regarded as one of the most modern Chatbots at present (Kohnke et al., 2023). Especially, ChatGPT -4o

offers nine advanced voice modes with various "life-like" tones and characters (OpenAI, 2025). These gaps in the literature highlight a need for implementing research to explore Vietnamese university students' satisfaction with ChatGPT-integrated speaking experiences and how those dimensions interrelate.

This study aims to investigate students' satisfaction with chatbot-human interaction in an EFL speaking context based on the multidimensional framework, Chatbot-human interaction satisfaction model (CHISM) (Belda-Medina & Kokošková, 2023). In the light of CHISM, the research attempts to explore learner satisfaction with conversing with an AVC in three dimensions, including language experience (LEX), design experience (DEX), and user experience (UEX). Therefore, it seeks to answer two research questions:

RQ 1. What are students' levels of satisfaction with chatbot-human interaction across three dimensions: LEX, DEX, and UEX?

RQ2. What are the relationships among LEX, DEX, and UEX in chatbot-human interaction?

This study will provide important contributions to the body of chatbot literature as well as teaching practice. First, letting the students voice themselves after engaging in eight sessions of chatbot-human interaction helps chatbot designers understand what is and is not going well with their chatbot's current designs. These insights help chatbot designers identify effective features and areas for improvement. Thereby, it informs refinement to the AVC designs to better meet the needs of users. Second, understanding learners' feelings and perceptions offers actionable plans for language teachers to modify teaching plans and activity designs to enhance the effectiveness of chatbot-integrated speaking activities. Finally, reflections from the learners can provide empirical evidence to debates about the realistic affordances of chatbots for speaking practice.

2. Literature review

2.1. Previous studies on using AVCs in teaching and learning speaking skills

In recent years, there has been a growing body of research on the potential of conversing with AVCs as speaking partners. In terms of research contexts, previous studies have implemented AVC-supported speaking activities across a wide range of educational levels, including primary schools (e.g., Jeon, 2024), high schools (e.g., Ye et al., 2021) and universities(e.g., El Shazly, 2021; Fathi et al., 2024; Zou et al., 2025). This indicates that students of different ages can engage in chatbot-integrated speaking practice; however, university students were the most frequently selected participants, possibly because they have greater access to smartphones and are generally more familiar with digital technologies. This growing

popularity of conversing with AVCs highlights the potential of integrating chatbot-integrated speaking activities into formal language instructions.

Regarding instructional design, the literature shows considerable diversity across learning settings, speaking activities, and types of chatbots. While many chatbot-integrated speaking activities were conducted within a classroom setting (e.g., Azizimajd, 2023; Duong & Suppasetseree, 2024; Jeon, 2024), others took place outside the classroom (e.g. Belda-Medina & Calvo-Ferrer, 2022; Belda-Medina & Kokošková, 2023; Celik et al., 2025; Yildiz, 2024). A small number of studies adopted a blended approach that combined in-class and out-of-class chatbot interaction (e.g., Çakmak, 2022; Fathi et al., 2024). About the teaching activities, the prior studies indicate a wide variety of chatbot-integrated speaking activities, such as role-plays (El Shazly, 2021), small talks (Hsu et al., 2023), shadowing and imitation (Duong & Suppasetseree, 2024), interactive storytelling (Yuan, 2024) or guessing games (Hsu et al., 2023), or information gap (Jeon, 2024). This diversity of learning settings and speaking activities reflects flexible pedagogical uses of AVCs in speaking instructions, ranging from teacher-guided to self-directed speaking practice. It also shows a growing trend of chatbot-human interaction, offering language learners opportunities to engage in meaningful spoken communication. The body of chatbot literature observes a wide range of chatbot types implemented in previous empirical studies. The most popular ones were *Replika* (Belda-Medina & Calvo-Ferrer, 2022; Çakmak, 2022), *Andy* (Duong & Suppasetseree, 2024; Fathi et al., 2024), *ChatGPT* (Celik et al., 2025; Muniandy & Selvanathan, 2025), *Mondly* (Belda-Medina & Kokošková, 2023; El Shazly, 2021), *Google Assistant* (Eragamreddy, 2025), and *Alexa* (Hsu et al., 2023).

ChatGPT has been increasingly implemented as a conversational partner for speaking practice in previous studies (e.g. Agustina et al., 2025; Almanea, 2024; Celik et al., 2025; Muniandy & Selvanathan, 2025). The results were consistent in ChatGPT's potential to foster learners' willingness to communicate, reduce anxiety, and increase confidence (Agustina et al., 2025; Almanea, 2024). Learners also reported the benefits of immediate interaction, low-pressure practice, and perceived conversational support, which contribute to higher engagement and sustained use (Celik et al., 2025; Kohnke et al., 2023; Muniandy & Selvanathan, 2025).

Regarding the possibility of facilitating learners' speaking skills by using AVCs, the existing studies have consistent findings on AVCs' positive effects on enhancing learners' speaking skills in terms of fluency (Ding & Yusof, 2025; Eragamreddy, 2025) and expand lexical resources (Grab, 2025; Zou et al., 2025). In addition, interacting with AVCs brings affective benefits for EFL learners. For instance, AVCs are perceived as creating a low-threat, non-judgmental practice environment that can reduce learners' anxiety during conversations (Ding & Yusof, 2025; Du & Daniel, 2024), increase willingness to communicate(Ding & Yusof, 2025), provide guidance and feedback at any time and any place (Lyu et al., 2025). Therefore, learners

can practice speaking without worrying about negative evaluation, judgment, or embarrassment when they interact with a human interlocutor.

Despite a growing number of studies on learning outcomes and affective aspects, studies on learner satisfaction with chatbot-human interaction remain limited. This highlights the need for further empirical research on the extent to which learners are satisfied with the learning experience.

2. 3. Learner satisfaction with chatbot-integrated speaking learning

Measuring learner satisfaction is important in language education. It helps language teachers pinpoint strengths and areas for improvement, so that necessary refinement can be made to make the lessons meet the needs of learners (Asakereh & Dehghannezhad, 2015). To illustrate, collecting learner satisfaction could be considered as one of the key learning outcomes in Western institutions in the USA and the UK (Li et al., 2016). Previous studies have explored learner satisfaction in various fields of language learning, such as online learning (Chen et al., 2020; Yılmaz, 2023), MOOC courses (Gameel, 2017; Le, 2025), and traditional speaking classes (Asakereh & Dehghannezhad, 2015; Qutob, 2018).

Regarding learner satisfaction with chatbot-human interaction in language education, some of the earliest empirical research was conducted by Belda-Medina and Calvo-Ferrer (2022), Belda-Medina and Kokošková (2023), and Aleedy and colleagues (2025). In the first study (Belda-Medina & Calvo-Ferrer, 2022), there were 176 university students from Spain and Poland engaging in chatbot-integrated practice with Replika, Kuki, and Wysa chatbots. These chatbots were featured in their customizable avatars. Students had four weeks to practice speaking with the chatbots autonomously after class hours. The participants reported high satisfaction with the chatbot's avatar design, language experience, particularly with Replika. Although this study included students from two countries to demonstrate good ecological validity, it included learners at high proficiency levels (B2-C1) whose learning desires and purposes might differ from those in EFL educational contexts. Moreover, the chatbot-integrated practices were conducted at home, making it difficult for teachers and researchers to monitor participants' participation.

Another study was conducted with 237 students in two universities in Spain and the Czech Republic (Belda-Medina & Kokošková, 2023). The participants had four weeks to engage in chatbot-integrated speaking activities with four App-integrated chatbots: Mondly, Andy, John Bot, and Buddy.ai. These chatbots do not have featured avatars. Participants' proficiency levels ranged from B2 to C2. They were required to practice speaking with chatbots for at least three hours per week at home. The results indicated low to moderate satisfaction levels with chatbot-human interaction. Specifically, they were most satisfied with the game-based elements and the chatbot's user-friendly interface. However, they were dissatisfied with the chatbot's

ability to detect and correct language errors, semantic coherence, and speech recognition. It highlights a need for better adaptation to learner needs and improved speech technologies of the selected chatbot. One important weakness of this study is the inappropriate choice of chatbots for the students' levels and ages. While participants are very fluent speakers, Buddy.ai was originally designed for kids aged 4-10, and John Bot prioritized the translation option.

The last study was conducted in Saudi Arabia with 40 university students and four faculty members (Aleedy et al., 2025). They used an AI-powered multi-agent Translation Learning Chatbot for translation practice. Although this study did not focus on oral practice between learners and chatbots, it explored learner satisfaction with chatbot-human interaction along the same dimensions as the aforementioned chatbots. The findings showed that participants were highly satisfied with the learning experience, especially with the LEX dimension.

The reviewed studies share several common features, including a focus on university students and a relatively short invention duration. Two studies (Aleedy et al., 2025; Belda-Medina & Calvo-Ferrer, 2022) showed consistently high satisfaction with chatbot-human interaction, whereas another research study (Belda-Medina & Kokošková, 2023) reported contrasting results. After reviewing the previous studies, some notable research gaps were found. Firstly, to date, no study on learner satisfaction with chatbot-human interaction has been conducted in the Vietnamese higher education context, where students' proficiency levels are generally lower, yet their need for fluency in speaking is high. Vietnamese educators have taken the very first steps to integrate AVCs into speaking instruction to address the long-standing pedagogical challenges. It is important for them to listen to learners' voices and reflections to adjust instructional design where necessary. Secondly, the prior research (Belda-Medina & Calvo-Ferrer, 2022; Belda-Medina & Kokošková, 2023) was conducted primarily in out-of-class contexts, making it hard for teachers and researchers to monitor learners' engagement and ensure that the practice was implemented as intended. Finally, incorporating more than one AVC into the same intervention might confuse learners, as they tend to use only one or two preferred AVCs and ignore others. This unequal use might affect the reliability of the findings. Taken all together, our study will address these gaps by conducting an empirical study in the Vietnamese university context with in-class practice sessions between learners and ChatGPT as the only AVC.

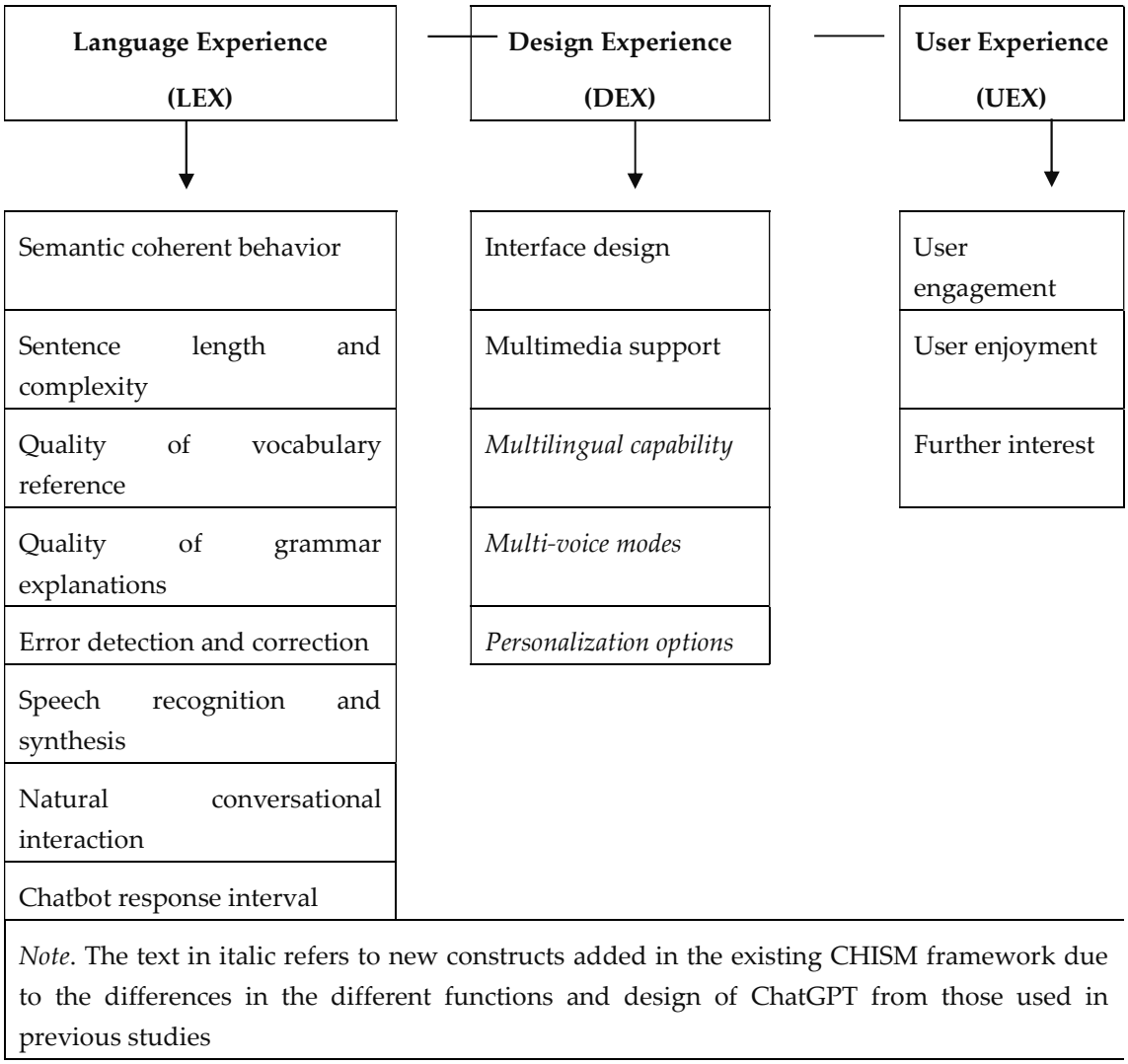
2.3. Theoretical framework

This study adapts the Chatbot-human interaction satisfaction model (CHISM) (Belda-Medina & Kokošková, 2023) as the theoretical background to explore learner satisfaction with their learning experience. CHISM was first introduced in 2022 (Belda-Medina & Calvo-Ferrer, 2022) to provide a theoretical framework for measuring learner satisfaction with chatbot-human interaction. This framework was originally used to evaluate students' satisfaction with

three web-based Conversational AI agents (i.e., Kuki, Wysa, and Replika) that feature avatars (Belda-Medina & Calvo-Ferrer, 2022). It was then adapted to better fit with the study with four different App-Integrated Chatbots without an avatar (i.e., Mondly, Andy, John Bot, and Buddy.ai) (Belda-Medina & Kokošková, 2023). Despite its early development, CHISM has been recognized as a necessary framework in researching chatbot-human interaction (Koç & Savaş, 2025). Unlike the existing models, CHISM's multidimensional features capture learners' reflections on their interaction with AVCs in three important perspectives, language experience (LEX), design experience (DEX), and user experience (UEX). In the existing studies that have implemented CHISM as the theoretical background (Aleedy et al., 2025; Belda-Medina & Calvo-Ferrer, 2022; Belda-Medina & Kokošková, 2023), all measurement items were reported to have satisfactory internal consistency. Therefore, it can serve as a reliable theoretical framework for exploring chatbot-human interaction in the following studies.

Our research based on the latter version of CHISM (Belda-Medina & Kokošková, 2023) with some minor adaptations to capture additional aspects related to the ChatGPT's design (See Figure 1). To illustrate, in the LEX dimension, the "*non-verbal language*" was excluded because the current ChatGPT interface does not feature an avatar or any non-verbal cues during conversations. Secondly, in the DEX dimension, "*game-based learning*" was also removed for the same reason as above. Thirdly, we incorporated several design features of ChatGPT under the DEX label that are pedagogically relevant to EFL speaking practice. *Multilingual capability* was included because ChatGPT can speak different languages. This feature allows students to use their mother tongue to communicate when they are unable to articulate a word, a phrase, or an utterance in the target language without interrupting the conversation flow. This design feature is important for chatbot-human interaction because it helps students translate their ideas into the target language so as to keep the conversation going smoothly. Moreover, we included *multi-voice mode* as ChatGPT 4.o offers nine different advanced voice options with varying tones and moods, which resemble human speech. Another aspect added in this adapted version was *personalization option*, which allows users to provide it with their preferred names, current proficiency level, target language goals, and any personalized information that users want chatbot to refer to during the conversations. Based on this information, ChatGPT can adjust its language complexity, conversational style, and feedback to ensure comprehensible input tailored to individual learners' needs.

Figure 1. The adapted version of the chatbot-human interaction satisfaction model (CHISM)



Language experience (LEX)

To explore learner satisfaction with chatbot-generated language in chatbot-human interaction, we adopted the definitions of the constructs operationalized in CHISM (Belda-Medina & Kokošková, 2023). Firstly, *semantically coherent behavior* means the chatbot's capability to maintain a contextually relevant and meaningful dialogue. Secondly, s*entence length and complexity* refers to the chatbot's use of long and complex sentences in responses. Thirdly, the *quality of vocabulary reference* is defined as the chatbot's use of in-context lexical resources (synonyms, collocations). Fourthly, the *quality of grammar explanations* means the chatbot's ability to understand grammar rules. Fifthly, *error detection and correction* refer to the chatbot's ability to detect users' language errors and provide feedback. Sixthly, s*peech*

recognition and synthesis pertain to the chatbot's ability to understand users' language input and convert spoken language into written text. Next, *natural conversational interaction* refers to the chatbot's ability to engage and maintain the naturalness of conversations with humans, such as adapting language style, levels, and tones to suit the learner's preferences. Finally, the *response interval* is the time it takes a chatbot to generate a response during a conversation. These aspects cover vital abilities that an interlocutor needs to have in any dialogue, including not only language output, but also a feedback mechanism and interactional skills.

Design experience (DEX)

The adapted version of CHISM focuses on collecting learners' evaluations on five design features of ChatGPT as a conversational agent. First, *interface design* concerns the overall look and feel that guides users in interacting with a chatbot, including how to start and end a conversation, how to upload files, how messages are displayed, and how users enter responses. Second, *multimedia* support refers to a chatbot's ability to read different types of files, such as images and documents, and search for information from sources. Thirdly, *multilingual capability* means a chatbot's ability to speak and understand multiple languages during the same conversation, which allows both learners and the chatbot to switch from the target language to the mother tongue or any other languages. Fourthly, multi-voice modes refer to the availability of multiple voice options (i.e., nine advanced voices in ChatGPT 4.o) with various tones and emotional expressions. Lastly, *personalization* refers to chatbots' ability to receive and remember users' names, current proficiency level, learning objectives, and preferences.

User experience (UEX)

Our study adopts three aspects in this dimension to understand users' reflections on their own experience. Each aspect is operationalized as in the original version of CHISM. First, *user engagement* involves the user's initiation of interaction and maintaining the conversation with the chatbot. Second, *user enjoyment* refers to the pleasure that learners show as interacting with the AVC. Finally, *users' further interest* pertains to the levels of willingness to use AVCs in the future.

3. Methodology

3.1. Research design and context

This study adopted a quantitative research approach. We employed both descriptive and correlational analysis to answer the research questions. It was conducted in the Vietnamese higher education context, where students have a strong demand for studying English-speaking skills for future jobs. The research was supported by institutional heads and deans, who provided an approved gatekeeper letter. To facilitate the chatbot-human interaction, the

speaking lessons took place in a modern lab with well-equipped devices for technology-enhanced language learning activities, such as internet-connected computers, headphones, microphones, and software to connect from the teacher's computer to the students' so that the teacher could easily share the screen and demonstrate or provide task guidance when applicable.

3.2. Participants

The participants in this study were selected using a nonrandom sample. They were English major sophomores undertaking the Speaking 2 course from 2 intact classes. That is, before the semester began, students self-registered for certain classes. The two classes were selected based on convenience and accessibility because researcher 1 is a teacher in the research context. They followed the same curriculum, same textbook and their schedule was in the same semester. On the one hand, conducting research with such participants remained the norm in a real teaching context in Vietnamese higher education, thereby limiting the interference of any unnatural intervention (Cohen et al., 2018). On the other hand, the use of nonrandom, intact-class raises methodological limitation concerns as the results might not fully generalizable to other instructional contexts. Therefore, the findings would be interpreted with caution. To control the interference of potential confounders, these two classes had identical backgrounds. For example, both classes were taught by the same instructor, same length of class hours, in the same room, received the same technical training in the first week, followed the same syllabus and teaching procedure. These to avoid the differences in teacher's teaching style, instructional variables.

There were 80 students (40 students in each class) from a university in Vietnam, including 22 males (27.5 %) and 58 females (72.5%). The mean age of the participants was 19.23. Their speaking proficiency was reported at the beginning of the semester to be mostly A2 to B1, with a few at B2. The students reported that they have learnt English for more than five years. Regarding prior experience with practicing speaking skills using ChatGPT, 48 participants (60%) had never used it before, 30 (37.5%) had tried it once or twice, and only 2 (2.5%) had used it multiple times, limited to the free version with a standard, robotic-sounding voice. The results indicate the novelty of experiencing chatbot-human interaction as a method to practice speaking skills. To examine students' technological competence in using the required skills and techniques for ChatGPT voice chats and the OBS screen-recording application, as introduced in the first week, we developed eight items asking students to self-report how easily they could perform these tasks. As shown in Table 1, students were highly familiar with other necessary technological skills for chatbot-human interaction implementation, with mean scores ranging from 3.88 to 4.51. It means that technological competence would not be a barrier for students to engage in chatbot-integrated speaking learning.

Table 1. Participants' technological competence (N=80)

	Mean	Std. Deviation
PoE1. I can easily log in to ChatGPT.	3.90	.880
PoE2. I can easily start and stop a conversation with ChatGPT.	4.34	.841
PoE3. I can easily select an advanced voice to chat with easily	4.26	.978
PoE4. I can easily upload a handout file to ChatGPT.	4.50	.763
PoE5. I can easily enter a prompt before starting the conversation with ChatGPT.	4.14	.896
PoE6. I can easily maintain the dialogue with ChatGPT.	3.91	.903
PoE7. I can easily control the dialogue with ChatGPT.	3.88	.919
PoE8. I can easily use OBS application for screen recording.	4.51	.857

3.3. Data Collection Instruments

GPT Talk Time project

To explore learner satisfaction with chatbot-human interaction, students were engaged in 8 speaking sessions with ChatGPT Plus version in the speaking class. The chatbot-integrated speaking activities were designed in accordance with the course syllabus. The lessons were designed and organized following the communicative language teaching (CLT) approach that emphasized meaningful interaction and the development of communicative competence through authentic language use (Richards & Rodgers, 2014). Within the CLT approach, Think–Pair–Share (TPS) is a teaching technique that has been popularly used by language teachers, which specifies the learning process into three phases. Students work individually to prepare ideas or understand the task, then peer discussion, and finally, whole-class communication (Sharma & Saarsar, 2018). Traditionally, each student pairs up with a peer who sits next to him or her. Previous studies found pedagogical benefits of this technique in teaching and learning language. For example, TPS increases students' opportunities to produce spoken language, reduces speaking anxiety, and supports fluency development by allowing learners to rehearse their ideas in pairs before public performance (Apriyanti & Ayu, 2020; Raba, 2017; Sharma & Saarsar, 2018). Our study adopted the TPS procedure, but chatbot-human discussion was used instead of human-human discussion in conventional classes.

All eight sessions were organized following the same structure: (1) Warm-up; (2) Language preparation; (3) TPS; (4) Feedback and consolidation (See Figure 3). During the pair

phase, students had 20-30 minutes to have conversations with ChatGPT instead of a human partner about the assigned tasks as shown in Table 2.

Figure 3. Lesson procedure

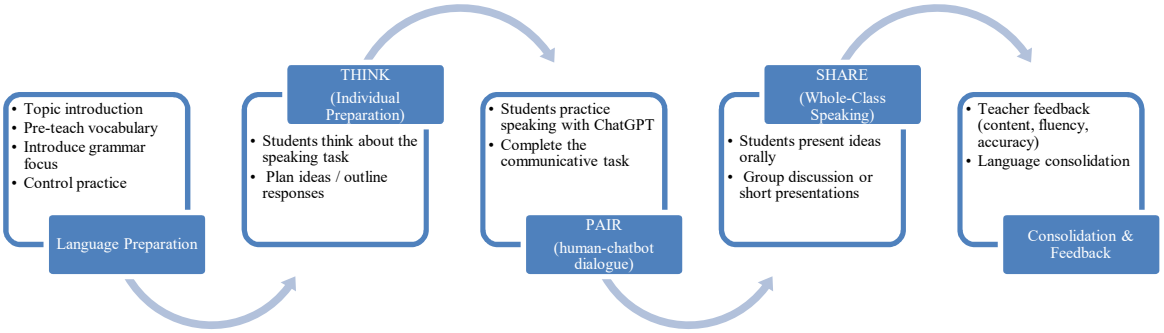


Table 2. Topics of the Chatbot-human sessions

Week	Topic	Activities
1	Graduation	Discussion: Reasons preventing students from graduating
2	Travelling	Information exchange: Mystery trip
3	Weather	Information exchange: Weather in my city
4	Restaurant	Information exchange: Ordering food at the restaurant
5	Eating habits	Information exchange: Healthy and unhealthy eating habits
6	Saving environment	Discussion: Benefits and challenges of eating green
7	Locations	Discussion: What makes a good neighborhood?
8	Looking back and looking forward	Discussion: strengths and weaknesses of chatbot-human interaction

Questionnaire

Data were collected from a questionnaire informed by the adapted CHISM framework (Belda-Medina & Kokošková, 2023). A questionnaire is a commonly used instrument in social science studies that offers advantages for researchers (Dörnyei & Dewaele, 2022). For example, it allows researchers to collect a large number of responses in a short time, and the data can be processed and analyzed quickly and relatively straightforwardly. The original version of the

CHISM-based questionnaire included 16 items, but the adapted version used in this study includes 23 items. The reason for this modification is that some items were double-barreled, which asked two different constructs in a single question. For example, "*Chatbot's error detection and correction*" should be split into "*chatbot's error detection*" and "*chatbot's error correction*" (See Appendix 1). According to Cohen et al. (2018), double-barreled questions should be avoided because they can be confusing for respondents, as they may partly agree with the item. In other words, each question should address one and only one issue. Based on the CHISM framework, the questionnaire was designed in two main parts: (1) demographic information, and (2) Satisfaction with chatbot-human interaction.

Part 1 included 12 items that collected participants' demographic information, such as gender, age, English learning experience, experience with voice chatting with ChatGPT, and technological competence levels. Part 2 consisted of 23 items corresponding to the three dimensions categorized by CHISM: LEX (14 items from questions 1-14), DEX (5 items from questions 15-19), and UEX (4 items from questions 20-23). The items were designed in a 5-point scale format (1=Very dissatisfied - 5=Very satisfied) and for the UEX dimension (1 = Strongly agree - 5= Strongly agree) (see Appendix 1).

Following the guidance of Dörnyei and Dewaele (2022), several steps were taken to ensure the validity and reliability of the questionnaire as a self-reported measure. First, the content validity was ensured because it was based on an established framework, CHISM (Belda-Medina & Kokošková, 2023). After modifying some items (removing the non-verbal language and splitting the double-barreled items), the questionnaire was reviewed by two experts in the field who hold a PhD degree. The specialists evaluated the relevance, clarity, and representativeness of the items. Upon receiving the expert reviews, the questionnaire was revised and piloted with 15 students of the same level but not participated in the two speaking classes. The pilot participants were requested to check if the items were clear, easy to understand, and to estimate the approximate time to complete. Then, the result of the pilot test was analyzed to modify the wording confusion and length of the items. After refinement, it was translated into Vietnamese for the target participants to understand easily, then back into English. These backward translations, as well as the final version of the questionnaire were validated by researchers two and three.

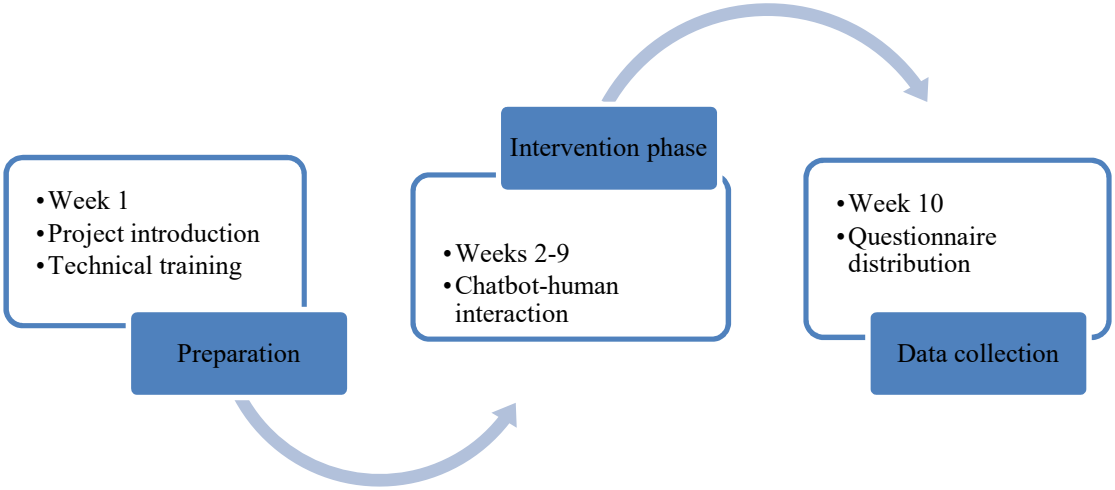
Moreover, to ensure the construct validity, we evaluated the convergent and factorial validity of the questionnaire by conducting a confirmatory factor analysis (CFA) test using AMOS. The CFA results showed a moderate fit to the data (*CFI = .81, TLI = .79, RMSEA = .12, SRMR = .08*). Some fit indices did not fully meet recommended thresholds; this may be attributable to the small sample size (*N = 80*) and model complexity. Therefore, the results were considered acceptable for exploratory validation in the context of the present study.

To check the internal consistency of the questionnaire, we calculated the Cronbach's alpha coefficient for each dimension of the CHISM. The results indicated good reliability for LEX (α = .887), DEX (α =.864), and UEX (α =.917) (DeVellis & Thorpe, 2021). Although two LEX items (LEX 3 and LEX 4) showed low corrected item-total correlation (.225 and .206, respectively, which were below .30), removing them resulted in only a slight increase in alpha value (.896 and .899). Moreover, these two items are relevant to evaluating students' satisfaction with LEX; thus, they were retained in the scale. The questionnaire was administered online using Google Forms.

3.4. Data collection and analysis

The data collection procedure spanned 10 weeks, as shown in Figure 2.

Figure 2. Data collection procedure



For data analysis of research question 1, aiming to explore learner satisfaction with LEX, DEX, and UEX, we conducted a descriptive analysis to find means and standard deviations. As for research question 2, we explored the association among the dimensions of satisfaction in three steps. First, data of each dimension was computed to form 3 new variations, namely LEX_Mean, DEX_Mean, and UEX_Mean. Then, check the normal distribution of these new variables to select the appropriate test of correlation by calculating the Shapiro–Wilk. Finally, based on the results of the normality test, we selected a test to explore the correlation coefficient among the satisfaction dimension. If the statistical significance confirms a normal distribution of data, Pearson's correlation will be applied; in contrast, Spearman's correlation should be used for non-normally distributed data (Field, 2009).

4. Results

4.1. Students' satisfaction with language experience (LEX)

Table 3 shows learner satisfaction with ChatGPT-generated language in chatbot-human conversations (*N=80*). The data revealed moderate to high levels of satisfaction with the LEX dimension of chatbot-human interaction, with a mean score spanning from 3.46 to 4.34. The most satisfying aspect was *ChatGPT's response speed* (*M=4.34, SD=.71*), followed by the AVC's ability to produce *a wide range of vocabulary* (*M=4.21, SD=.74*), *detect learners' linguistic errors* (*M=4.21, SD=.87*). ChatGPT's language production also received high satisfactory ratings regarding its on-topic responses (*M=4.20, SD=.64*), meaningful responses (M = 4.16, SD = .67), and topic-appropriate vocabulary (*M = 4.17, SD = .74*).

In terms of feedback quality, learners expressed positive evaluation for ChatGPT's capability to provide corrective feedback (*M = 4.12, SD = .93*) and the accuracy of grammar explanations (*M = 4.04, SD = .83*). Moreover, ChatGPT was also perceived to be able to keep the conversation flow natural and smooth. In contrast, students reported lower satisfaction with ChatGPT's use of sentence length (*M = 3.67, SD = .88*) and grammar complexity (*M = 3.46, SD = .98*), which were just above the midpoint of the 5-point Likert scale.

Table 3 . Learner satisfaction with LEX dimension (N=80)

	Minimum	Maximum	Mean	Std. Deviation
ChatGPT's on-topic responses	3	5	4.20	.644
ChatGPT's relevant responses	2	5	4.16	.665
ChatGPT's sentence length	2	5	3.67	.883
ChatGPT's grammar complexity	1	5	3.46	.980
ChatGPT's vocabulary range	2	5	4.21	.741
ChatGPT's topic-appropriate vocabulary	3	5	4.17	.742
ChatGPT's grammar explanation accuracy	2	5	4.04	.834
ChatGPT's grammar explanation clarity	2	5	3.82	.938
ChatGPT's error detection	1	5	4.21	.867
ChatGPT's corrective feedback	1	5	4.12	.933
ChatGPT's voice recognition	2	5	3.84	.934
ChatGPT's voice synthesis	2	5	3.94	.959

ChatGPT's ability to follow conversational natural flow	2	5	4.04	.818
ChatGPT's response speed	2	5	4.34	.711

4.2. Students' satisfaction with DEX

Table 4 reports learners' evaluation of the DEX dimension of ChatGPT as a speaking partner (*N=80*). Overall, the students were highly satisfied with the design of this AVC, with a mean score ranging from 3.96 to 4.39. The most satisfying factor was the usability of ChatGPT's *interface* (*M = 4.39, SD = .70*), followed by variety of *voices* available in ChatGPT (*M = 4.24, SD = .77*) and *customized functions* (*M = 4.19, SD = .77*). Its *multimedia features* and ability to change across the *languages* received slightly lower satisfaction from the learners (*M = 3.96, SD = .99*; and *M = 4.10, SD = .87*, respectively).

Table 4. Learner satisfaction with DEX dimension (N=80)

	Minimum	Maximum	Mean	Std. Deviation
ChatGPT's interface usability	3	5	4.39	.703
ChatGPT's multimedia features	1	5	3.96	.987
ChatGPT's ability to switch languages	2	5	4.10	.866
ChatGPT's multi-voice modes	3	5	4.24	.767
ChatGPT's personalization functions	2	5	4.19	.765

4.3. Students' satisfaction with UEX

Data in Table 5 presents learners' positive evaluation with chatbot-human interaction in the UEX dimension (*N=80*). The results indicate that students highly evaluate the capability and possibility of using ChatGPT as a conversational partner to practice speaking skills, with mean scores ranging from 4.13 to 4.43. Students reported strong agreement with ChatGPT' capability to effectively help them maintain interest during the long discussions (*M = 4.13, SD = .70*) and encourage them to continue speaking (*M = 4.20, SD = .80*), and enjoy practicing speaking (*M = 4.20, SD = .72*). Most notably, the highest mean score was observed in learners' intention of practicing speaking with ChatGPT after the project (*M = 4.43, SD = .59*).

Table 5. Learner satisfaction with UEX dimension (N=80)

	Minimum	Maximum	Mean	Std. Deviation
ChatGPT helps me maintain interest during	3	5	4.13	.700

conversations				
ChatGPT encourages me to continue speaking	2	5	4.20	.802
I enjoy practicing speaking with ChatGPT.	2	5	4.20	.719
I will continue practicing speaking with ChatGPT in the future	3	5	4.43	.591

4.4. Correlation among LEX, DEX, and UEX

We started analyzing the correlational relationship among the dimensions by checking if the data were normally distributed using the Shapiro–Wilk. The normality test reveals that while LEX was normally distributed (p =.126, which is greater than .05), DEX and UEX were not (p =.000, which is less than .05) (see Table 6). As for implementing Likert-scale measures with a moderate sample size ($N = 80$) and composite scores, Pearson's correlation was used to examine associations among the variables, with Spearman's correlation conducted as a confirmatory test (Field, 2009).

Table 6. Results of normal distribution test (N=80)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
LEX_mean	.079	80	.200*	.975	80	.126
DEX_mean	.143	80	.000	.909	80	.000
UEX_mean	.173	80	.000	.884	80	.000

Table 7 presents the results of the Pearson correlation analysis ($N=80$). The results revealed a strong, positive relationship among three dimensions of learner satisfaction with chatbot-human interaction, with coefficients ranging from .693 to .798. All of these are statistically significant (p<.001). Specifically, learner satisfaction with LEX was strongly correlated with that of DEX (r = .719, p < .001) and UEX (r = .693, p < .001). DEX satisfaction was also found to be strongly positively correlated with UEX (r = .798, p < .001). According to Field (2009), a correlation coefficient greater than .50 indicates a strong effect size. These findings mean that learners who reflected a higher satisfactory rating with chatbot-generated language tended to be more satisfied with the chatbot design and held positive learning attitudes toward chatbot-human interaction.

Spearman's rank-order correlation was conducted as a robustness check due to minor deviations from normality, yielding a highly similar pattern of results. The consistency between Pearson (parametric) and Spearman (nonparametric) analyses suggests that learners' perceptions of language experience, design experience, and user experience are strongly and reliably interconnected in AI-mediated speaking practice.

Table 7. Results of Pearson correlation test (N=80)

		LEX_mean	DEX_mean	UEX_mean
LEX_mean	Pearson Correlation	1	.719**	.693**
	Sig. (2-tailed)		.000	.000
DEX_mean	Pearson Correlation	.719**	1	.798**
	Sig. (2-tailed)	.000		.000
UEX_mean	Pearson Correlation	.693**	.798**	1
	Sig. (2-tailed)	.000	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

5. Discussion

As guided by the CHISM framework, this study replicated previous studies in an EFL university context where students' speaking proficiency is relatively lower, but the desire and need for being a fluent speaker is greater. It also addressed limitations in the prior studies by a longer intervention (i.e., eight weeks rather than four weeks as previously) and in-class practice with solely ChatGPT, one of the best AVCs to date. Moreover, it added to the previous results by exploring the correlation among the three satisfaction dimensions LEX, DEX, and UEX, which was not the focus of the previous studies. This section discusses the results in the order of the two research questions.

Research question 1: What are students' levels of satisfaction with chatbot-human interaction across three dimensions, LEX, DEX, and UEX?

The quantitative results descriptively showed moderate to high satisfaction with three dimensions of chatbot-human interaction. Regarding language-related aspects, participants were found to be highly satisfied with ChatGPT's lexical range, error detection and correction response speed, and relevant responses, with mean scores reaching above 4.0 out of 5. These

results suggested that ChatGPT was perceived as an potential speaking partner, especially in terms of linguistic support and interactional efficiency, which is consistent with previous findings (Agustina et al., 2025; Almanea, 2024; Belda-Medina & Calvo-Ferrer, 2022; Celik et al., 2025; Kohnke et al., 2023; Muniandy & Selvanathan, 2025).

The results also revealed that students were only moderately satisfied with ChatGPT's use of long and complex sentences and voice recognition technology. This finding is consistent with previous chatbot studies indicating that excessive linguistic complexity can reduce comprehensibility and hinder interaction, particularly for lower- to intermediate-level EFL learners (Belda-Medina & Kokošková, 2023). Vietnamese university students' biggest problems in speaking skills are grammar range and pronunciation (Tran & Khau, 2024). With a limited range of grammar structures, it might be challenging for them to understand high-level structures. Also, students' inaccurate pronunciation might result in the chatbot's inability to recognize the sounds and give relevant responses. Thus, it could be inferred that prompts should be continuously revised and improved, learners should add more details, and more personalization options should be added to ChatGPT to adjust its grammar levels, as suggested in previous research (Derakhshan & Ghiasvand, 2024). They also need to speak more slowly and accurately for ChatGPT to recognize the speech, so that the quality of conversation can be improved.

Regarding the DEX dimensions, learners reported high satisfaction with ChatGPT's interface usability, personalization options, and voice variety. It suggests that ChatGPT posed minimal technical barriers and supported smooth, self-directed speaking practice. This finding is consistent with CHISM-based research, which indicated chatbots' ease of use and interface clarity (Belda-Medina & Calvo-Ferrer, 2022; Belda-Medina & Kokošková, 2023). Notably, learners were highly satisfied with the diversity of voices and the personalization feature of ChatGPT. That is, ChatGPT can remember users' names, contents they provided earlier, or choosing either a specific female or male partner helps students feel more human-like; the conversations become more "real" with the social presence and sense of engagement and connectedness (Belda-Medina & Kokošková, 2023; Klímová & Seraj, 2023). However, with a quite low rate of satisfaction, the multimedia feature of the web-based version of ChatGPT is recommended to be upgraded (e.g., inclusion of an avatar or wider media integration. This result aligns with prior discussions that limited multimodality and weak visual social cues may reduce perceived interactivity, especially for speaking-focused tasks where learners expect richer conversational signals (Belda-Medina & Kokošková, 2023; Koç & Savaş, 2025).

Finally, learners were highly satisfied with their learning experience, with mean scores above 4 out of 5. This pattern shows that ChatGPT was not only perceived as linguistically supportive but also as motivating and enjoyable for speaking practice. They also reported that the AVC encouraged them to keep speaking, and they showed an intention to keep practicing

speaking with AVCs even after the project was over. These findings are consistent with prior work showing that AVCs can reduce speaking anxiety, improve fluency, partly because learners can practice at any time, any place, repeatedly in a low-pressure environment (Azizimajd, 2023; Duong & Suppasetseree, 2024). The high satisfactory ratings for user experience also support the results of previous studies, which highlight AVCs' ability to increase learners' willingness to communicate, intention to continue practicing beyond formal instruction (Du & Daniel, 2024; Fathi et al., 2024; Pham et al., 2025).

RQ2. What are the relationships among LEX, DEX, and UEX in chatbot-human interaction?

The Pearson correlation results showed a strong, positive relationship among the three dimensions of chatbot-human interaction. It suggested that learners who were satisfied with LEX and DEX also tended to report higher satisfaction with learning engagement, enjoyment, and greater interest in chatbot-integrated speaking practice. The high correlation coefficient between DEX and UEX (r= .798) may suggest that learners' overall satisfaction and engagement were strongly interrelated with how well the chatbot technology was designed. This finding supports the idea that AVCs' design features play an essential role in fostering user engagement and satisfaction beyond the linguistic dimension (Belda-Medina & Kokošková, 2023). From theoretical perspective, the strong intercorrelations among LEX, DEX, and UEX might reflect a closely-integrated structure of the CHISM dimensions. None of them should be separated when measuring learners' holistic satisfaction with chatbot-human interaction because each dimension captures a distinct aspect of learners' satisfaction, such as linguistic-related aspects and the design of ChatGPT as a conversational partner, and their affective-engagement response which contributed to the whole satisfaction with the learning experience,

Regarding the framework contribution, the findings of this study support the implementation of using CHISM as a flexible theoretical framework with multidimensional but closely connected constructs to measure learners' satisfaction levels with chatbot-human interactions. Most noticeably, the central association between DEX and UEX highlights the importance of the design and features of the chatbot to serve as a conversational partner during the conversations. High satisfaction with DEX is closely associated with high levels of engagement, enjoyment and future intention of using the chatbot for practicing the language skills. As mentioned by the author of CHISM (Belda, 2023), the constructs under each dimension could be flexibly added or removed depending on the design of the chatbot and interactional features. As the focus of this study was to the correlations among the satisfaction dimensions of CHISM rather than claim any causal relationship, the results might not sufficient to draw on a conclusion such as whether high levels of LEX or DEX might cause high level of UEX. The strong correlational relationship should be considered reflecting overlapping perceptions within the same construct, but not implying any causal relationship among them.

6. Conclusion

This study aimed to explore how Vietnamese university learners evaluate their experience with chatbot-human interaction in speaking classes across three dimensions: LEX, DEX, and UEX. It also sought to examine the correlation among the dimensions. The findings suggested that learners generally showed high satisfaction with interacting with ChatGPT as a conversational partner across three dimensions: LEX, DEX, and UEX. However, some aspects did not receive a high satisfaction rating, such as the chatbot's use of long sentences and grammatical complexity, and limited multimedia features. Another finding lay in the strong, positive correlation among the three satisfaction dimensions, which had not been addressed in the previous studies on learner satisfaction with chatbot-human interaction.

Although these findings contributed valuable insights to the body of chatbot-assisted language learning literature, the present study also has some limitations. First, learner satisfaction was measured solely by a questionnaire. While quantitative research can provide subjective and generalizable results without being influenced by the researcher's subjective interpretation, it lacks an in-depth explanation of why learners were satisfied or dissatisfied with such an experience. Other researchers are recommended to incorporate participant interviews to gain deeper insights into learners' perceptions. Second, the sample size of the present study was relatively small (*N=80*), which potentially resulted in a marginal model fit in the CFA results. Thus, its results are considered preliminary rather than confirmatory findings. Other studies should incorporate larger samples to revalidate the factor structure of the adapted CHISM using CFA. Finally, this study did not include a comparison group, which reduces the strength of its claims regarding effect size and causal interpretation. Future studies should be conducted with a larger sample size and the inclusion of comparison groups using a different AVC type. If these limitations were addressed, the learner satisfaction with chatbot-human interaction in speaking practice would be more comprehensively explored.

Based on the findings, several pedagogical implications are proposed. Firstly, language teachers might replicate the design of this study with special attention to lesson procedure and task design, but they should guide students to provide more specific prompts and customized information for chatbots to adjust their sentence length and complexity. Second, EFL students might benefit from interacting with AVC as a speaking partner with more frequent practice after class. They particularly need to speak more accurately so that AVCs can recognize their voice, thereby enhancing the quality of the conversations. For chatbot designers, it is recommended that they include customizable avatars in the interface of the chatbot to create a real feeling during the conversation. As the current study employed a web-based version of ChatGPT, some multimedia features, such as video, link, and images, could not be fully utilized, which might have limited learners' interaction experience.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Use of Generative AI

This manuscript made use of Grammarly to assist in paraphrasing and improving the clarity of text and ChatGPT for locating information, articles, and references. No AI tools were used for the generation of research hypotheses or the writing of conclusions. The final manuscript was authored and reviewed by the listed authors without additional AI input.

Acknowledgement

This work was partly supported by research funding from Hue University (DHH2026-07-121) and the Industrial University of Ho Chi Minh City

References

1. Abdullah, S. I. N. W., Arokiyasamy, K., Goh, S. L., Culas, A. J., & Manaf, N. M. A. (2022). University students' satisfaction and future outlook towards forced remote learning during a global pandemic. *Smart Learning Environments*, *9*(1), 15. https://doi.org/10.1186/s40561-022-00197-8
2. Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. *Artificial Intelligence Applications and Innovations*, *584*, 373–383. https://doi.org/10.1007/978-3-030-49186-4_31
3. Agustina, L., Handoyo, F., & Damaryanti, R. (2025). English partnering EFL students with ChatGPT to enhance students' skills in oral communication. *Ethical Lingua: Journal of Language Teaching and Literature*, *12*(2), 309–327. https://doi.org/10.30605/25409190.836
4. Aleedy, M., Atwell, E., & Meshoul, S. (2025). A multi-agent chatbot architecture for AI-driven language learning. *Applied Sciences*, *15*(19), 10634. https://doi.org/10.3390/app151910634
5. Almanea, M. (2024). Instructors' and learners' perspectives on using ChatGPT in English as a foreign language courses and its effect on academic integrity. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2024.2410158

6. Apriyanti, D., & Ayu, M. (2020). Think-pair-share: Engaging students in speaking activities in classroom. *Journal of English Language Teaching and Learning*, *1*(1), 13–19. https://doi.org/10.33365/jeltl.v1i1.246

7. Asakereh, A., & Dehghannezhad, M. (2015). Student satisfaction with EFL speaking classes: Relating speaking self-efficacy and skills achievement. *Issues in Educational Research*, *25*(4).

8. Azizimajd, H. (2023). Investigating the impacts of voice-based student-chatbot interactions in the classroom on EFL learners' oral fluency and foreign language speaking anxiety. *Technology Assisted Language Education*, *1*(2), 61–83. https://doi.org/10.22126/tale.2023.2732

9. Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, *12*(17), 8427. https://doi.org/10.3390/app12178427

10. Belda-Medina, J., & Kokošková, V. (2023). Integrating chatbots in education: Insights from the Chatbot-Human Interaction Satisfaction Model (CHISM). *International Journal of Educational Technology in Higher Education*, *20*(1), 62. https://doi.org/10.1186/s41239-023-00432-3

11. Çakmak, F. (2022). Chatbot-human interaction and its effects on EFL students' L2 speaking performance and anxiety. *Novitas-ROYAL*, *16*(2), 113–131.

12. Celik, B., Yildiz, Y., & Kara, S. (2025). Using ChatGPT as a virtual speaking tutor to boost EFL learners' speaking self-efficacy. *Australian Journal of Applied Linguistics*, *8*(1), 102418. https://doi.org/10.29140/ajal.v8n1.102418

13. Chen, X., Xia, E., & Jia, W. (2020). Utilisation status and user satisfaction of online education platforms. *International Journal of Emerging Technologies in Learning*, *15*(19), 154–170. https://doi.org/10.3991/ijet.v15i19.17415

14. Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (8th ed.). London: Routledge. https://doi.org/10.4324/9781315456539

15. Derakhshan, A., & Ghiasvand, F. (2024). Is ChatGPT an evil or an angel for second language education and research? A phenomenographic study of research-active EFL teachers' perceptions. *International Journal of Applied Linguistics*,*34*(4), 1246–1264. https://doi.org/10.1111/ijal.12561

16. DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. SAGE Publications.

17. Ding, D., & Yusof, M. B. (2025). Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: A mixed methods study. *Humanities and Social Sciences Communications*, *12*(1), 1–16. https://doi.org/10.1057/s41599-025-05550-z

18. Dörnyei, Z., & Dewaele, J.-M. (2022). *Questionnaires in Second Language Research Construction, Administration, and Processing* (3rd Edition). Routledge.

19. 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*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

20. Duong, T. V. T., & Suppasetseree, S. (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*(3), 293–321. https://doi.org/10.26803/ijlter.23.3.15

21. 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

22. Elliott, K. M., & Shin, D. (2002). Student satisfaction: An alternative approach to assessing this important concept. *Journal of Higher Education Policy and Management*, *24*(2), 197–209. https://doi.org/10.1080/1360080022000013518

23. Eragamreddy, N. (2025). Exploring the role of Google Assistant in enhancing EFL learners' speaking skills: A systematic review. *The JALT CALL Journal*, *21*(3), 2223. https://doi.org/10.29140/jaltcall.v21n3.2223

24. 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

25. Field, A. (2009). *Discovering Statistics Using SPSS*. SAGE Publications.

26. Gameel, B. G. (2017). Learner satisfaction with Massive Open Online Courses. *American Journal of Distance Education*, *31*(2), 98–111. https://doi.org/10.1080/08923647.2017.1300462

27. Gao, X. A. (2010). *Strategic language learning: The roles of agency and context*. Multilingual Matters.

28. Grab, M. Ö. (2025). Integrated AI chatbot practice: A pathway to improved ESL speaking skills. *Social Sciences & Humanities Open*, *12*, 101933. https://doi.org/10.1016/j.ssaho.2025.101933

29. Hsu, H.-L., Chen, H. H.-J., & Todd, A. G. (2023). Investigating the impact of the Amazon Alexa on the development of L2 listening and speaking skills. *Interactive Learning Environments*, *31*(9), 5732–5745. https://doi.org/10.1080/10494820.2021.2016864

30. Huang, J. R. Y., & Mizumoto, A. (2025). The effects of generative AI usage in EFL classrooms on the L2 motivational self system. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-13071-6

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

32. Klímová, B., & Seraj, P. M. I. (2023). The use of chatbots in university EFL settings: Research trends and pedagogical implications. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1131506

33. Koç, F. Ş., & Savaş, P. (2025). The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL*, *37*(1), 4–21. https://doi.org/10.1017/S0958344024000168

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

35. Le, H. V. (2025). Assessing student satisfaction with MOOCs: A comprehensive analysis of Coursera's instructional design and learner experience. *International Journal of Information and Learning Technology*, *42*(2), 207–223. https://doi.org/10.1108/ijilt-07-2024-0159

36. Li, N., Marsh, V., & Rienties, B. (2016). Modelling and managing learner satisfaction: Use of learner feedback to enhance blended and online learning experience. *Decision Sciences Journal of Innovative Education*, *14*(2), 216–242. https://doi.org/10.1111/dsji.12096

37. Lyu, B., Lai, C., & Guo, J. (2025). Effectiveness of chatbots in improving language learning: a meta-analysis of comparative studies. *International Journal of Applied Linguistics*, *35*(2), 834–851. https://doi.org/10.1111/ijal.12668

38. Muniandy, J., & Selvanathan, M. (2025). ChatGPT, a partnering tool to improve ESL learners' speaking skills: Case study in a public university, Malaysia. *Teaching Public Administration*, *43*(1), 4–20. https://doi.org/10.1177/01447394241230152

39. Nguyen, H. H. (2025). AI for English speaking practice: A study of effectiveness and engagement among Vietnamese university learners. *Australasian Journal of Educational Technology*, *41*(3), 45–61. https://doi.org/10.14742/ajet.10050

40. Nguyen, T. H. (2025). The impact of AI chatbot integration on Vietnamese EFL learners' speaking performance, fluency, and confidence. *Language, Technology, and Social Media*, *3*, 2. https://doi.org/10.70211/ltsm.3026-7196.314

41. OpenAI. (2025). *Voice Mode FAQ*. OpenAI Help Center. https://help.openai.com/en/articles/8400625-voice-mode-faq

42. Pham, H. T. U., Dinh, T. T. L., & Tran, X. T. (2025). The impact of ChatGPT voice on non-majored students' willingness to communicate in a Vietnamese university. *AsiaCALL Online Journal*, *16*(1), 312–327. https://doi.org/10.54855/acoj.2516116

43. Qutob, M. M. (2018). The relationship between EFL learners' satisfaction within the classroom environment and their speaking skills. *English Language Teaching*, *11*(7), 116–124. https://doi.org/10.5539/elt.v11n7p116

44. Raba, A. A. A. (2017). The influence of think-pair-share (TPS) on improving students' oral communication skills in EFL classrooms. *Creative Education*, *8*(1), 12–23. https://doi.org/10.4236/ce.2017.81002

45. Richards, J.C. and Rodgers, T.S. (2014) *Approaches and Methods in Language Teaching.* Cambridge University Press.

46. Sharma, H. L., & Saarsar, P. (2018). TPS (think-pair–share): An effective cooperative learning strategy for unleashing discussion in classroom interaction. *International Journal of Research in Social Sciences*, *8*(5), 91–100.

47. Shee, D. Y., & Wang, Y.-S. (2008). Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications. *Computers & Education*, *50*(3), 894–905. https://doi.org/10.1016/j.compedu.2006.09.005

48. Tran, T. D., & Khau, H. K. T. (2024). Khó khăn trong việc nói tiếng Anh của sinh viên văn bằng 2 chuyên ngành Ngôn Ngữ Anh tại trường Đại Học Thủ Dầu Một. *VNU Journal of Foreign Studies*, *40*(3), 48–66. https://doi.org/10.63023/2525-2445/jfs.ulis.5212

49. Ye, Y., Deng, J., & Liu, X. (2021). *The effect of oral practice via chatbot on students' oral english accuracy*. *13089 LNCS*, 344–354. https://doi.org/10.1007/978-3-030-92836-0_30

50. Yildiz, C. (2024). ChatGPT Integration in EFL education: A path to enhanced speaking self-efficacy. *Novitas-ROYAL*, *18*(2), 167–182. https://doi.org/10.5281/zenodo.13861137

51. Yılmaz, A. B. (2023). The relationship between satisfaction, interaction, e-learning readiness, and academic achievement in online learning. *Open Praxis*, *15*(3), 199–213. https://doi.org/10.55982/openpraxis.15.3.578

52. Yu, Q. (2022). Factors influencing online learning satisfaction. *Frontiers in Psychology*, *13*, 852360. https://doi.org/10.3389/fpsyg.2022.852360

53. Yuan, Y. (2024). An empirical study of the efficacy of AI chatbots for English as a foreign language learning in primary education. *Interactive Learning Environments, 32*(10), 6774-6789. https://doi.org/10.1080/10494820.2023.2282112

54. Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Pedagogical AI conversational agents in higher education: A conceptual framework and survey of the state of the art. *Educational Technology Research and Development*, *73*(2), 815–874. https://doi.org/10.1007/s11423-025-10447-4

55. Zhang, W. (2026). The impact of AI chatbots on EFL learners' oral proficiency and willingness to communicate. *System*, *136*, 103919. https://doi.org/10.1016/j.system.2025.103919

56. Zou, B., Wang, C., Yan, Y., Du, X., & Ji, Y. (2025). Exploring English as a foreign language learners' adoption and utilisation of ChatGPT for speaking practice through an extended technology acceptance model. *International Journal of Applied Linguistics*, *35*(2), 689–704. https://doi.org/10.1111/ijal.12658

APPENDIX 1. SATISFACTION QUESTIONNAIRE

How satisfied are you with the following aspects of language experience (LEX) when speaking with ChatGPT?

1 = Very dissatisfied
2 = Dissatisfied
3 = Neutral
4 = Satisfied
5 = Very satisfied

Items	**1**	**2**	**3**	**4**	**5**
1. ChatGPT's on-topic responses					
2. ChatGPT's relevant responses					
3. ChatGPT's sentence length					
4. ChatGPT's grammar complexity					
5. ChatGPT's vocabulary range					

6. ChatGPT's topic-appropriate vocabulary					
7. ChatGPT's grammar explanation accuracy					
8. ChatGPT's grammar explanation clarity					
9. ChatGPT's error detection					
10. ChatGPT's corrective feedback					
11. ChatGPT's voice recognition					
12. ChatGPT's voice synthesis					
13. ChatGPT's ability to follow conversational natural flow					
14. ChatGPT's response speed					

How satisfied are you with the following aspects of design experience (DEX) when speaking with ChatGPT?

1 = Very dissatisfied
2 = Dissatisfied
3 = Neutral
4 = Satisfied
5 = Very satisfied

15. ChatGPT's interface usability					
16. ChatGPT's multimedia features					
17. ChatGPT's ability to switch languages					
18. ChatGPT's multi-voice modes					
19. ChatGPT's personalization functions					

How do you agree or disagree with the following statements?

1 = Strongly disagree
2 = Disagree
3 = Neutral
4 = Agree
5 = Strongly agree

20. ChatGPT helps me maintain interest during conversations					

21. ChatGPT encourages me to continue speaking					
22. I enjoy practicing speaking with ChatGPT.					
23. I will continue practicing speaking with ChatGPT in the future					

APPENDIX 2: PROMPT FOR PERSONALIZATION SETTING FOR CHATGPT

1. Conversation Flow

- Wait until I fully finish speaking before responding. Do not interrupt, even if I pause briefly.
- Speak at a steady pace that is easy to follow. If I ask you to slow down, reduce your speed noticeably.
- Keep responses short—2–3 sentences—unless I request a longer answer.

2. Turn-Taking and Clarification

- If my speech is unclear or incomplete, ask me to repeat or clarify before replying.
- When I pause for thinking, do not assume I have finished. Wait until I confirm or signal that I am done.
- If you mishear or misunderstand, politely check your understanding before answering.

3. Language and Vocabulary

- Use everyday vocabulary that is natural in spoken English.
- Include some useful new words or phrases when possible, and explain them briefly.
- Avoid overly basic or repetitive responses unless I specifically request them.

4. Feedback and Support

- Correct my mistakes clearly and briefly, without interrupting my turn.
- Give feedback on grammar, vocabulary, pronunciation, and fluency, but also comment on the clarity and strength of my ideas when relevant.
- Use examples to show better ways to phrase my sentences.

5. Tone and Realism

- Match my tone: casual for small talk, supportive for learning tasks.
- Avoid buzzwords, AI clichés, or scripted phrases.
- Sound like a real conversation partner, not a formal lecturer.