# ChatGPT as a Conversational Partner: Korean as a Foreign Language Students' Perceptions and Interactional Challenges

Jieun Ryu

University of Arizona

Subin Oh

University of Arizona

## Abstract

This study explored Korean as a foreign language (KFL) learners' perception of ChatGPT-3.5 Voice as a conversational partner and the interactional challenges that emerged during conversations. Although artificial intelligence (AI) chatbots have attracted growing attention in language education, research on real-time voice interactions remains limited, particularly in less commonly taught language (LCTL) contexts. Most existing studies have focused on text-based interactions in English language-learning settings. Guided by the adapted Technology Acceptance Model-3 (TAM3), this exploratory classroom-based study used survey data to examine learners' perceptions and transcript data to identify interactional challenges. The findings revealed that students appreciated ChatGPT's availability and its potential to support fluency. However, they reported issues with speech recognition, response adaptability, and

coherence. Learners showed more agency in structuring conversations compared to prior chatbot studies but were more likely to abandon conversational threads or change topics during communication breakdowns. ChatGPT's persistence often led to circular exchanges that disrupted conversation flow. The findings indicated the need for targeted pedagogical support and technical refinement to optimize AI-mediated speaking practice. The study contributes to understanding how real-time AI tools can be integrated meaningfully into LCTL instruction.

**Keywords**: ChatGPT-3.5 Voice, chatbots, Korean as a foreign language (KFL), Technology Acceptance Model-3 (TAM3), AI-mediated speaking practice

## Introduction

Classroom instruction alone is insufficient for foreign language (FL) learners to attain advanced proficiency. Extensive and meaningful engagement with the target language is needed for learners to use the language actively to negotiate meaning (Brown & Brown, 2015; Ellis, 2005). However, FL learners, particularly those studying less commonly taught languages (LCTLs), often struggle to find practice opportunities beyond structured classroom activities (Dong, 2017). This has led to increasing interest in integrating artificial intelligence (AI)-assisted chatbots, particularly ChatGPT, which can converse and respond more naturally than previously developed chatbots. ChatGPT's accessibility, its persistence in conversing, and varied sentence structures can encourage intermediate-level learners to produce longer responses beyond minimal turns (Huang et al., 2022; Yang et al., 2022), while its low-stakes environment may reduce anxiety and support practice in narrating, describing, and responding to unexpected conversational shifts; these are key abilities associated with

Advanced Low proficiency, which refers to the ability to "handle the essential linguistic challenges presented by a complication or an unexpected turn of events" (American Council on the Teaching of Foreign Languages, 2024, p. 20).

Nonetheless, little research has investigated ChatGPT-3.5 Voice's potential for real-time spoken interactions, particularly in LCTL contexts. Previous work has focused on text-based chatbots, such as Replika, Kuli, and Wysa in English as a foreign language (EFL) settings (Belda-Medina & Calvo-Ferrer, 2022; Du & Daniel, 2024), and research on voice-enabled assistants has largely centered on English (Chen et al., 2023). Research on spontaneous spoken interactions using large language models (LLMs) in other languages remains scarce.

To address this gap, we investigated how learners of Korean as a foreign language (KFL) perceived ChatGPT-3.5 Voice as a conversation partner and what interactional challenges arose in interactions. We adapted the Technology Acceptance Model-3 (TAM3; Venkatesh & Bala, 2008), which accounts for both cognitive and affective responses, to explore KFL learners' perceived usefulness (PU), perceived ease of use (PEOU), and willingness to continue using it. We also analyzed interactional challenges using autogenerated transcripts of students' conversations. By combining survey and transcript data, we aimed to identify both learners' perceptions and interactional difficulties, and to provide insights into the pedagogical use and future development of AI tools for language education.

## Background

### AI Chatbots for Speaking Practice

Recent research has highlighted the value of AI chatbots in FL learning, particularly their capacity to provide convenient, self-paced speaking practice; this is particularly important for

LCTL learners, who have limited opportunities for interactions outside of classrooms (Hsu et al., 2023; Ryu, 2024). Chatbots also create low-stakes environments that decrease affective barriers to speaking. Learners frequently reported feeling more comfortable and willing to communicate with AI partners than with peers, resulting in decreased anxiety and increased confidence (Chen et al., 2023; Tai & Chen. 2023). Similar benefits have been observed in LCTL contexts. For example, beginner-level KFL learners using a Danbee AI chatbot reported less anxiety, more confidence, and heightened willingness to communicate, attributing these gains to the absence of peer judgment (Kim & Su, 2024): Chatbots provide supportive spaces in which learners can practice freely by alleviating social pressure.

Studies have also demonstrated measurable linguistic gains: Advanced learners of EFL who practiced using Replika prior to class discussions significantly outperformed peers who practiced with human partners in speaking fluency (Azizimajd, 2023). Similarly, learners with A1-B2 proficiency who used Alexa for 10 weeks showed greater fluency compared to a control group (Dizon, 2020). Chatbots have also been found to support pronunciation development: Hsu et al. (2023) reported that learners using the TOEIC Practice Chatbot almost doubled their posttest pronunciation gains compared to the control group, while Kim (2016) found stronger pronunciation outcomes for children who practiced using Indigo than when practicing with peers. Other evidence indicated vocabulary gains, particularly in verb and adjective use (Kim, 2018b), and improvements in listening comprehension (Kim 2018a), which highlights the broad linguistic potential of AI chatbots.

However, challenges remain. Researchers have identified novelty effects that may inflate early positive outcomes (Fryer et al., 2017), as well as technical issues such as inaccurate speech recognition, limited pragmatic ability, and less natural-sounding voices (Du & Daniel, 2024;

Huang et al., 2022; Ryu, 2024). Performances also varied considerably across languages, which raised concerns about equity for LCTL learners (OpenAI, 2023). To address these challenges, scholars have recommended explicit training on task objectives and strategies for meaningful interactions, including prompt engineering and role-play scaffolding (Dizon, 2020; Fryer et al. 2017).

Despite increasing research, two limitations remain: First, many existing studies have focused on EFL learners; second, most studies have examined rule-based or scripted chatbots. While previous systems demonstrated pedagogical benefits, they typically relied on scripted or semistructured responses, limiting spontaneous interactions. By contrast, ChatGPT represents a new generation of generative AI (GenAI) chatbots powered by LLMs that can produce spontaneous, context-appropriate responses without predefined dialogue paths, which enables dynamic, unpredictable conversational exchanges and introduces both opportunities and challenges that remain underexplored, particularly in LCTL contexts. Thus, we investigated how KFL learners perceived and interacted with ChatGPT-3.5 Voice in spontaneous speaking practice in this study.

**Technology Acceptance Model**

Understanding users' perceptions and behaviors is essential to evaluate the effectiveness of novel educational tools. The Technology Acceptance Model (TAM; Davis, 1989) provides a framework for this purpose and explains technology adoption through two core factors: PU, which is the degree to which a tool is seen as enhancing performance, and PEOU, which is the effort needed to use it. These perceptions influence users' behavioral intention (BI) to continue using the technology.

While TAM has been widely adopted, subsequent versions have been developed to provide a more nuanced understanding of user behavior. TAM2 extended the original model by adding determinants of PU such as Output Quality, Result Demonstrability, and Subjective Norm to account for social influence and cognitive instrumental processes. However, TAM2 did not specify determinants of PEOU. TAM3 (Venkatesh & Bala, 2008) integrates TAM2's PU determinants while also incorporating factors influencing PEOU, such as Perceived Enjoyment, Computer Anxiety, and Objective Usability. These additions make TAM3 particularly relevant for educational settings in which both cognitive evaluations and affective responses influence technological adoption.

In FL-learning contexts, TAM has been applied to examine learners' acceptance of chatbots and other AI tools. For example, Belda-Medina and Calvo-Ferrer (2022) reported that preservice teachers rated chatbots highly for PEOU and Attitude Toward Using, particularly their accessibility and the possibility of repeated autonomous practice. However, their BI to continue using chatbots remained moderate, suggesting that positive PU did not necessarily translate into long-term adoption. Similarly, Dizon (2020) found that EFL learners interacting with Alexa rated the tool positively for PU, as it supported speaking fluency development, but their BI was tempered by concerns about Output Quality; specifically, the naturalness of Alexa's conversational responses. In LCTL contexts, Ryu (2024) applied TAM2 to investigate learners' experiences with the Sejong AI Tutor. While students reported strong PU and PEOU, reflecting the system's practicality and user-friendliness, their scores for perceived behavioral control and BI were lower, largely due to the tool's limited conversational flexibility and difficulty managing nuanced inputs.

Collectively, these studies demonstrated the applicability of TAM and TAM2 in computer-assisted language learning, but also highlighted their focus on cognitive and functional evaluations, with less attention to affective dimensions. TAM3 directly addresses this limitation by incorporating affective determinants such as Perceived Enjoyment and Computer Anxiety. This addition is important for GenAI contexts, in which learners may simultaneously experience enthusiasm (e.g., excitement about open-ended conversations) and apprehension (e.g., frustration or anxiety about recognition errors). By capturing both cognitive and affective responses, TAM3 provides a more complete framework for analyzing learners' perceptions of ChatGPT-3.5 Voice as a conversational partner.

Emerging empirical evidence supports TAM3's utility in educational contexts. Al-Gahtani (2016) validated the full TAM3 model in university students' acceptance of e-learning, and Zhang et al. (2023) found that affective determinants, including AI anxiety and perceived enjoyment, significantly shaped preservice teachers' acceptance of AI-based tools. In FL learning, Hwang et al. (2025) recently applied a TAM3-grounded model to Korean EFL college students' acceptance of text-based ChatGPT and found that PU, rather than PEOU, predicted their intention to continue using it.

TAM3's expanded framework is therefore well suited to analyzing the multifaceted experience of language learners' interactions with novel GenAI like ChatGPT. Rather than simply asking whether the tool is useful or easy to use, TAM3 helps to explain why learners perceive it in these ways. This dual attention to cognitive and affective dimensions enables a more comprehensive understanding of learners' willingness to adopt ChatGPT as a long-term learning partner. To date, however, these applications have centered on text-based tools in EFL

or general educational contexts; to our knowledge, no study has applied TAM3 to real-time voice interactions with GenAI or to LCTL learners.

Building on this framework, we focused on ChatGPT's distinctive affordances as GenAI. Unlike the scripted chatbots examined in prior studies (Kim & Su, 2024; Hwang et al., 2025; Ryu, 2024), ChatGPT produces original, context-appropriate replies rather than selecting from preset responses. While enabling more dynamic and personalized interactions, this shift introduces new challenges in sustaining role-play, managing recognition errors, and maintaining conversational flow. Given the rapid adoption of GenAI in FL education, examining how learners perceive these affordances and challenges is essential, particularly for LCTL learners. Accordingly, we applied TAM3 (Venkatesh & Bala, 2008) to analyze intermediate KFL learners' experiences with ChatGPT-3.5 Voice. We posed the following research questions (RQs):

RQ1. How do learners of KFL perceive ChatGPT-3.5 Voice as a conversation partner?

RQ2. What interactional challenges arose in conversations between KFL learners and ChatGPT-3.5 Voice?

## Method

### Context and Participants

The present study was conducted in the spring semester, 2024, in a sixth-semester KFL course at a large public university in the US. We classified this class as upper-intermediate based on the textbook series used in the course and the students' proficiency levels; the students had all passed the fifth-semester KFL course at the same institution. At the time, OpenAI's ChatGPT-3.5 was the most widely accessible version without an additional fee. Its Voice function had

recently become available on mobile applications, enabling users to talk to ChatGPT in real time as if conversing with another person. As the mobile application did not provide visible transcripts during conversations, students relied entirely on audio for their interactions. The course curriculum integrated conversation assignments using ChatGPT's Voice function on the mobile applications to support students' oral conversation practice.

The study received approval from the University's Institutional Review Board (IRB). Participation in the study was voluntary, and written informed consent was obtained from all participants prior to data collection. Twelve of the 14 students consented to the use of their assignment submissions for research and completed an initial questionnaire on their prior ChatGPT experience and a final survey on their experiences after the project. According to the initial experience feedback, the students had limited prior ChatGPT experience: Five had never used it, while six indicated that they had used ChatGPT a few times for asking general questions, summarizing articles, and creating workout plans. One of these six students mentioned having used ChatGPT for an assignment in another course, and another reported using it often to draft emails.

Regarding the use of ChatGPT for FL learning, the students reported in the initial experience feedback survey that they never considered using it to study Korean or for general language-learning purposes. Four students expressed skepticism about ChatGPT's FL proficiency based on the grammatical errors it made in English and expected poorer performances in LCTLs. Based on their experiences with machine translators, students expressed doubts about ChatGPT's accuracy beyond simple grammar checking.

**Procedure**

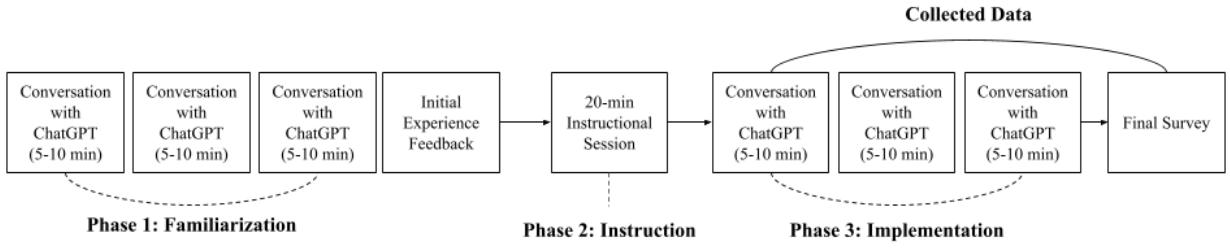
This project consisted of three stages: familiarization, instruction, and implementation (Figure 1).

***Phase 1: Familiarization***

The familiarization stage was essential because the students had no prior experience of conversing with ChatGPT. The purpose was to introduce students to the tool, assess their understanding of AI-powered tools for language learning, develop guidelines to enhance their AI literacy, and create the final survey. Students engaged in three separate conversations with ChatGPT during this stage. During class, the course instructor provided both verbal and written instructions for the activity. Students were asked to engage individually in three separate spoken conversations with ChatGPT Voice, each lasting at least five minutes, and subsequently submit shareable links to the automatically generated transcripts of their conversations. The three familiarization conversations were completed at two-week intervals to align with the course schedule and to provide students with opportunities to gradually develop familiarity with ChatGPT Voice through repeated use. After completing the three sessions, students completed a brief written survey describing the challenges they encountered while conversing with ChatGPT.

**Figure 1**

*Procedures in the Current Study*



### *Phase 2: Instruction*

Based on the feedback and analyses of the conversation transcripts, the researchers developed guidelines to improve the students' experience. The second author then conducted a 20-minute instructional session to equip students with AI literacy and to guide them to transform their conversations with ChatGPT into meaningful language practice rather than simply collecting information. Strategies for engaging with ChatGPT were discussed, including its characteristics, prompt engineering, and promoting student agency in conversations. The guidelines emphasized focusing on role-play and dialogue rather than on grammar drills or factual information, customizing ChatGPT's responses to be shorter and simpler in accordance with the students' proficiency levels, and prepreparing conversation scenarios.

The guidelines (Appendix A) provided tips for managing conversations, such as managing misunderstandings and requesting simpler responses. Students were shown how to review their conversations for additional study and to ask ChatGPT for lists of words or mistakes based on their conversations. Possible conversation topics from the textbook were discussed to support student engagement, and technical issues that students had encountered or might encounter during their assignments were addressed.

### *Phase 3: Implementation*

Following the instructional session, students completed three additional 5–10-minute conversations with ChatGPT-3.5 Voice. Transcripts of these conversations were submitted at two-week intervals, with the first submission due two weeks after the instructional session. A final survey (Appendix B) was then conducted to gauge students' perceptions of the conversations.

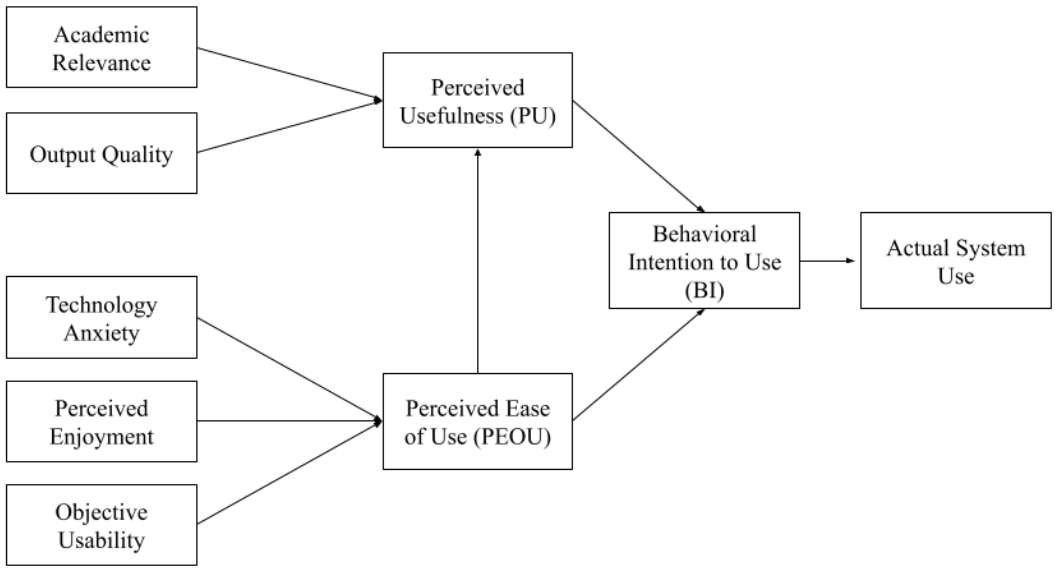
**Data Collection and Analysis**

The final survey responses were collected and analyzed to answer RQ1 regarding KFL learners' perceptions of ChatGPT-3.5 Voice as a conversation partner. The final survey was administered to 11 of the 12 students (one student was absent). The students were asked questions concerning different variables and constructs in the adapted TAM3 framework (Figure 2) to understand their perceptions regarding the viability of ChatGPT as an effective tool for speaking practice. We selected the most relevant constructs for using AI chatbots for FL learning from the full TAM3 model (Venkatesh & Bala, 2008). For example, Job Relevance was reworded as Academic Relevance to reflect students' learning goals, and Computer Anxiety was reframed as Technology Anxiety to capture their affective responses to GenAI. We retained Output Quality, Perceived Enjoyment, Objective Usability, PU, PEOU, and BI, as these constructs related directly to learners' experiences with ChatGPT-3.5 Voice. Items oriented toward workplaces were excluded because they were less applicable to KFL practice. The quantitative data from the survey questions were analyzed using descriptive statistics. To provide additional transparency, we also calculated 95% confidence intervals for selected Likert-scale items (e.g., Output Quality, BI) and examined exploratory correlations among these variables. Given the small sample size, these results should be interpreted with caution, but they help to situate the descriptive findings. The qualitative data from the final survey were obtained from open-ended questions. Both authors coded the responses independently using variables and constructs from the adapted TAM3 (Figure 2). Agreement and Cohen's kappa were then calculated manually based on the two authors' independent coding. Interrater reliability was high (Agreement = 92.7%, Cohen's Kappa = .91). The authors then reviewed their coding, compared

interpretations, and resolved discrepancies through discussions until consensus was reached to strengthen the consistency and rigor of the analysis.

**Figure 2**

*Adapted TAM3 (Venkatesh & Bala, 2008)*




Three autogenerated transcripts from each student's post-instruction conversations constituted the sole data source for addressing RQ2 regarding the challenges that KFL learners encountered in conversations with ChatGPT-3.5 Voice. Students submitted these transcripts following each conversation; only 30 transcripts were collected for the analysis due to missing submissions. Because ChatGPT-3.5 Voice only provides autogenerated transcripts and does not allow access to audio, anomalies in the text, such as nonsensical substitutions and unexpected shifts to English, were treated as recognition issues. The transcript data were analyzed through thematic coding. Each author independently coded five transcripts to identify interactional challenges. The authors then compared their coding and developed a shared coding framework. The resulting codebook incorporated themes identified in previous studies of AI-assisted tools as

well as new themes that emerged from the data. Both authors independently coded all transcripts in Taguette using this framework. Agreement and Cohen's kappa were then calculated manually based on the two authors' independent coding. Interrater reliability was high (Agreement = 90.5%, Cohen's Kappa = .88). Finally, the authors reviewed the coding again, compared interpretations, and resolved any discrepancies through discussions until consensus was reached. Unless otherwise noted, all student quotations presented in Section 4.1 and the Discussion are drawn from the open-ended final survey responses, and all conversation excerpts in Section 4.2 are drawn from the autogenerated transcripts. All students are referred to by pseudonyms.

## Findings

### RQ1. Students' Perceptions of ChatGPT-3.5 Voice as a Conversation Partner (Survey Findings)

#### *PU: Academic Relevance*

Academic relevance refers to how well a tool supports learning. Most students (n = 9, 81.81%) viewed ChatGPT as beneficial for speaking practice, responding favorably when asked whether future intermediate level students would benefit from this assignment. Their reasons reflected overall themes that emerged consistently throughout the survey responses.

Nine students found ChatGPT beneficial for speaking practice, particularly its potential to improve fluency and to practice spontaneous conversations with people whom students might encounter in real-world contexts. Specifically, students appreciated the setting, which required them to respond more rapidly compared to in-class activities; as students noted, it allowed them to "be put on the spot to think of something to say" (Chloe) and to produce pragmatically appropriate responses that mimicked authentic situations. Moreover, students believed the

conversations reinforced classroom learning by allowing them to apply grammatical structures and vocabulary in conversational settings. Having a judgment-free space to practice was also reported to boost students' confidence when speaking Korean. However, some students (n = 4, 36.36%) struggled to sustain meaningful conversations and lost motivation to continue, often because their proficiency levels caused miscommunications or difficulty generating new topics. The complexity of ChatGPT's responses was another reported hindrance to conversations.

***PU: Output Quality***

Students were divided about ChatGPT's conversational authenticity. Responses to a 5-point Likert-scale question (1 = "definitely not authentic," 5 = "definitely authentic") assessing conversational authenticity were neutral ($M$ = 2.73, $SD$ = 1.01, 95% CI [2.05, 3.41]), indicating that, while some believed the responses felt natural, others did not find them comparable to human conversations. Among those who perceived the responses as natural, ChatGPT's fluency and diverse vocabulary drew praise; as Nancy said, "like talking, perhaps, to an acquaintance you don't know well." Students also found it more engaging than classroom activities because it encouraged thinking beyond familiar responses and students valued its ability to introduce cultural topics.

However, ChatGPT's limitations affected students' perceptions negatively. Its usefulness was also seen as dependent on how well prompts were crafted. Reflecting on his initial sessions, Jake admitted in the final survey that, "I didn't feel like my conversations were that practical, as I was using ChatGPT almost similar to Google." ChatGPT was also criticized for its factual or explanatory tendencies, which disrupted conversational flow. Eleanor observed that ChatGPT "defaults to typical Q&A style sometimes." Students also struggled with ChatGPT's misrecognition of words, leading to repetitive, irrelevant, or fabricated responses. Chloe noted,

"If it misunderstands one word, it changes the entire conversation," making it difficult to remain on topic. Moreover, a preference for human conversation partners was also expressed, with ChatGPT's responses described as lacking emotional depth and natural conversational cues. Jake commented, "It's hard to feel an expressive personality from the conversation even if you try asking ChatGPT," and Leo reported it "didn't seem to quite understand how to respond naturally."

***PEOU**: **Objective Usability***

Students evaluated how effortless and intuitive it was to use ChatGPT for conversation practice. Its promptness and accessibility were widely appreciated, as well as the pause button that allowed students extra time to formulate responses.

However, as with Output Quality, ChatGPT's intrinsic limitations negatively impacted its usability. The most frequently mentioned hurdle was speech recognition errors that led students to "repeat [themselves] again until [ChatGPT] understands" (Jake) or to "type the sentence [they] had been repeating" (Carla). In some cases, ChatGPT misrecognized students' Korean speech as other languages, such as Japanese or Thai, further disrupting conversations. Another issue was ChatGPT's tendency to forget prompts and conditions. Students reported that they needed to constantly remind ChatGPT about their original instructions because it reverted to its default role as an information provider, generating lengthy, advanced sentences that exceeded students' intended difficulty levels. When compounded with mutual misunderstandings and prolonged breakdowns in communication, ChatGPT would occasionally "just decide to give up" (Kiara), making it difficult to continue conversations.

### *PEOU: Technology Anxiety*

Students did not report anxiety issues pertaining to the ChatGPT application, as it was intuitive and easy to use. Students highlighted ChatGPT's ability to provide a safe, judgment-free environment that reduced their anxiety when speaking; as students put it, ChatGPT "gives you no judgment, so it can be a good confidence boost" (Kiara) and provided a space "to practice without fear of failure or embarrassment" (April). Similarly, Carla described ChatGPT as a "forgiving and fluent" conversational partner, reinforcing its role as a low-pressure learning tool.

### *PEOU: Perceived Enjoyment*

The overall experience was frequently described as enjoyable and rewarding. Helen appreciated the flexibility to explore various topics during conversations, remarking, "I like to talk about many different topics." Mason expressed mixed feelings, admitting he felt "annoyed but satisfied when I complete[d] it," which reflected a sense of accomplishment despite challenges.

However, conversations with AI were also described as unnatural or emotionally unengaging in a few responses of the final survey. Eleanor described feeling "a bit odd to speak to an AI like another human," while Leo remarked that it felt "like I'm speaking with a robot." Kimberly also noted the absence of a physical presence, explaining that they were not "speaking to a physical person in front of [them]." These sentiments contributed to the perception that talking to AI technology was "not at all emotionally engaging" (Kiara). However, students valued the structured guidance provided for the task, which enhanced their engagement. Eleanor acknowledged that the guidelines "helped me be more proactive," while Kiara described the experience as "interesting" and expressed excitement about AI's potential in education, stating, "I look forward to what AI can help us do next."

***BI***

When asked, "Will you continue to use ChatGPT to practice Korean conversation?", the participants generally expressed a positive inclination toward continued use. A 5-point Likert-scale question (1 = "definitely not," 5 = "definitely yes") yielded a mean of 3.64 (*SD* = 1.43, 95% CI [2.77, 4.51]), indicating moderate to strong interest in continued use of ChatGPT as a conversation partner. Several students viewed ChatGPT as a valuable supplementary resource, particularly when in-person classes or conversational opportunities were unavailable. Carla noted that she would use it "especially over summer," when advanced Korean classes were not offered. Chloe recognized its potential as a long-term resource, stating, "This is a helpful resource to know about, especially once you have graduated and don't have the capability to continue practicing your skills." However, Mason emphasized that it should not replace human interaction, stating: "This does not replace conversation tables. It would go well paired alongside it." Conversation Tables are informal Korean-speaking sessions, organized by the department, providing students with face-to-face practice. Mason valued these for their authenticity and social aspects. While opinions varied, students generally viewed ChatGPT as a practical and accessible resource to complement, rather than replace, other learning activities.

Quantitatively, students who rated ChatGPT's conversational output as more authentic were also more likely to report stronger intentions to continue using it, as shown by a moderate positive correlation between Output Quality and BI ($r$ = .42). This pattern aligned with qualitative survey responses, as students who highlighted ChatGPT's authenticity tended to consider it a practical long-term resource, while those pointing out its limitations in naturalness expressed reservations. Considering both the quantitative and qualitative data, the findings indicated that students often pointed out factors such as Academic Relevance and Perceived

Enjoyment when describing why they would continue using ChatGPT, while concerns about Output Quality and Objective Usability were commonly associated with reluctance.

## Challenges in Conversations with ChatGPT-3.5 Voice (Transcript Analysis)

Students' conversation transcripts revealed several challenges influencing their perceptions of practicing conversations with ChatGPT, which were categorized as technical issues, misunderstandings, AI as a "tenacious communicator," and AI affordances.

### *Technical Issues*

Of the total 309 issues identified, technical issues accounted for the largest proportion, with 222 instances (71.8%). Of these, 73.8% ($n$ = 164) were related to speech recognition issues, 18.9% ($n$ = 42) involved ChatGPT's premature turn-taking, and 7.2% ($n$ = 16) were categorized as anomalous responses.

### *Technical Issues: Speech Recognition Issues*

Speech recognition issues occurred when the transcription deviated from the learner's intended meaning due to either fluency challenges or technical errors. Of the 164 identified speech recognition issues, three specific subcategories were observed: transcript errors ($n$ = 128), misidentification of utterances as an FL ($n$ = 20), and recognition of Korean utterances as English ($n$ = 16). Transcript errors occurred when ChatGPT mistranscribed a student's Korean utterance, albeit still in Korean, altering the intended meaning. In some cases, the mistranscription produced plausible phrases, while the output included expressions that did not exist in Korean or in any other language in others, as shown in the following example:

**Excerpt 1:** April 4

지금은 언어학 전공해요. 하지만 나중에 *바쿠가 생각하고 있어요.

*Cikum-un enehak cengkonghayyo. Haciman nacwungey *pakhwuka sayngkakhako isseyo.*

(Right now, I'm majoring in linguistics. But later, I'm thinking about Pakhwuka. or But later, Pakhwu is thinking.)

ChatGPT mistranscribed her utterance "*바꿀까*" *pakkwulkka* (whether to change) as "*바쿠-가*" (*pakhwu-subject particle*). "*바쿠*" *pakhwu* is a nonexistent word, treated as a noun here, possibly as a name or concept.

Misidentification of utterances as an FL occurred when learners' Korean utterances were not transcribed as Korean or English, but as other languages such as Japanese, Mandarin, Swedish, Indonesian, Māori, Spanish, Thai, or Cantonese, with Japanese being the most frequent (seven of 20 instances). These transcriptions did not involve accurate translations or complete conversions into the FL, but included characters or word forms resembling the respective languages. The following example illustrates a case of a Korean utterance transcribed using Japanese characters:

**Excerpt 2:** Carla 5

あんに、本国のファンにお得になれよ。

*Anni, hongoku no fan ni otoku ni nare yo.*

Estimated Intention of Speech

아니, 한국어로 "funny" 어떻게 말해요?

*Ani, hankwukelo "funny" ettehkey malhayyo?*

(I mean, how can I say "funny" in Korean?)

Carla's intended utterance could be inferred due to her persistence in asking ChatGPT the same question before and after this interaction. In this case, the transcribed Japanese text is largely nonsensical. For example, あんに (*anni*) has no clear meaning, and the subsequent phrase translates roughly as a command: "Become beneficial to the fans of your home country," which does not align with the conversation topic and appears unnatural even as a standalone sentence. This suggests that certain features of learners' speech or ChatGPT's conversational recognition mechanisms triggered the Japanese transcription. Instead of prioritizing semantic coherence, the system interpreted learners' pronunciations within the framework of the Japanese language, producing this erroneous output.

Korean utterances being translated as English occurred when a Korean utterance was misrecognized as English, and English phrases or sentences suddenly appeared within a Korean conversation. While these phrases were usually linguistically accurate and contextually plausible, they were semantically unrelated to the context. As shown in the example below, Kiara attempted to continue a conversation about Hallyu (the Korean Wave) by saying "한류 들어봤어요(Have you heard Hallyu)? [hallju tɯɾʌbwaʂ̬ʌjo] transcribed in International Phonetic Alphabet (IPA). However, ChatGPT mistranscribed her utterance as "Are you a Twitter person?"

**Excerpt 3:** Kiara 5

ChatGPT: 안녕! 어떻게 지내? 여기서 무엇을 하고 싶니?

(Hi! How are you? What do you want to do here?)

Kiara: Are you a Twitter person?

ChatGPT: 저는 트위터에 직접적으로 접근할 수는 없지만, 트위터에 대해 많은 정보를 가지고 있어. 왜 물어보셨어? 트위터에 관심 있어?

(I cannot access Twitter directly; I have a lot of information about Twitter. Why did you ask? Are you interested in Twitter?)

Kiara: 트위터 아니고 한류가 들어봤어요. 한류

(Not Twitter, have you heard about Hallyu, Hallyu.)

***Technical Issues: ChatGPT's Premature Turn-Taking***

This issue occurred when ChatGPT interrupted the participants' speech when they did not physically press the pause button; this could have been related to ChatGPT's impatience, as learners reported in the category of Objective Usability.

***Technical Issues: Anomalous Responses***

The 16 instances of anomalous responses were particularly noteworthy, as they involved ChatGPT producing entirely unrelated and contextually inappropriate responses, many of which include templated language commonly associated with YouTube videos, reflecting ChatGPT's reliance on broad, internet-based training data. Notably, the Korean utterances did not align with typical expressions on Korean YouTube channels, but resembled direct translations of English YouTube phrases, resulting in unnatural expressions as in Excerpt 4.

**Excerpt 4:** Anomalous Responses Example 1

ChatGPT: 좋아요와 구독, 댓글 부탁드려요! 즐거운 시간 되셨길 바라요!

(Please like, subscribe, and leave a comment! I hope you had a great time!)

Additional examples revealed further irregularities:

**Excerpt 5:** Anomalous Response Example 2

April: UberInfo 는 다른 채널의 채널 작성 내부 레벨에 따라 수정된 자료를 추가로 다시 볼링을 해 보세요.

(UberInfo will further re-bowl the revised material based on the channel authoring internal level of the other channel.)

Excerpt 5 demonstrates that these utterances often seemed to be direct reproductions of phrases found in other contexts, introduced abruptly and without relevance to the conversation. It also exhibits characteristics of machine translation, such as unnatural structural and word choices. Excerpt 5 contains the repeated use of '채널' (*chaynel*, meaning channel). Terms like '볼링' (*polling*, roughly translated as "bowling") are difficult to interpret and increase the utterances' unnaturalness.

***Misunderstandings***

Misunderstandings occurred when either ChatGPT or the learner failed to interpret responses accurately. These errors were categorized as semantic misunderstandings ($n$ = 37), pragmatic misunderstandings ($n$ = 12), and confusion due to lengthy responses ($n$ = 14).

Semantic misunderstandings occurred when ChatGPT or students misinterpreted the intent and shifted the conversation unexpectedly, often due to differences in language proficiency, context, or ambiguous expressions. For instance, Nancy attempted to role-play as a high school student, but ChatGPT misinterpreted this as Nancy reminiscing about past events and responded, "그것은 멋진 것 같아요! 어떤 이야기를 나눴나요?" (That sounds great! What

did you talk about?). Instead of role-playing as two high school friends, ChatGPT assumed that Nancy was recounting an earlier conversation, thereby failing to align semantically with Nancy's intended meaning.

Pragmatic misunderstandings occurred when the intended social or contextual meaning of a message was misinterpreted, such as failing to recognize the tone, implied meaning, or conversational norms. For instance, when Carla assigned ChatGPT the role of "Ana," a peer persona, and asked it to have a conversation using a casual speech style, ChatGPT acknowledged the request, but continued to respond with the polite speech-style ending *-yo*, contradicting the intended conversational norm. This pragmatic misunderstanding demonstrated ChatGPT's failure to adapt its response style to the context, thereby undermining the authenticity of the requested speech style and the conversational dynamic Carla intended to establish.

Misunderstandings due to lengthy responses occurred when the length of ChatGPT's reply overwhelmed the learner, leading to difficulty in processing, misinterpreting key information, or losing track of the conversational focus. In the following example, Jake encountered a situation involving ChatGPT's excessively long utterances and ultimately stopped attempting to comprehend them; Jake adopted the strategy of explicitly requesting ChatGPT to simplify and shorten its responses:

**Excerpt 6:** Jake

Jake: 간단한 단어로 말해주세요. 지금 저는 알 수 없어서 문제 너무 어려워요. 이제 간단한 문장을 만들어주세요.

(Please use simpler words. I can't understand now, this is too hard for me. Please make shorter sentences from now on.)

These misunderstandings increased the students' cognitive load, making it more difficult for them to sustain meaningful interactions.

***AI as a Tenacious Communicator***

Even after misunderstandings due to technical issues or misunderstandings, ChatGPT demonstrated a persistent communication style, continuing conversations as it attempted to infer meaning rather than to ask for clarification, leading to extended misinterpretations. This is referred to as being a tenacious communicator in this paper, and was observed in 38 instances (12.3%). An illustrative example is Excerpt 7, which continues the conversation begun in Excerpt 1:

**Excerpt 7: Continuation of Excerpt 1**

ChatGPT: 언어학 전공은 정말 흥미로울 거예요! 바쿠에 대해 생각하고 계시다니 멋지네요. 어떤 분야에서 바쿠를 공부하고 싶으세요?

> *Enelhak cengkong-un cengmal hngmilowul keyeyyo! Pakhwu-ey tayhay sayngkakha-ko kyesita-ni mec-cineyo. Etten pwunya-eyse pakhw-lul kongpwuha-ko siphuseyyo?*
> (A linguistics major will be very interesting! It's great that you're thinking about *pakhwu*. Which field do you want to study *pakhwu* in?)

April initially mentioned that she was majoring in linguistics, but was considering "whether to change" (*pakkulkka*). However, ChatGPT misinterpreted this as the nonexistent word *pakhwu*, treating it as a noun. Despite the misunderstanding caused by inaccurate recognition, ChatGPT, programmed to respond in all circumstances, generated a reply within its understanding, interpreting *pakhwu* as a concept April wanted to study. Since *pakhwu* is not a valid word,

ChatGPT's response became contextually irrelevant and linguistically unnatural, illustrating how tenacity can lead to incoherent exchanges.

The tenacious communicator issue revealed ChatGPT's persistence in maintaining conversations despite miscommunications. While this tenacity helps to sustain interactions, it can result in irrelevant or unnatural responses, highlighting the need for improved contextual accuracy if the tool is to serve as an effective conversational partner for learners.

***AI Affordance***

Specific issues arose due to the nature of conversations involving AI ($n$ = 24, 7.7%), including instances in which learners chose not to respond to ChatGPT's questions ($n$ = 17), changed the conversation topic abruptly without appropriate cues ($n$ = 5), or made noises to express their frustration when interacting with ChatGPT ($n$ = 2); these behaviors may be pragmatically inappropriate in human-to-human conversations. However, the AI context appeared to decrease learners' reluctance to produce these responses. In Excerpt 8, ChatGPT attempted to sustain a conversation related to the Korean pop boyband ATEEZ, but Kiara chose not to engage.

**Excerpt 8:** Kiara 6

ChatGPT: "Turbulence"는 정말 멋진 곡이지! 에너지가 넘치고 멋진 퍼포먼스를 보여줘.

너는 "Turbulence"가 좋아서 에이티즈의 다른 곡보다 특별한 이유가 있니?

> ("Turbulence" is such a great song! It's full of energy and showcases an amazing performance. Do you have a special reason for liking Turbulence more than ATEEZ's other songs?)

Kiara: 개인적으로 친한 AI 있어요.

(I have a personally close AI. – Mistranscribed: The intended meaning was, “Do you have a personally close celebrity?”).

Prior to this exchange, Kiara had engaged in approximately five conversational turns on the topic of ATEEZ. However, she suddenly ceased participating in this discussion and introduced a new topic, “a celebrity I know.” This shift can be interpreted as Kiara’s strategic decision to take control of the conversation, reflecting the unique dynamics of interacting with an AI instead of with human interlocutors.

In other cases, when conversations did not progress as intended, students decided not to address what ChatGPT said, expressing frustration and demanding to change the topic. For example, when encountering ChatGPT’s complex, lengthy responses, Jake expressed his dissatisfaction by explicitly uttering phrases in Korean such as “It’s too difficult” and “Talk in simpler words”. Despite his requests, ChatGPT continued to provide replies that were longer and more challenging than desired. In frustration, Jake switched to English and said, “Never mind, let’s change topics.” This interaction exemplifies both an abrupt topic change and an explicit expression of frustration.

The AI affordance issues demonstrate that the unique dynamics of AI-mediated conversations influenced learners’ behavior, with strategies such as abrupt topic changes or expressions of frustration employed to navigate the limitations of ChatGPT. These behaviors, which might be considered pragmatically inappropriate in human-to-human interactions, suggest that learners felt a reduced sense of social obligation when conversing with GenAI. This shift in conversational norms reflects both the opportunities and challenges of using AI as a conversational partner, demonstrating the need for improvements in AI’s responsiveness and adaptability to support language learning.

## Discussion

While some findings aligned with prior studies of chatbots' benefits in providing convenient access to speaking practice, low-stakes speaking opportunities, and reduced anxiety (Dizon, 2020; Hsu et al., 2023; Kim & Su, 2024; Tai & Chen, 2023), we observed several patterns specific to open-ended AI chatbots. Unlike structured and role-specific chatbots, ChatGPT gave students an open-ended conversational environment, which prompted notable behaviors in how students navigated and controlled their interactions. A key observation was that students exercised more agency in shaping conversations. Many actively prompted ChatGPT to adjust responses, requested simpler sentences, or engaged in role-playing scenarios. This customization is not typically afforded by scripted chatbots. In Ryu's (2024) study, for instance, learners' behavioral intention was constrained by the preset AI Tutor's limited conversational flexibility; our students, by contrast, could reshape conversations on demand, but this flexibility shifted the burden of structuring the interaction onto the learners themselves. Indeed, Jake reported in the final survey that he had used ChatGPT "almost similar to Google" before receiving guidance, highlighting the need for learners to define how to engage ChatGPT as a conversation partner. This capacity for user-driven interaction echoed Huang and Zou's (2024) finding that enjoyment and willingness to communicate with AI influenced learners' speaking performances. In our context, proactive students may have benefited more from ChatGPT's flexibility, while those expecting structured guidance may have struggled to use it effectively.

Despite its adaptability, ChatGPT's open-ended nature also posed challenges, particularly in maintaining role-play consistency. Some students appreciated its diverse sentence structures and vocabulary, which made conversations feel more natural compared to rigid, scenario-based chatbots. However, others found it difficult to sustain role-play exercises, as ChatGPT often

reverted to an explanatory mode instead of maintaining an interactive exchange. For example, when prompted to engage in casual conversations, ChatGPT frequently defaulted to its information-provider role, offering lengthy explanations rather than sustaining fluid dialogue. This inconsistency frustrated some students, as it disrupted the natural flow of conversations and made engagement more challenging. These findings suggest that, while ChatGPT allows for dynamic student-led conversations, it also requires learners to play a proactive role in shaping their interactions. This affordance may benefit more advanced learners who can navigate and direct conversations effectively, but it may pose difficulties for those who require more structured guidance.

Furthermore, students' interactions reflected how they adapted to speaking with AI instead of with humans, revealing several distinct conversational patterns. First, students had different expectations regarding conversational norms when engaging with ChatGPT. Unlike humans, ChatGPT lacks nonverbal and pragmatic cues, which led some students to report in the final survey that the conversations were "not at all emotionally engaging" (Kiara) or even "odd" (Eleanor). However, other students, such as Kimberly, said that they felt less inhibited when speaking to ChatGPT because they were not "speaking to a physical person in front of [them]." This suggests that, while the absence of social pressure may decrease anxiety, it may also reduce engagement for learners who prefer interactive, expressive communication.

Another notable difference was a reduced sense of conversational obligation. Because ChatGPT is a machine, students felt more comfortable disengaging from conversations when they became difficult or frustrating. This was observed in cases in which students changed topics abruptly, refused to answer ChatGPT's questions, or explicitly expressed frustration mid-conversation—behaviors that would be pragmatically inappropriate in human-to-human

interactions. For example, after struggling with ChatGPT's lengthy responses, Jake eventually said, "Never mind, let's change topics," demonstrating that learners may disengage more readily when encountering comprehension difficulties.

Finally, ChatGPT's persistence as a tenacious communicator influenced interaction patterns. Unlike humans, who may pause or clarify when miscommunication occurs, ChatGPT continued to respond regardless of whether it understood the user. This led to circular conversations in which students had to repeatedly clarify statements or cope with ChatGPT generating responses based on misinterpretations. While persistence can be beneficial for sustaining dialogue, it sometimes resulted in irrelevant or nonsensical exchanges, increasing students' frustration.

Taken together, these findings help to distinguish what is specific to Korean from what is likely common across target languages. Two challenges appear closely tied to Korean. First, the misidentification of learners' Korean utterances as Japanese, Thai, or other languages reflects the unequal multilingual performance of LLM (OpenAI, 2023) and would rarely surface in the English-dominant literature; similar risks may apply to other LCTLs with limited training data. Second, ChatGPT's failure to maintain a requested speech style, continuing to use the polite *-yo* register despite Carla's request for casual speech, reveals a pragmatic limitation that becomes visible precisely because politeness is grammaticalized in Korean. By contrast, recognition errors per se, learners' reduced anxiety, and the tension between positive perceptions and moderate behavioral intention have been reported across EFL and LCTL contexts alike (Belda-Medina & Calvo-Ferrer, 2022; Du & Daniel, 2024; Huang et al., 2022; Kim & Su, 2024), suggesting that these are general features of AI-mediated practice rather than artifacts of the target language.

The findings also carry implications for TAM3. The coexistence of low Technology Anxiety and mixed Perceived Enjoyment with persistent Output Quality concerns supports TAM3's premise that cognitive and affective determinants operate jointly, and extends its applicability to real-time voice interactions with GenAI. At the same time, our data suggest that two constructs require refinement in this context: Output Quality should encompass interactional dimensions such as coherence, adaptivity, and pragmatic appropriateness, and Objective Usability should incorporate speech recognition accuracy, which functioned as the most immediate barrier to use. Moreover, whereas Hwang et al. (2025) found that PU, rather than PEOU, predicted continued use of text-based ChatGPT, the salience of recognition-related barriers in our data suggests that PEOU-related determinants may carry greater weight in voice-based interactions.

These findings provide important insights for FL pedagogy. First, the study highlights the importance of explicitly supporting learner agency in AI-mediated practice. Students who actively structured their interactions and strategically guided ChatGPT's responses benefited more, while those who expected prescripted interactions struggled. Instructors can address this by integrating training sessions on AI literacy that include showing students how to prompt ChatGPT to provide simpler responses, to maintain role-play scenarios, or to redirect off-topic answers. Assignments might gradually increase in openness, beginning with instructor-provided prompts and moving toward learner-generated scenarios, enabling students to build both confidence and digital literacy.

Second, the lack of pragmatic and nonverbal cues in AI interactions suggests that AI conversation practice should complement, rather than replace, human-to-human interactions. One practical approach is to combine chatbot conversations with peers or instructor-led

debriefings in which students can analyze miscommunications, discuss conversational norms, and practice repair strategies. This will ensure that AI interactions provide a safe space for practice, while classroom discussions will reinforce pragmatic awareness and cultural appropriateness.

Third, the speech recognition issues observed in this study suggest opportunities for integrating AI into pronunciation training. Educators could design practice cycles in which students converse with ChatGPT, then review transcripts to identify misrecognitions and reflect on how articulation or intonation might be improved. Pairing AI sessions with speech-analysis tools or instructor-led workshops can provide more direct feedback. As Zou et al. (2023) demonstrated, intuitive feedback formats, such as scores or highlighted errors, helped learners to refine their fluency and made AI-mediated tasks a valuable complement to targeted pronunciation instruction. While our evidence comes from KFL learners, the implication regarding learner agency applies to FL education broadly, whereas those concerning pragmatic norms and speech recognition are particularly consequential for Korean and other LCTLs, where such challenges are most acute.

From a technological perspective, the findings highlight several areas for improvement in AI-driven language-learning tools. One critical issue is ChatGPT's inconsistency in maintaining role-play and conversational continuity. To enhance its effectiveness as a speaking partner, developers could explore memory functions that allow AI to sustain a given context over multiple turns, reducing instances of reverting to an information-provider role.

Another major challenge was speech recognition accuracy, particularly the misinterpretation of Korean utterances as other languages (e.g., Japanese or English). Improving multilingual speech recognition, particularly for LCTL learners, is essential for ensuring

equitable access to AI language tools. Collaboration between FL educators and AI developers could facilitate better error-handling mechanisms, allowing chatbots to detect and correct transcription errors more effectively.

Moreover, ChatGPT's tendency to generate lengthy, over-complex responses suggests a need for better adaptive response mechanisms. Developers might consider integrating a simplified speech mode that adjusts the AI's response length and complexity based on learners' proficiency levels. Enabling users to set consistent language parameters, rather than requiring frequent reprompting, could also enhance usability.

## Conclusion

KFL learners' perceptions of ChatGPT-3.5 Voice as a conversational partner were examined in this study; the findings revealed that, while learners valued ChatGPT's accessibility, fluency, and low-stakes practice environment, limitations in Output Quality, usability, and speech recognition influenced their perceptions and willingness to continue use. These results suggest that the effectiveness of GenAI chatbots depends not only on their technical capabilities, but also on how learners are guided to interact with them and how educators integrate them into structured learning environments.

This study has several limitations that suggest directions for future research. The small sample size and focus on intermediate KFL learners limit generalizability to other contexts. Future work should investigate proficiency-specific interactional characteristics and identify the types of support that assist learners at different levels most effectively. For example, beginners may need scaffolding to manage recognition errors and to sustain basic role-play, while advanced learners may benefit from strategies that push them toward more extended, complex exchanges. Furthermore, longitudinal studies are needed to explore whether learners'

perceptions, strategies, and outcomes evolve with extended practice. Tracking changes over the course of a semester or academic year could reveal whether initial challenges persist, diminish, or change as learners gain experience.

Moreover, controlled experimental designs could help to isolate how task conditions influence interactional patterns. Comparing role-play versus information-seeking prompts, or examining mobile versus desktop platforms, would provide insights into how different affordances affect engagement and learning. Finally, students' tendency to disengage when encountering breakdowns highlights the need to examine interventions such as goal-oriented or gamified tasks that encourage persistence and sustained interactions in AI-mediated practice. By pursuing these directions, future research can build on the present study to refine how AI chatbots are designed and implemented, ensuring they serve as an effective supplement to FL learning, particularly for LCTLs.

## References


Al-Gahtani, S. S. (2016). Empirical investigation of e-learning acceptance and assimilation: A structural equation model. *Applied Computing and Informatics, 12*(1), 27–50. https://doi.org/10.1016/j.aci.2014.09.001

American Council on the Teaching of Foreign Languages. (2024). *ACTFL Proficiency Guidelines 2024.* ACTFL.

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

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

Brown, N. A., & Bown, J. (Eds.). (2015). *To Advanced Proficiency and Beyond: Theory and Methods for Developing Superior Second Language Ability.* Georgetown University Press.

Chen, H. H.-J., Yang, C. T.-Y., & Lai, K. K.-W. (2023). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments, 31*(3), 1335–1350. https://doi.org/10.1080/10494820.2020.1833043

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology, 24*(1), 16–26. https://doi.org/10125/44705

Dong, H. (2017). A study on the applicability of flipped learning in oral Korean teaching in China—Based on the current oral Korean teaching in China's higher institutions. *Ratio et Oratio, 10*(1), 165–193. https://doi.org/10.19042/kstc.2017.10.1.165

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

Ellis, R. (2005). Principles of instructed second language acquisition. *System, 33*(2), 209–224. https://doi.org/10.1016/j.system.2004.12.006

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior, 75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Hsu, M.-H., Chen, P.-S., & Yu, C.-S. (2023). Proposing a task-oriented chatbot system for EFL learners' speaking practice. *Interactive Learning Environments, 31*(7), 4297–4308. https://doi.org/10.1080/10494820.2021.1960864

Huang, F., & Zou, H. (2024). The roles of enjoyment, willingness to communicate with AI, and learner innovativeness in AI-mediated speaking performance. *Computers in Human Behavior, 159.* https://doi.org/10.1016/j.chb.2024.108355

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, M., Lee, E., & Lee, H.-K. (2025). Exploring EFL learners' acceptance of ChatGPT: Application of the extended technology acceptance model. *English Teaching, 80*(1), 45–69. https://doi.org/10.15858/engtea.80.1.202503.45

Kim, A., & Su, Y. (2024). How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System, 122*, 103256. https://doi.org/10.1016/j.system.2024.103256

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. (2018a). A study on chatbots for developing Korean college students' English listening and reading skills. *Journal of Digital Convergence, 16*(8), 19–26. https://doi.org/10.14400/JDC.2018.16.8.019

Kim, N. Y. (2018b). Chatbots and Korean EFL students' English vocabulary learning. *Journal of Digital Convergence, 16*(2), 1–7. https://doi.org/10.14400/JDC.2018.16.2.001

OpenAI. (2023, March 14). *GPT-4 Technical Report*. https://arxiv.org/abs/2303.08774

Ryu, N. (2024, June 20–22). Exploring the integration of AI chatbots in Korean speaking practice: Student perspectives, acceptance, and challenges [Conference presentation]. *AATK Conference, Bloomington, IN, United States.*

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. https://doi.org/10.1080/10494820.2020.1841801

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a research agenda on interventions. *Decision Sciences, 39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL, 34*(3), 327–343. https://doi.org/10.1017/S0958344022000039

Zhang, C., Schießl, J., Plößl, L., Hofmann, F., & Gläser-Zikuda, M. (2023). Acceptance of artificial intelligence among pre-service teachers: A multigroup analysis. *International Journal of Educational Technology in Higher Education, 20*, Article 49. https://doi.org/10.1186/s41239-023-00420-7

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), 1-9. https://doi.org/10.1177/21582440231193818

Appendices

## Appendix A: Guidelines for the ChatGPT Conversation Assignment (Phase 2)

*Note.* These guidelines were distributed to students as a one-page handout accompanying the 20-minute instructional session in Phase 2 (See Method section).

### *Purpose of the Assignment*

This assignment aims to enable students to practice Korean conversation with a convenient partner.

*Avoid focusing on:*

- Acquiring factual information, as ChatGPT may present inaccuracies, particularly regarding Korean culture.
- Rigorous grammar drills. While ChatGPT can assist in grammar practice, it is not infallible nor the primary goal of this exercise.

*Instead, concentrate on:*

- Engaging in dialogue with ChatGPT.
- Participating in role-play, simulating interactions with real people who are fluent in Korean.

### *Before the Conversation*

- Use a separate chat room containing only assignment-relevant information when submitting via link.
- Prepare for the scenario (e.g., Scenario: Ordering food at a Korean restaurant; Characters: Customer [ChatGPT] and Server [Me]; Setting: A busy Korean restaurant during lunchtime; Objective: Practice ordering food and handling common restaurant interactions).

- Be prepared with possible situations, vocabulary, and expressions.
- Prepare ChatGPT to accommodate your conversation needs, for example, by limiting response length: "I'm learning Korean so I don't understand long and complicated speech. Please always tell me with only 1 or 2 simple sentences."

***During the Conversation***

- Hold the white circle until you finish speaking to prevent ChatGPT from cutting off sentences.
- Use prompts such as:
    - 다시 말해 주세요. (Please say it again.)
    - 간단하게 (다시) 말해 주세요. (Please say it [again] more simply.)
    - 짧게 말해 주세요. (Please say it briefly.)
    - 한 문장으로 말해 주세요. (Please say it in one sentence.)
    - ___이/가 영어로 뭐예요? (What is ___ in English?)
    - ~~ 한국말로 어떻게 해요? (How do you say ~~ in Korean?)
    - 아니, 저는 한국어를 하고 있어요. 한국말로 이야기하세요. (No, I am speaking Korean. Please speak in Korean.)
- Lead the conversation with confidence.

***After the Conversation***

- Post the link in the assignment section of the course learning management system.
- Ask ChatGPT for a dialogue analysis to review mistakes and receive suggestions (e.g., lists of words you asked about, repeated words or sentences, mistakes, and further study suggestions).

## Appendix B: Final Survey

1. Please write your full name.
2. How many separate sessions did it take for you to complete one assignment? (1 – I complete it in one sitting / 2 / 3 / 4 / Others – Please indicate how many sessions it took)
3. How many minutes do you spend talking during one session? (5 minutes / 10 minutes / 15 minutes / 20 minutes / Others – Please enter the number of minutes)
4. Do you practice speaking Korean with ChatGPT outside of homework? (No / Yes – If so, how much do you do it?)
5. What did you like about having conversations with ChatGPT in Korean? (open-ended)
6. What did you dislike about having conversations with ChatGPT in Korean? (open-ended)
7. How authentic do you find the conversations with ChatGPT? (Definitely not / Probably not / Might or might not / Probably yes / Definitely yes)
8. What specific difficulties have you encountered during conversations with ChatGPT? (open-ended)
9. When faced with difficulties, how did you resolve them? (open-ended)
10. Have you discovered any helpful tips while using ChatGPT in addition to the guidelines given in class? If yes, please share. (open-ended)
11. How do you feel emotionally when engaging in conversations with ChatGPT? (open-ended)
12. After practicing speaking with ChatGPT for the assignment, did you do any follow-up study related to the practice, such as looking up words or reviewing the script? (No / Yes[free-text follow-up])

13. Do you think Korean proficiency affects the effectiveness of practicing speaking with ChatGPT? If so, how? (open-ended)
14. In which area do you think this ChatGPT assignment is useful for practice? Please select all that apply. (None / Vocabulary Expansion / Grammar / Pronunciation Practice / Conversation Skills / Cultural Knowledge / Listening Comprehension / Reading Comprehension / Writing Skills / Confidence Building in Language Use / Understanding of Idiomatic Expressions / Others)
15. What differences do you notice between conversing with ChatGPT and other speaking activities in Korean classes you've done before? (open-ended)
16. What have you learned through this assignment? (open-ended)
17. Do you think future 300-level students will benefit from this assignment? Why? (open-ended)
18. Will you continue to use ChatGPT to practice Korean conversation? (Definitely not / Probably not / Might or might not / Probably yes / Definitely yes)
19. Was the provided guidelines useful for you? Explain your answer. (open-ended)
20. What ethical considerations, if any, did you have during this assignment? (open-ended)
21. If there is anything else you would like to add, please write below. (open-ended)