The impact of a chatbot-assisted flipped approach on EFL learner interaction

Jaeho Jeon[1] and Seongyong Lee[2*]
[1]Indiana University Bloomington, USA // [2]University of Nottingham Ningbo China, China //
jaehojeon21@gmail.com // seongyonglee77@gmail.com
[*]Corresponding author

(Submitted February 12, 2023; Revised February 6, 2024; Accepted May 10, 2024)

ABSTRACT: Flipped learning is gaining prominence in EFL classrooms, where it has been shown to provide various advantages. However, the literature has not adequately addressed the integration of L2 interaction outside the classroom into flipped learning design. Accordingly, given the ubiquitous and interactive nature of chatbot technology, we designed a chatbot-assisted flipped approach to increase learners' opportunities for L2 practice. To examine the impact of this approach, we investigated how the use of chatbots as conversational partners in the out-of-class phase of flipped learning affected students' interactive performance and perceptions of interaction. We randomly assigned 87 elementary school EFL learners to experimental (n = 42) and control (n = 45) groups. For six weeks, both groups were taught using the same flipped approach except that after viewing a video, the experimental students practiced a pre-class dialogue with chatbots while the control group practiced the dialogue using a worksheet. Employing a sequential, explanatory mixed-methods design, we collected two sets of quantitative data (pre- and post-interactive speaking tests and an interaction questionnaire) and one set of qualitative data (semi-structured interviews). The quantitative data analyses showed that the experimental group outperformed the control group in terms of interactive speaking performance and the degree to which they perceived they had participated in learner-learner and learner-instructor interactions. Thematic analysis of the qualitative data supplemented the quantitative results by elucidating how the chatbots facilitated the learners' interactive speaking performance and perceptions of interaction. This study highlights the potential of chatbot technology to bridge the out-of-class and in-class phases of flipped learning.

Keywords: Flipped learning, Artificial intelligence, Chatbot, Customized chatbot, Interaction

1. Introduction

According to the Flipped Learning Network (2014),

"Flipped learning is a pedagogical approach in which direct instruction moves from the group learning space to the individual learning space, and the resulting group space is transformed into a dynamic, interactive learning environment where the educator guides learners as they apply concepts and engage creatively in the subject matter" (p. 1).

Research has demonstrated the positive effects of flipped learning on academic, intra- and inter-personal, and affective outcomes across different subject areas and levels of learners (Akçayır & Akçayır, 2018; Bond, 2020; Bredow et al., 2021). Several studies have affirmed a major benefit of flipped learning as providing extended opportunities for interactive activities in the classroom supported by the learners' acquisition of needed knowledge outside the classroom (e.g., Hung, 2017; Hwang et al., 2015; Lin & Hwang, 2018). As an interactive instructional environment is especially important for language learners, researchers have investigated the benefits of flipped learning that are unique to language instruction, such as accommodating interaction opportunities to facilitate students' acquisition of communicative language skills and promoting willingness to communicate in a target language (Jiang et al., 2022b; Zou et al., 2022).

To explore interactions in flipped learning settings, Moore's (1989) interaction framework comprising three types of interaction (learner-content, learner-learner, and learner-instructor) has been widely used (Bernard et al., 2009; Hwang & Chen, 2023). Based on this framework, scholars have found that flipped learning provides benefits to learners both outside and inside the classroom (e.g., Förster et al., 2022; Lin & Hwang, 2018). For example, Lin et al. (2023) found that a well-designed flipped course promotes these three types of interaction leading to students' enhanced perceptions of their learning and course satisfaction. That is, learner-content interaction outside the classroom enables students to engage in more active learner-learner and learner-instructor interactions in the classroom. Thus, flipped learning should be designed in such a way that learner-content

interaction is seamlessly connected to learner-learner and learner-instructor interactions (Akçayır & Akçayır, 2018; Hwang et al., 2015).

Moreover, when designing second or foreign language (L2) flipped learning, it is critical to note that interactive practice is an essential component of the learning process in learning a language (Long, 1996). As Jiang et al. (2022b) point out, learners in L2 classrooms, whether or not they are flipped, must practice interaction in the target language, making it an "essential portion of the curriculum content" (p. 1240). However, recent reviews of the relevant literature indicate that the integration of L2 interaction into the design of the flipped classroom has not been adequately addressed, particularly in the out-of-class context. Generally, L2 scholars have simply followed general learning theories, and students have been instructed to watch educational videos or read materials without further guidance, resulting in students' inadequate preparedness for in-class L2 interaction (Jiang et al., 2022b; Zou et al., 2022). This issue is especially pronounced in EFL contexts, such as South Korea (the setting of this research), where the social environment provides few opportunities for learners to practice English speaking and listening. In such settings, relying only on unidirectional instructional materials is not sufficient to prepare students for L2 interaction in flipped classrooms.

In response to this concern, we explored a flipped approach to EFL instruction utilizing chatbot technology, a type of artificial intelligence (AI) agent capable of executing human-like interactions, to provide a simulated interlocutor with whom students could interact in English outside their classrooms (Smutny & Schreiberova, 2020), and examined its impact on learners' L2 interactions. The use of chatbot technology for language learning has recently garnered increasing attention, particularly since the release of ChatGPT (Jeon & Lee, 2023; Kohnke et al., 2023). Along with this development, the availability of customized chatbot-building platforms, such as Dialogflow and more recent platforms based on large language models (LLMs) like GPTs, (see OpenAI, 2023) makes it possible for users to easily build chatbots tailored to their specific needs with little to no coding knowledge (Jeon et al., 2023a). This technological advancement might also enable language teachers to create and use their own customized chatbots for pedagogical purposes.

Therefore, we posit that the integration of chatbot technology into a flipped classroom design could address the shortcomings of flipped learning such as insufficient provision of interaction opportunities before class (Jiang et al., 2022b; Zou et al., 2022). In this research, following Hew et al.'s (2023) and Lee et al.'s (2020) recommendation of using chatbots customized to meet students' learning needs, we chose to use Google's Dialogflow to create chatbots to support the instructional goals and content of the target classrooms. Thus, students could be prepared for in-class activities by not only watching videos but also engaging iteratively in structured speaking practice with chatbots.

2. Literature review

2.1. Flipped learning for EFL classrooms

A flipped classroom approach usually involves synergistically connected out-of-class and in-class phases (Akçayır & Akçayır, 2018; Förster et al., 2022), in which, as noted above, self-directed learning in the out-of-class phase replaces traditional teacher-centered instruction. The underlying rationale of flipped learning is that by giving students time to gather information and complete assignments outside of class, usually with the aid of technology, teachers can engage them in meaningful interactions and active learning experiences during class (Bishop & Verleger, 2013). In the field of language learning, researchers have found positive effects of flipped classrooms on the development of specific language skills, including oral presentation, writing, vocabulary, and grammar (Chen & Hwang, 2020; Hung, 2017; Wu et al., 2017, 2020), and on affective domains, such as motivation, confidence, and willingness to communicate (Chen Hsieh et al., 2017; Hung, 2017; Wu et al., 2020).

While these studies demonstrate the efficacy of flipped learning as a pedagogical approach to language learning, Jiang et al. (2022b) pointed out that the literature on the design of the flipped approach, particularly the out-of-class phase, does not fully address the interactional aspect of language learning. This indicates significant room for improvement in terms of innovative flipped learning design. In line with Zou et al. (2022), Jiang et al. further argued that the technology examined in flipped research has been mostly one-way, such as online videos and quizzes, which mainly serve the purpose of "simplistic preparedness checking" (p. 1244). This form of pre-class work neither supports the interactive nature of L2 learning nor sufficiently prepares students for active L2 interaction in the classroom (Long, 1996). This highlights the need for research in L2 flipped learning, where the interactive nature of L2 learning is pursued during the out-of-class phase as well.

To address this concern, some researchers have investigated ways to provide L2 interaction experiences in the out-of-class phase of flipped learning using social media platforms (e.g., Lin & Hwang, 2018; Wu et al., 2020). Lin and Hwang, for example, found that participating in an online community on Facebook in which they watched videos and shared comments in the target language before class helped university-level EFL students improve their oral English performance. Similarly, Chen Hsieh et al. (2017) investigated the use of LINE to encourage Taiwanese university English learners in a mobile-assisted flipped course to share their oral recordings and receive feedback from their teacher and peers before class. Results of pre- and post-tests of English idioms and questionnaires showed positive effects on students' English acquisition of idiomatic English and motivation. Tseng et al. (2018) encouraged secondary-level learners to engage in social interactions via Twitter before attending class. Results of pre- and post-multiple-choice tests showed significant improvement in their language skills. Overall, these studies affirm the potential benefits of using technology to expand learners' opportunities for target language interaction prior to their attendance in class.

Nonetheless, our review of the research on flipped learning in the field of L2 education has revealed two issues. First, social media platforms employed in most studies have certain limitations. In some cases, the platforms were used asynchronously outside the classroom and so failed to provide synchronous L2 interaction (e.g., Lin & Hwang, 2018). The asynchronous use of technology, although interactive to a certain extent, does not capture the highly interactive nature of language learning (Canale & Swain, 1980), which can "distinguish flipped FL/L2 education from other flipped courses (e.g., flipped STEM education)" (Jiang et al., 2022b, p. 1240). Furthermore, teachers will still need to undertake the demanding task of providing individual feedback on students' practice. Even when utilized synchronously in some cases (Chen Hsieh et al., 2017; Tseng et al., 2018; Wang & Qi, 2018), the type of technology exacerbated some of the constraints intrinsic to synchronous communications (Akçayır & Akçayır, 2018), such as the difficulty of arranging a common time block for meeting outside the classroom and managing the teacher's workload. As Wang and Qi stated, being available to provide students with assistance outside of the classroom demanded "teachers' considerable commitment" (p. 66).

Second, we found that previous research has not adequately addressed elementary school students. Only a limited number of studies have explored their perceptions of flipped learning, and research with experimental components remains notably scarce (e.g., Zou, 2020). For example, Zou investigated the perceptions of EFL elementary school students toward flipped learning. The results revealed that 95% of the participants had positive attitudes towards it, with high pre-class assignment completion rates of around 90% and improved grades. While such research demonstrates the potential of flipped learning for elementary-level EFL instruction, there remains a pressing need for more research in the elementary school setting, as also highlighted by systematic reviews on flipped language classrooms (Turan & Akdag-Cimen, 2020; Zou et al., 2022). This need is echoed by Zou as well, who stated, "most previous studies were conducted among secondary or university students, and it is uncertain whether primary students who were younger and had comparatively lower language proficiency levels can also benefit from flipped classroom" (p. 215).

Overall, the literature suggests the need for more investigations into flipped learning at earlier grade levels and for a technological solution to the problem of supporting active L2 practice and synchronous interaction activities that can capture the interactive nature of L2 learning, without placing an undue burden on teachers. By adopting a quasi-experimental research design, we aimed to address both issues. We proposed that a line of research on computer-assisted language learning (CALL) which shows that chatbots can be used to enhance L2 learning (Fryer et al., 2020; Hwang et al., 2022; Jeon, 2024) can also be applied to elementary-level L2 flipped classrooms. In the following section, we examine the evidence of the suitability of chatbot technology for flipped learning.

2.2. Chatbots for EFL learning

Research on language learning has demonstrated the importance of interactions using the target language practice on learning outcomes, such as speaking proficiency, motivation, and willingness to communicate in L2, as well as challenges in providing opportunities for such interactions (Dizon, 2020; Hwang et al., 2022). Studies of the uses of technology in language learning suggest that access to chatbot technology, now ubiquitously available (Fryer et al., 2020; Hwang et al., 2022), may provide a solution to the problem of adequate L2 interaction. With regard to flipped learning, teachers may be able to design learner-chatbot interaction activities for out-of-class phases that support active human-human interaction in the classroom.

Some studies have provided insights into the potential of using speech-based chatbots to prepare students for class participation (e.g., Dizon, 2022; Jeon, 2024; Jeon et al., 2023b). For example, Tai (2022) compared the effects of human-chatbot interactions and human-human interactions to investigate how the use of chatbot

technology outside the classroom affected the oral proficiency of EFL learners. The results showed that chatbot activities outside the classroom significantly improved oral proficiency while having a more positive effect than interaction with L2 human partners. In an investigation of the impact of chatbot technology outside the classroom on EFL students' motivation, Jeon (2022) found that participants valued the immediate feedback provided by chatbot technology and stated in interviews that they were willing to continue using the technology for language learning. Overall, the literature suggests that such affordances of chatbot technology as providing immediate feedback and creating a ubiquitous, anxiety-reduced positive learning environment may support the concept of flipped learning, particularly for bridging pre- and in-class phases, by allowing students to study at their own pace and in their own place for as much as they choose (Jeon, 2024; Jeon et al., 2023a).

Noting the potential that speech-based technology provides, a small but growing number of language researchers have been exploring a broad range of uses of speech technologies such as automatic speech recognition (ASR) and chatbots in flipped classrooms (Lin & Mubarok, 2021; Jiang et al., 2021; 2022a). For example, Jiang et al. (2021) compared the effects of out-of-class tasks with and without accompanying ASR technology on oral language complexity of university-level EFL students in China. Analysis of recordings of the students' oral discourse on a specified topic in English at four time points revealed that students who utilized ASR technology improved significantly more than those who did not. The researchers concluded that "by self-correcting the errors with the ASR-based feedback, students could familiarize themselves with in-class activities in terms of oral expressions and improve their oral performance" (p. 116). Of particular relevance to the current study, Lin and Mubarok (2021) investigated the effectiveness of combining chatbot use with mind-mapping activities with that of using chatbots alone in a five-week flipped university EFL speaking class. Analysis of three voice recordings of English oral performance and student-chatbot interaction records produced during the course revealed that students using chatbots with mind-map activities demonstrated more interactive behavior patterns when conversing with chatbots and showed greater improvement in their L2 oral presentation performance than students who used chatbots only. Besides the ubiquity and interactivity of chatbots as key benefits, this study demonstrates the enhanced efficacy of a mind-map guided AI chatbot approach.

Despite these contributions, gaps remain in research concerning speech technology-based flipped learning. First, previous research has not included the assessment of participants' interactive L2 oral performance using a pre-post experimental design; thus, it has not offered direct insights into how the use of technology affects interactive speaking skills. For example, Lin and Mubarok (2021) evaluated recordings of students' solo, non-interactive performances on specific topics. Jiang and colleagues (2021; 2022a) collected and analyzed in-class peer-to-peer interaction data at different time points, but without a pre-post research design. Second, the impact of speech technology on various types of interactions that can occur in a flipped course has not been comprehensively addressed, while each study tends to focus only on certain types of interactions. For example, Lin and Mubarok analyzed students' interactive behaviors during conversations with chatbots in a flipped EFL course. In another flipped classroom design, where EFL learners completed an additional pre-class task using ASR technology, Jiang and colleagues (2021; 2022a) examined learner-learner interactions in the classroom setting. A comprehensive understanding of how the use of chatbot technology affects the different types of interactions that occur in flipped learning—including learner-learner, learner-instructor, and learner-content interactions (Lin et al., 2023)—will help determine how we can design more effective flipped learning environments with this technology. Third, as Zou et al. (2022) noted, research on flipped classrooms for language learning in elementary education settings is scarce, and studies on speech technology-based flipped learning at this level are even scarcer, thus making it uncertain whether elementary students can also benefit from it. We have addressed these three gaps in the literature by investigating the impact of using chatbots in a flipped EFL elementary classroom on learners' interactive speaking performance and how perceptions of different types of interactions, including learner-learner, learner-instructor, and learner-content interactions, are influenced by the use of chatbots. The following research questions (RQs) guided this study:

- RQ1. Does a chatbot-assisted flipped approach have a significant effect on EFL learners' interactive speaking performance?
- RQ2. Does a chatbot-assisted flipped approach have a significant effect on EFL learners' perceptions of interaction in the language course?
- RQ3. What are EFL learners' overall experiences with and perceptions of a chatbot-assisted flipped approach, and how are these affected by the use of chatbots?

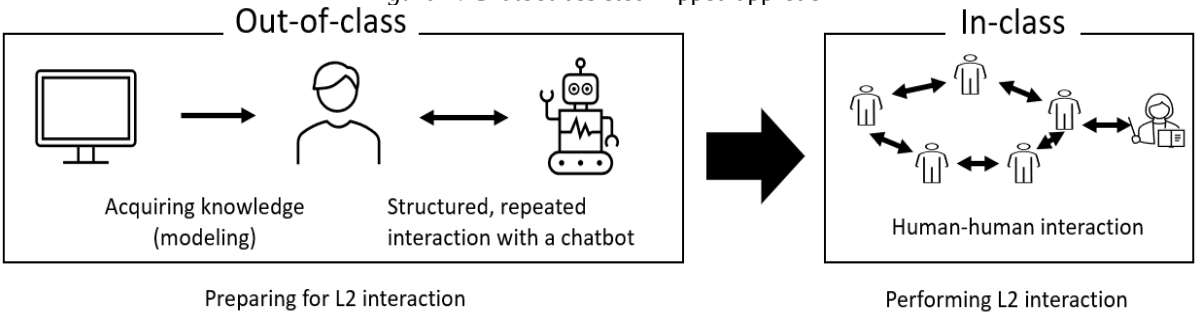
3. Theoretical framework for chatbot use in an L2 flipped classroom

Our investigation of the use of chatbots in an L2 flipped classroom is based on skill acquisition theory (Dekeyser, 2010) and a multi-faceted approach to the use of chatbots (Jeon, 2024). According to skill acquisition theory, students can progress from effortful to more automatic L2 use through modeling and repeated practice with timely feedback (Lyster & Sato, 2013). As DeKeyser (2010) highlighted, this principle is especially applicable in beginner-level EFL courses, a stage at which students require deliberate practice in turning declarative language knowledge into language performance (DeKeyser, 2010). According to this theoretical perspective, simply modeling language use in videos viewed before a class is insufficient preparation for in-class activities. Rather, video-viewing needs to be supplemented with opportunities for students to practice L2 interaction paired with feedback in a structured and iterative manner (Jiang et al., 2022b; Zou et al., 2022).

However, students in the EFL context lack opportunities to interact with others in the L2, particularly outside the classroom (Fryer & Carpenter, 2006). To address this issue, research suggests that learners can take advantage of the ubiquity of the technology and use chatbots as conversational partners while still experiencing many of the features that human-human interaction provides (Bibiuw et al., 2019; Tai, 2022). In fact, as Fryer and Carpenter pointed out, some learners, particularly beginner EFL learners, may prefer chatbots over human interlocutors because "chatbots are willing to repeat the same material with students endlessly; they do not get bored or lose their patience" (p. 10). In addition, they provide consistent feedback to learners (Hwang et al., 2022; Jeon, 2022, 2024). Thus, chatbot technology may provide the opportunities for practice prescribed by skill acquisition theory, particularly during the pre-class phase of a flipped classroom.

Meanwhile, an effective instructional design for using chatbots to support language learning is a multifaceted approach to chatbot use, where the combination of both human-chatbot and human-human interactions is pursued (Jeon, 2024). In the field of CALL, research has demonstrated that providing supplementary interaction opportunities with chatbots is more effective in both cognitive and affective domains than relying exclusively on human interlocutors (e.g., Dizon, 2020; Tai, 2022). For example, Dizon (2020) found that students who engaged in interaction with chatbots as part of classroom activities scored higher on speaking tests than their counterparts who interacted only with human interlocutors. Jeon (2024) suggested that this approach, besides encouraging students to have a favorable perception of using chatbots, can lower their speaking anxiety during subsequent human-human interactions, which is essential for active L2 interaction between human interlocutors (Jiang & Dewaele, 2020). Thus, for our research we adopted a multifaceted approach to the use of chatbots to support language learning that combines both human-chatbot and human-human interactions (Jeon, 2024; Jeon & Lee, 2023). We also posited that a multifaceted approach to the use of chatbots might be especially effective with elementary school students, who, as Wang et al. (2023) found, tended to use AI ineffectively for learning on their own. They observed that many elementary EFL students interacted with AI chatbots passively and used them for the repetition and memorization characteristic of rote learning. In line with a multifaceted approach to chatbot use, Wang et al. called for an approach in which human teachers and peers play a complementary role in students' learning with AI, ensuring that learners actively use the knowledge gained from interactions with chatbots in meaningful human-human communication. The chatbot-assisted flipped approach suggested in the current study (see Figure 1) may address this concern.

Figure 1. Chatbot-assisted flipped approach



4. Research design

4.1. Participants

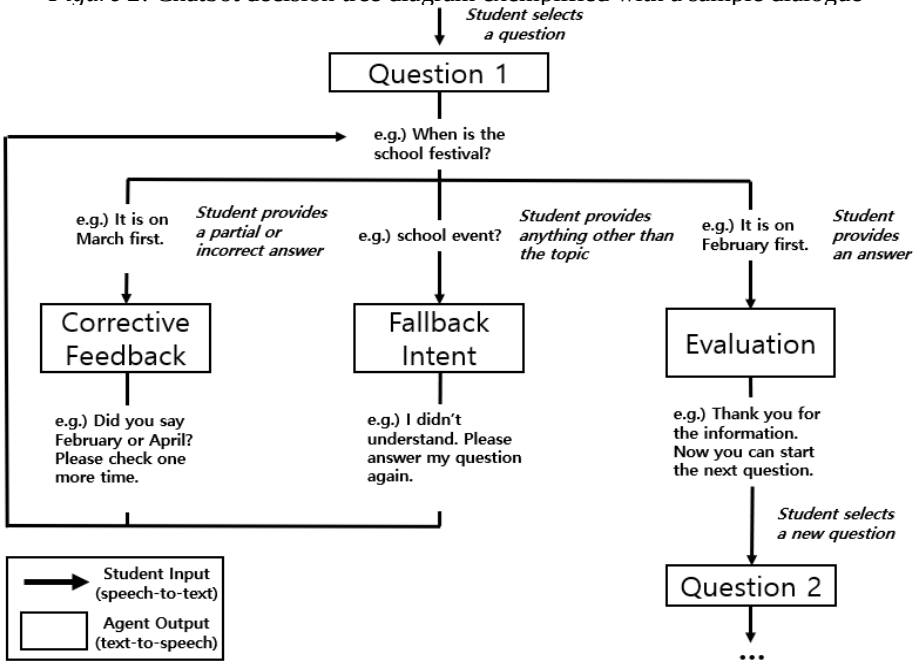
This study was conducted in an English course in a public elementary school in South Korea. Participants were recruited through convenience sampling. The research team outlined the study's protocol to students and their parents and obtained their consent. Consequently, 87 EFL students aged 11 to 12 in four intact classes participated in this study. At the time of this research, the students had experienced about three to four years of formal EFL learning. Also, all were confirmed to be familiar with online language learning, and most (n = 84) had experienced one semester of flipped learning. The teacher, who had eight years of EFL teaching experience and had been teaching the participants for more than a year at the time of this study, was also recruited. We confirmed that the teacher had adequate knowledge of and previous experience with flipped learning.

The four classes were randomly divided into two groups: two constituted the experimental group (n = 42, male = 23, female = 19), who utilized chatbots after video-viewing for their flipped course, and the other two constituted the control group (n = 45, male = 25, female = 20), who used worksheets for speaking practice instead of chatbots. A preliminary diagnostic assessment was administered to determine the participants' English proficiency levels and prepare appropriate proficiency measurement methods. Each student engaged in a 10-minute conversation with the teacher that included exchanging greetings, expressing emotions, and discussing hobbies. The teacher rated each student's English proficiency according to the guidelines of the Common European Framework of Reference for Languages (CEFR) and reported that the majority (n = 85) were at the A1 level. Two students, one from each group, were found to be at the A2 level. The teacher further elaborated that all participants were able to converse in English but with some mistakes and hesitations to recall vocabulary items. Regarding prior chatbot use experience for language learning, the teacher stated that she had used conversational agent systems with the participants more than 20 times before the study, indicating that both the teacher and the participants were already familiar with chatbot technology for language learning.

4.2. Flipped learning environment and chatbot design

E-Hakseupteo, an online education management platform developed by the Korean Ministry of Education was used in this study. By logging onto the website, the students could access the course and start online learning activities. The online learning materials were posted on the course bulletin board, and the students in both groups could watch around 15-minute videos and download relevant materials on their own before attending offline classes. The video contents consisted of (1) the introduction of target vocabulary and sentences and (2) follow-up explanations regarding grammar and examples of its use. The students were asked to understand and remember vocabulary and sentences presented in a video and guided to apply them to a subsequent task.

Figure 2. Chatbot decision tree diagram exemplified with a sample dialogue



We developed chatbots for this study using Dialogflow, a chatbot builder provided by Google, and grounded the development process in skill acquisition theory; that is, the chatbots were able to conduct structured, repeated conversations with the students regarding each topic presented in a video while providing consistent feedback on the students' utterances. We used the target dialogues provided in the course textbook as a guide and followed an approach similar to Ayedoun's et al. (2019) in designing the chatbot's dialogue flow, content, response types, and feedback, taking into account the probable usefulness of chatbot responses and the feasibility of their implementation from a technical standpoint (Ayedoun et al., 2019). As shown in Figure 2, the chatbots were specifically designed to provide corrective feedback, fallback intent, and evaluation. Depending on the type of student input, interaction continued until a target response was provided to the chatbots. The chatbots were integrated with the Google Assistant interface into tablet PCs, which provided speech-to-text and text-to-speech technologies. The research team distributed the devices to the experimental group before the commencement of the experiment. During the experimental process, the teacher regularly reminded both groups of students of their pre-class learning to ensure that the students were prepared for in-class learning.

4.3. Measuring tools

To address RQ1, which concerns the students' interactive speaking performances, we compared the results of pre- and post-interactive speaking tests. Adopting the formats used in Divekar et al. (2022), we developed and administered semi-structured interactive speaking tests to both groups (see Appendix 1 for a sample). At a designated quiet area of the school, one of us conducted a three-part conversation (interview, role-play, and interview based on the role-play) with each participant on the three topics covered in the treatment. The same topics were utilized in both the pretest and posttest but in a different order. The test took around 10-15 minutes per individual, and each conversation was recorded. To evaluate the participants' interactive speaking performance, we employed the rating scale for the elementary level of the General English Proficiency Test (GEPT), which is equivalent to level A2 on the CEFR (see Appendix 2). Two raters evaluated the recorded files using a six-band scoring scale of 0 to 20 for the following five metrics: fluency, appropriateness of content, pronunciation, vocabulary, and grammar. The individual proficiency scores were then averaged. Inter-rater reliability was confirmed to be high for both pre- and post-test sessions (kappa = 0.81, 0.84). An analysis of covariance (ANCOVA) was implemented to compare the experimental and control groups.

To address RQ2, which concerns perceptions of interactions in the course, we adopted Lin's et al. (2022) interaction questionnaire. Two EFL professionals confirmed the content validity of the instrument. We then translated the questionnaire into the students' L1, piloted it with five students of the same age as the participants, and finalized it by taking their comments into account. The questionnaire consisted of 13 items covering interaction types that can occur in a flipped course (see Appendix 3): three learner-content interaction items, five learner-instructor interaction items, and five learner-learner interaction items. All items used a seven-point Likert scale ranging from strongly disagree (1) to strongly agree (7). Internal reliability was determined by examining Cronbach's alpha scores for each construct, all of which were verified to be higher than 0.80. The students completed the questionnaire in their classrooms after the treatment sessions.

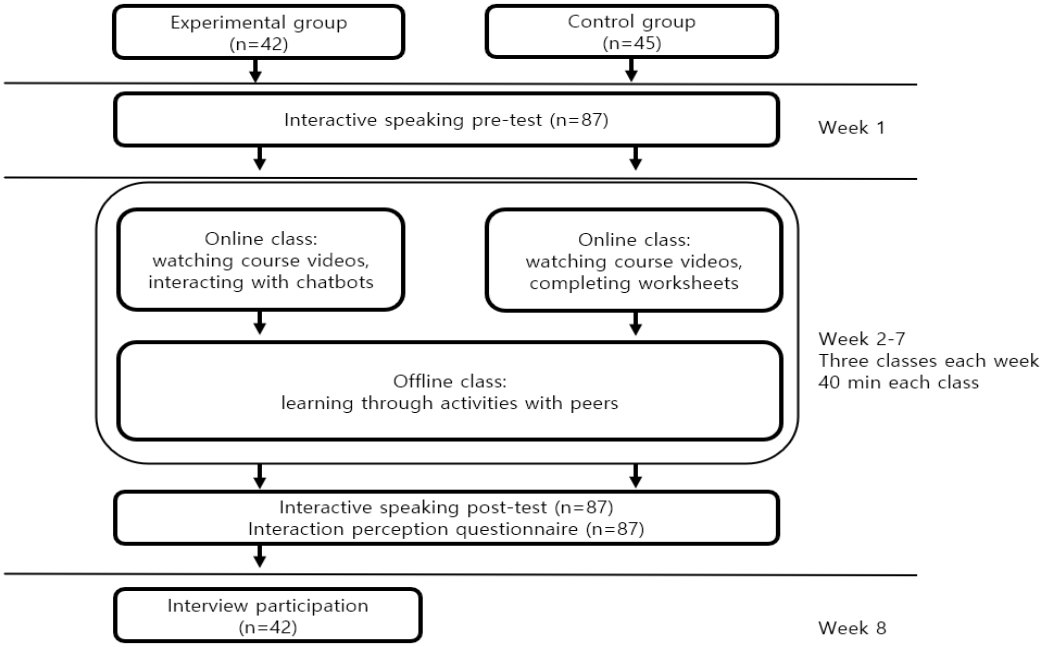
Last, to answer RQ3 concerning students' overall experiences with and perceptions of the chatbot-assisted flipped approach, a week after the last treatment session, a semi-structured interview was conducted with each participant in the experimental group (n = 42) at a designated quiet place of the school (see Appendix 4 for sample interview questions). The interview protocol began with broad questions regarding the course. Then, follow-up questions were focused on how the use of chatbots affected the interviewee's experiences and perceptions. The interviews, which lasted between 8 and 15 minutes, were audio-recorded and transcribed verbatim. We analyzed the contents of the transcripts using consensus coding to identify repeated patterns of meaning related to how the chatbots affected students' experiences and perceptions of flipped learning (Braun & Clarke, 2021). As a result, three themes were identified as will be discussed in the subsequent sections.

4.4. Procedure

Figure 3 shows the experimental procedure of this study. Before the flipped course started, the experimental and control groups took an interactive speaking pre-test, which was different from the diagnostic test. During the six-week EFL course, both groups followed the same sequence of first participating in pre-class activities and then attending in-class activities. During the period, students in both groups followed the sequence three times a week, resulting in a total of 18 in-class 40-minute sessions. Three topics were taught following the communicative language teaching (CLT) method, each topic for two weeks. As this study was conducted as part of the public elementary school's English course, the topics were selected based on the school's English

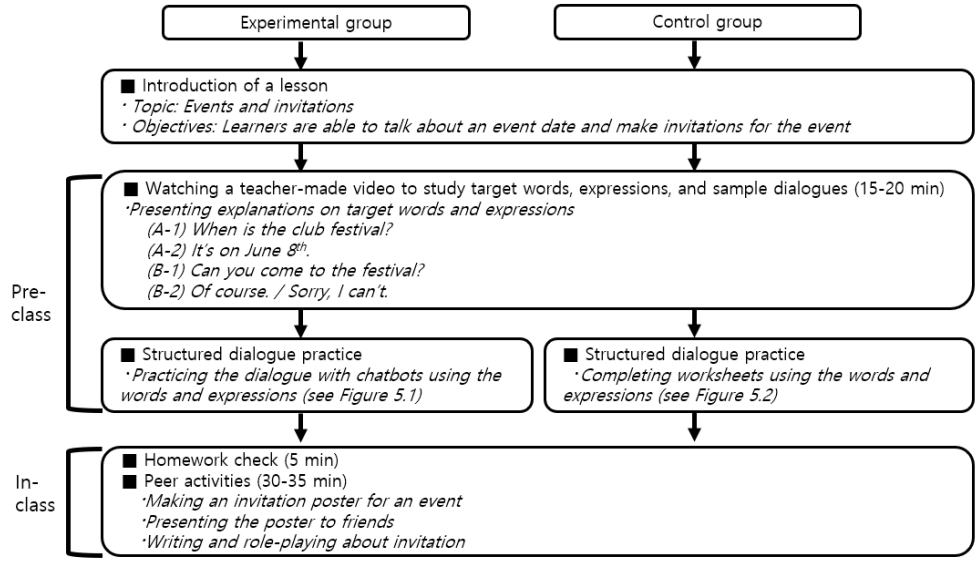
curriculum. Although most of the participants could talk about the topics in English at a basic level, as confirmed by the diagnostic test, they made some mistakes and had difficulty maintaining fluency. As the primary goal of the intervention was to help students improve their current English-speaking proficiency, the day following the last session, all students took an interactive speaking post-test and completed an interaction questionnaire. The experimental group participated in semi-structured interviews one week after the last session.

Figure 3. Study design



Overall, regarding lesson plans, the two groups followed the same flipped instructional sequence, learning content, and learning objectives; however, the format of the materials used during the pre-class phase differed. Figure 4 provides a flipped learning lesson sample used in the study. Regarding lesson plans for both groups, pre-class instruction was conducted through the use of technology and involved individual work, which included understanding and remembering knowledge acquired through video-viewing and applying that knowledge to an additional task. Specifically, teacher-made videos were used, and students in both experimental and control groups were introduced to target words, sentences, and model dialogues.

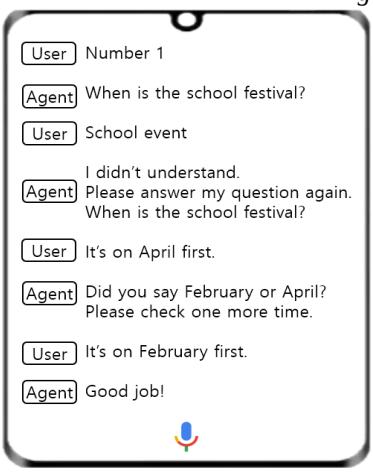
Figure 4. Flipped learning lesson sample



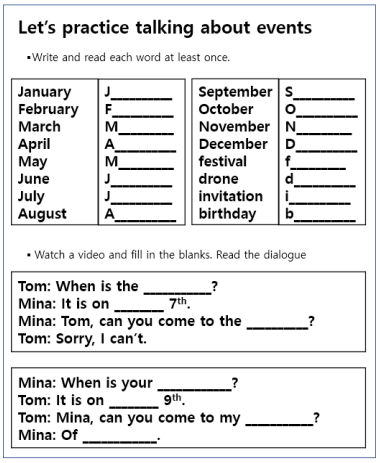
After watching a teacher-made video, different formats of material were used for each group (see Figure 5). To be specific, students in the experimental group were required to perform L2 dialogue practice with chatbots at least once, using English expressions presented in the video. Students in the control group were required to

complete speaking practice by using a worksheet, which involved reading the dialogues aloud at least once to prepare for spoken interactions with peers and the teacher in the classroom. Despite the difference in format, both activities aimed to provide structured speaking practice with a focus on preparing for spoken interactions that would occur in the subsequent face-to-face classroom. In-class instruction that was provided for both groups after pre-class teaching comprised in-person peer activities, such as role-playing, discussion, and peer assessment. The students used the knowledge that they acquired before class when participating in these activities.

Figure 5. Pre-class task format



(a) A snapshot of the chatbot interface



(b) A sample worksheet for the control group

5. Results

5.1. RQ1. Does a chatbot-assisted flipped approach have a significant effect on EFL learners' interactive performance?

ANCOVA was used to analyze the differences between the interactive speaking tests of the two groups using the pre-test scores as the covariate and the post-test scores as the dependent variable, which enabled comparison of the post-test results by excluding the impact of the pre-test scores. Before conducting ANCOVA, we implemented an independent *t*-test to confirm whether the two groups had similar interactive speaking skills. As Table 1 shows, no significant difference was found between the two groups' interactive speaking pre-test results with $t = 1.54$ ($p > .05$), indicating that the two groups were homogeneous in terms of interactive speaking performance prior to the experimental treatment.

Table 1. Descriptive statistics for the interactive speaking pre-test

	Group	*N*	Mean	*SD*	*t*
Pre-test	Experimental group	42	7.98	2.74	1.54
	Control group	45	7.00	3.17	

The test of homogeneity shows that the regression slopes of the two groups' interactive speaking performance tests are homogeneous with $F = 0.57$ and $p > .05$, indicating the suitability of the data analysis for using ANCOVA. The ANCOVA results for the two groups' interactive speaking tests are shown in Table 2, which reveals a significant statistical difference between the two groups' interactive speaking performance with $F = 7.84$ ($p < .01$). Specifically, participants who engaged in chatbot-assisted interaction practice before physically attending class (adjusted mean = 11.07) significantly outperformed those who followed a traditional flipped approach (adjusted mean = 9.49).

Table 2. Analysis of covariance for interactive speaking performance

	Group	*N*	Mean	*SD*	Adjusted mean	Std. error	*F*	η^2_p
Post-test	Experimental group	42	11.1	2.44	11.07	0.41	7.84^{**}	0.08
	Control group	45	9.49	2.80	9.49	0.39		

Note. $^{**}p < .01$.

5.2. RQ2. Does a chatbot-assisted flipped approach have a significant effect on EFL learners' perceptions of interaction in the language course?

The questionnaire used in the study consisted of three interaction constructs: learner-learner, learner-instructor, and learner-content. Independent *t*-tests were administered to examine the impact of the chatbot-assisted flipped approach on students' perceptions of the three patterns of interactions (see Table 3).

The results showed significant differences between the two groups in perceptions of the degree of learner-learner and learner-instructor interactions with $t = 2.36$ ($p < .05$), and $t = 2.19$ ($p < .05$), respectively, indicating that the experimental group perceived a greater degree. In contrast, no significant difference was found between the two groups in perceptions of learner-content interaction although the mean score was higher for the experimental group.

Table 3. Descriptive statistics for three interaction types by group

	Experimental group			Control group			
Constructs	*N*	Mean	*SD*	*N*	Mean	*SD*	*t*
Learner-learner interaction	42	5.31	1.37	45	4.64	1.25	2.36*
Learner-instructor interaction	42	5.02	1.35	45	4.60	1.30	2.19*
Learner-content interaction	42	4.90	1.41	45	4.73	1.30	0.59

Note. $^{*}p < .05$.

5.3. RQ3. What are EFL learners' overall experiences with and perceptions of a chatbot-assisted flipped approach, and how are these affected by the use of chatbots?

Through thematic analysis of the interview transcripts, three themes were identified regarding how learners in the experimental group experienced using chatbots, which offer additional insights into how the chatbots supported language learning and interactions (see Table 4).

Table 4. Interview summary

Theme	Code	Frequency
Interaction practice management	Self-paced interaction practice	24
	Repeated interaction opportunities	10
	Total	34
Enhanced motivation for interaction	Enjoyable and anxiety-reduced interaction with a chatbot	32
	Enhanced motivation for interaction with human interlocutors	8
	Total	40
Enriched in-class participation	Extended time for collaborative activities	5
	Enhanced quality time to collaborate with a teacher and peers	21
	Total	26

Note. When one student made more than two statements regarding the same code, it was counted as one.

5.3.1. Theme 1. Interaction practice management

A frequently mentioned theme ($n = 34$) was the management of interaction practice made possible by the use of chatbots, specifically self-pacing ($n = 24$) and unlimited repetition ($n = 10$). With regard to self-pacing, the students stated that having the chatbots to use at any time enabled them to practice according to their own schedules and to determine the amount to be learned at each session and how much practice they needed. For example, S2 stated, "I could check if I was doing well with the chatbot because the chatbot said that I was wrong when I said something wrong. On the other hand, if I pronounced some words right, the chatbot understood my utterance and I could move forward." Also, 10 students mentioned that they were able to practice a dialogue repeatedly without anxiety about interlocutors. Specifically, S3 reported, "I did not care about making mistakes and just practiced speaking English as much as I wanted."

5.3.2. Theme 2. Enhanced motivation for interaction

The most frequently mentioned theme was enhanced motivation for interaction ($n = 40$), mainly in relation to the stress-free nature of independently practicing with non-judgmental chatbots ($n = 32$) in low anxiety-inducing

environments outside the classroom. For example, S9 commented, "Unlike times when I spoke English in front of other students, I did not need to be afraid of my friends. I did not care if I made mistakes with the chatbot." Eight students also reported that practicing with the chatbots increased their motivation to interact with their peers in the classroom. Having become familiar with the use of target expressions, they felt prepared to test their skills with others. S4 stated, "After practicing with the chatbot, I became confident …[and] wanted to find out if my friends or teacher could also understand my English. I think this is the reason why I participated more eagerly in the class…."

5.3.3. Theme 3. Enriched in-class participation

Several students (n = 26) observed that moving dialogue practice, which was repetitious and time-consuming, outside the classroom gave them more in-class time for collaborative activities with the teacher and peers. Five students stated that, unlike their experiences in previous courses, they had more opportunities to learn through interactions with their peers in class. S10 stated, "We mostly spent time with friends studying English, which I really liked about this course." In addition to the increase in shared activities as a result of the flipped design, 21 students found that coming to class prepared to engage fully in the activities enhanced the quality of their in-class participation. S6 reported his changed attitude toward collaborative activities by stating, "Before, we first had to [take class time to] memorize English [words and phrases] in order to complete activities like the Poster-drawing activity (e.g., activity in week 7). However, thanks to the chatbot, I was able to study some English expressions necessary for the activity in advance, so I was able to focus on the activity itself."

6. Discussion

The results of the pre- and post-interactive speaking tests showed that the mean scores of both groups improved, supporting the efficacy of flipped classroom design for language learning (Chen Hsieh et al., 2017; Chen & Hwang, 2020; Wu et al., 2020; Zou et al., 2022), specifically in speaking skills (e.g., Amiryousefi, 2019; Lin & Hwang, 2018; Lin & Mubarok, 2021). Furthermore, consistent with previous research on flipped learning with speech technology (Lin & Mubarok, 2021; Jiang et al., 2021), this study provides empirical evidence that incorporating chatbot technology in the flipped design can lead to greater improvement in L2 interactive speaking performance than relying on the conventional out-of-class pattern of completing worksheets after viewing videos. Furthermore, in line with investigations of different methods of providing opportunities for interaction in the target language (Jiang et al., 2022a; Lin & Hwang, 2018; Tseng et al., 2018; Wang & Qi, 2018), this study highlights the effectiveness of enabling students to practice target language interaction with a chatbot before interacting with a teacher and peers in class, demonstrating that chatbots can play a valuable pedagogical role (Fryer et al., 2020; Tai, 2022).

While both groups demonstrated improvement, the between-group comparisons of the post-test scores and the interaction questionnaire show that the experimental group outperformed the control group in the interactive speaking post-test and reported more positive perceptions of learner-learner and learner-instructor interactions. Based on student interviews reported in the section related to RQ3, we posit that these discrepancies between the groups reflect different levels of preparation achieved by each group. According to Wang and Qi (2018), the flipped classroom, and in particular the L2 flipped classroom, must encourage "mastery" and not merely "previewing" of knowledge and skills prior to class, as "learning achieved before class lays the foundation for success in class" (p. 51). Consistent with this perspective, while both groups in this study were encouraged to practice L2 dialogues before class, participants in the experimental group experienced L2 interaction in a manner highly pertinent to forthcoming classroom activities with feedback from chatbots, whereas those in the control group were mainly exposed to linguistic content and procedures without opportunities to receive feedback.

Regarding theoretical rationale, the superior outcomes of the experimental group in this study support the tenets of skill acquisition theory (DeKeyser, 2010), which proposes three essential components of language learning, modeling, repeated practice, and feedback, to achieve proceduralization (conversion of effortful to more automatic skill use) of the target language (Lyster & Sato, 2013). In our approach, modeling occurred when the students first watched a video of the teacher explaining and demonstrating L2 interaction. They then participated in repeated practice with automatic feedback from the chatbots. As illustrated in Figure 1, we proposed that the use of chatbots along with video-viewing created an environment that promoted the proceduralization process more effectively than only video-viewing or practice without feedback. The experimental group's superior performance indicated that they experienced more effective L2 learning conditions for proceduralizing the target language features than the control group.

The student interviews further confirmed that the use of chatbots prepared the students to participate more meaningfully in collaborative activities and interactions with their teacher and peers. As mentioned by S6 (see Theme 3 in Section 5), because they had already practiced the target dialogue with the chatbots and received corrective feedback, the experimental students felt well-prepared and confident in the classroom. This outcome is also supported by the results of the interaction questionnaire. As shown in Table 3, students in the experimental group reported higher levels of perceived participation in learner-learner and learner-instructor interactions in class than those in the control group. These findings can also be explained by a multifaceted approach to chatbot use (Jeon, 2024). Specifically, as stated by Jeon (2024), interacting with the chatbots might have reduced the students' L2 speaking anxiety during subsequent human-human interactions, which is a crucial element for active L2 interaction between interlocutors. This could thereby facilitate higher perceived levels of participation in learner-learner and learner-instructor interactions for the experimental group. In summary, our study confirmed that practicing the target language with chatbots can promote language development, as shown in the results for speaking performance, and provide a foundation for further benefits in subsequent human-human interactions, as indicated by the results of interaction perceptions. In addition, our approach, which features students' out-of-class use of chatbots, differs from those of the majority of previous studies on in-class use of chatbots for language learning (see Jeon et al., 2023a, b for a review of chatbot studies), thus expanding the range of pedagogical applications for chatbots.

However, learner-content interaction results did not reveal a significant difference between the two groups. It should be noted that the learner-content interaction items in the questionnaire were focused on watching videos (Lin et al., 2023), which was common to both groups, implying that this was the venue for measuring learner-content interaction; the incorporation of chatbots in the pre-phase of the flipped classroom did not significantly affect it. We present two possible explanations for this result: the particular role that the chatbots played as teaching agents and the sequence of activities (video-viewing → chatbot activity). The chatbots' role for this study was strictly confined to producing structured conversational practice with guiding feedback (Kuhail et al., 2023). As hypothesized by Lee and Jeon (2022) and Kuhail et al. (2023), chatbots that can also serve as motivational agents capable of producing affectively engaging conversation will be more likely to augment learner-content interaction, suggesting the need for further investigation of this aspect. Also, regarding the sequence of activities, we recommend a future study in which video-viewing and chatbot activities cycle back to video-viewing (video-viewing → chatbot activity → video-viewing) to see whether chatbot activities can motivate students to engage more deeply with the content presented in the videos (Fidan & Gencel, 2022).

7. Limitations

Some limitations of the current study may inform the design of future research. First, while we endeavored to ensure comparable instructional design for pre-class learning between the two groups by offering structured, independent speaking practice opportunities to prepare learners for in-class interactions, it must be acknowledged that issues of comparability still exist. The comparability issue in this study arises from the difference between the groups in whether or not interactive spoken practice opportunities were provided, and the extent to which these opportunities affected the interactive spoken tests used for the post-test. In other words, the interactive nature of chatbot-assisted pre-class tasks, compared to speaking practice with worksheets, might have given the experimental group an advantage in the interactive speaking post-test. Thus, we acknowledge that this comparability issue limits the validity and generalizability of the findings to a certain degree. In this regard, while the current study demonstrates the effectiveness of the chatbot-assisted flipped approach compared to a worksheet-based traditional flipped approach, future research might include control group conditions where interactive activities with peers, a teacher, or other interactive technologies are provided in the out-of-class phase. This will allow for fairer comparisons between interactions with and without chatbot use and provide more valid and generalizable findings.

Second, the length of the experiment and the time interval between the pretest and posttest can be considered as areas for further improvement for future research. Specifically, in this study, students participated in a total of 18 sessions that continued for six weeks. According to Bond's (2020) systematic review of previous flipped learning studies, this six-week length, as well as the number of classes, is more than in "the majority of research focused on one, two, or three classes, within one or two units of work" (p. 29). However, according to Bond's recommendation, flipped learning research with a timeframe longer than three months needs to be considered to ensure that the confirmed effect is also observed in longitudinal studies that incorporate chatbots in flipped learning. Similarly, the time interval between the pretest and posttest was only six weeks, and the same topics were covered in both tests, indicating that there might have been a practice effect. To minimize this possibility and increase the generalizability of the study, future studies should employ a longer-term experimental design of

more than three months to ensure the robustness of findings and their applicability across broader contexts. A delayed posttest should also be conducted to examine the longer-term influence of the approach.

In addition to these two major issues, some minor considerations need to be taken into account. First, interaction transcript data including learner-learner, learner-instructor, learner-chatbot interactions were not collected and should be in future studies to provide a more complete picture of how interactions took place. Second, because raters for the speaking tests determined participants' scores holistically, we were unable to analyze the results of each of the five metrics employed. As suggested in Yesilcinar (2019), future studies should delve into which specific aspects of speaking skills are most enhanced by the experimental approach.

8. Conclusion

This study of chatbot use in flipped learning presents important implications for the fields of language learning and educational technology. First, it extends research on innovative technologies to enhance flipped learning (e.g., Lin & Mubarok, 2021) by providing evidence that the use of chatbots can increase its efficacy. Second, it broadens the range of flipped learning research by adding to the sparse empirical evidence of how a flipped approach can be implemented in elementary EFL education (Zou et al., 2022). Third, this study opens diversified avenues for the use of chatbots in the field of CALL. Unlike previous studies in which chatbots were used exclusively in either the classroom or informal settings (e.g., Dizon, 2022), we demonstrated the potential of using chatbots as a link between pre-class and in-class learning in a flipped approach. Last, we demonstrated the adaptability of chatbot technology by using Google's Dialogflow to develop a chatbot customized to a public school's language curriculum (Hew et al., 2023; Lee et al., 2020) and providing a theoretically-based rationale for adapting it to flipped learning.

Based on the findings, we suggest the following recommendations for researchers and teachers. For researchers, having established the adaptability of chatbot technology, we recommend further investigation into various, accessible forms of this technology to support its integration into educational settings. For example, as user-friendly chatbot-building platforms based on LLMs, such as GPTs, become increasingly available and thus make developing customized chatbots much easier (Jeon et al., 2023a, b), another important avenue for future inquiry is investigating ways to incorporate customized LLMs into a flipped approach and their impact on students' in-class learning. Such explorations, in turn, suggest the need for future research focused on developing and examining teacher education programs on the technical and pedagogical aspects of chatbot creation and use (Jeon et al., 2023a, b; Kohnke et al., 2023). Additionally, this study demonstrates only one method of applying a multifaceted approach to chatbot use in the context of flipped learning. Therefore, future research should explore various ways of utilizing chatbots in collaboration with human-human activities, both within and beyond the context of flipped learning (Jeon, 2024; Lee et al., 2023). Next, for teachers, this research highlights the potential benefits of chatbots, such as ubiquity and automatic feedback, and how they can be utilized in a flipped course. The use of chatbot technology for flipped learning will help minimize the labor intensity required of teachers, especially for pre-class learning. Beyond the context of flipped learning, it is also recommended that when using chatbots, teachers should incorporate them alongside opportunities for interaction with peers.

Declaration

We have no known conflict of interest to disclose and copyrighted material has not been used in this submission.

References

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education, 126*, 334-345.

Amiryousefi, M. (2019). The incorporation of flipped learning into conventional classes to enhance EFL learners' L2 speaking, L2 listening, and engagement. *Innovation in Language Learning and Teaching, 13*(2), 147-161

Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education, 29*, 29-57.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research September, 79*(3), 1243-1289.

Bishop, J. L., & Verleger, M. A. (2013, June). The flipped classroom: A survey of the research. *Paper presented at the 120th ASEE annual & exposition*. Atlanta, USA.

Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A systematic review. *Computers & Education, 151*, 103819.

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide.* SAGE Publications.

Bredow, C. A., Roehling, P. V., Knorp, A. J., & Sweet, A. M. (2021). To flip or not to flip? A meta-analysis of the efficacy of flipped learning in higher education. *Review of Educational Research, 91*(6), 878-918.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Chen Hsieh, J. S., Wu, W. C. V., & Marek, M. W. (2017). Using the flipped classroom to enhance EFL learning. *Computer Assisted Language Learning, 30*(1-2), 1-21.

Chen, M. R. A., & Hwang, G. J. (2020). Effects of a concept mapping-based flipped learning approach on EFL students' English speaking performance, critical thinking awareness and speaking anxiety. *British Journal of Educational Technology, 51*(3), 817-834.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE.

DeKeyser, R. (2010). Practice for second language learning: Don't throw out the baby with the bathwater. *International Journal of English Studies, 10*(1), 155-165.

Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J. A., & Braasch, J. (2022). Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. *Computer Assisted Language Learning, 35*(9), 2330-2360.

Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology, 24*(1), 16-26.

Dizon, G., Tang, D., & Yamamoto, Y. (2022). A case study of using Alexa for out-of-class, self-directed Japanese language learning. *Computers & Education: Artificial Intelligence, 3*, 100088.

Fidan, M., & Gencel, N. (2022). Supporting the instructional videos with chatbot and peer feedback mechanisms in online learning: The effects of learning performance and intrinsic motivation. *Journal of Educational Computing Research*. Advance online publication. https://doi.org/10.1177/07356331221077901

Flipped Learning Network. (2014). *The four pillars of F-L-I-P*. http://flippedlearning.org/definition-of-flipped-learning

Förster, M., Maur, A., Weiser, C., Winkel, K. (2022). Pre-class video watching fosters achievement and knowledge retention in a flipped classroom. *Computers & Education, 179*, 104399.

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology, 10*(3), 8-14.

Fryer, L., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology, 24*(2), 8-22.

Hew, K.F., Huang, W., Du, J., & Jia, C. (2023). Using chatbots to support student goal setting and social presence in fully online activities: learner engagement and perceptions. *Journal of Computing in High Education, 35*, 40-68.

Hung, H. T. (2017). Clickers in the flipped classroom: Bring your own device (BYOD) to promote student learning. *Interactive Learning Environments, 25*(8), 983-995.

Hwang, G. J., & Chen, P. Y. (2023). Effects of a collective problem-solving promotion-based flipped classroom on students' learning performances and interactive patterns. *Interactive Learning Environments, 31*(5), 2513-2528.

Hwang, G.-J., Lai, C.-L., & Wang, S.-Y. (2015). Seamless flipped learning: A mobile technology-enhanced flipped classroom with effective learning strategies. *Journal of Computers in Education, 2*(4), 449-473.

Hwang, W.-Y., Guo, B.-C., Hoang, A., Chang, C.-C., & Wu, N.-T. (2022). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2095406

Jeon, J. (2022). Exploring a self-directed interactive app for informal EFL learning: A self-determination theory perspective. *Education and Information Technologies, 27*, 5767-5787.

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.

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies, 28*(12), 15873-15892.

Jeon, J., Lee, S., & Choe, H. (2023a). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education, 206*, 104898.

Jeon, J., Lee, S., & Choi, S. (2023b). A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2023.2204343

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian Journal of Educational Technology, 37*(2), 110-131.

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2022a). Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom. *Journal of Computer Assisted Learning, 39*, 125-140.

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., Liu, K. S. X., & Park, M. (2022b). A scoping review on flipped classroom approach in language education: Challenges, implications and an interaction model. *Computer Assisted Language Learning, 35*(5-6), 1218-1249.

Jiang, Y., & Dewaele, J. -M. (2020). The predictive power of sociobiographical and language variables on foreign language anxiety of Chinese university students. *System, 89*, 102207.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537-550.

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973-1018.

Lee, J. H., Shin, D., & Noh, W. (2023). Artificial intelligence-based content generator technology for young English-as-a-foreign-language learners' reading enjoyment. *RELC Journal*, *54*(2), 508-516.

Lee, J. H., Yang, H., Shin, D., & Kim, H. (2020). Chatbots. *ELT Journal*, *74*(3), 338-344.

Lee, S., & Jeon, J. (2022). Visualizing a disembodied agent: young EFL learners' perceptions of voice controlled conversational agents as language partners. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2067182

Lin, C. J., & Hwang, G. J. (2018). A learning analytics approach to investigating factors affecting EFL students' oral performance in a flipped classroom. *Education Technology & Society*, *21*(2), 205-219.

Lin, C., & Mubarok, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Educational Technology & Society, 24*(4), 16-35.

Lin, G. Y., Wang, Y. S., & Lee, Y. N. (2023). Investigating factors affecting learning satisfaction and perceived learning in flipped classrooms: the mediating effect of interaction. *Interactive Learning Environments*, *31*(9), 5759-5780.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of research on language acquisition* (pp. 413-468). Academic Press.

Lyster, R., & Sato, M. (2013). Skill acquisition theory and the role of practice in L2 development. In M.G. Mayo, J. Gutierrez-Mangado, & M.M. Adrian (Eds.), *Contemporary approaches to second language acquisition* (pp. 71-92). John Benjamins.

Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education, 3*(2), 1-7.

OpenAI. (2023, December 15). *Introducing GPTs*. OpenAI. https://openai.com/blog/introducing-gpts

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook messenger. *Computers & Education, 151*, 103862.

Tai, T. Y. (2022). Effects of intelligent personal assistants on EFL learners' oral proficiency outside the classroom. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2075013

Tseng, M., Lin, C. H., & Chen, H. (2018). An immersive flipped classroom for learning Mandarin Chinese: Design, implementation, and outcomes. *Computer Assisted Language Learning, 31*(7), 714-733.

Turan, Z., & Akdag-Cimen, B. (2020). Flipped classroom in English language teaching: A systematic review. *Computer Assisted Language Learning, 33*(5-6), 590-606.

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li. L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers and Education. 194*, 104703.

Wang, Y., & Qi, G. Y. (2018). Mastery-based language learning outside class: Learning support in flipped classrooms. *Language Learning & Technology, 22*(2), 50-74.

Wu, W. C. V., Hsieh, J. S. C., & Yang, J. C. (2017). Creating an online learning community in a flipped classroom to enhance EFL learners' oral proficiency. *Educational Technology & Society, 20*(2), 142-157.

Wu, W. C. V., Yang, J. C., Chen Hsieh, S. J., & Yamamoto, T. (2020). Free from demotivation in EFL writing: The use of online flipped writing instruction. *Computer Assisted Language Learning, 33*(4), 353-387.

Yesilçinar, S. (2019). Using the flipped classroom to enhance adult EFL learners' speaking skills. *PASAA: Journal of Language Teaching and Learning in Thailand, 58*, 206-234.

Zou, D. (2020). Gamified flipped EFL classroom for primary education: Student and teacher perceptions. *Journal of Computers in Education, 7*(2), 213-228.

Zou, D., Luo, S., Xie, H., & Hwang, G. J. (2022). A systematic review of research on flipped language classrooms: Theoretical foundations, learning activities, tools, research topics and findings. *Computer Assisted Language Learning, 35*(8), 1811-1837.

Appendix 1. Interactive speaking test sample (adopted from Divekar et al., 2022)

1. Interview
Interviewer: 식당의 위치를 묻고 싶을 때, 어떻게 영어로 물을 수 있나요?
(What would you say when you ask for directions to a restaurant?)
Student: Where is the restaurant?

2. Role-play
Interviewer: Hi, I am looking for the school, where is the school?
Student: (looking at provided map) Go straight and turn right.
Interviewer: How many blocks?
Student: Two blocks, and it is on your left.
Interviewer: Thank you.

3. Interview based on the role-play
Interviewer: 어디로 가는 길을 설명했나요?
(Where did you give directions to?)
Student: School.
Interviewer: 지도에 어떤 건물들이 있나요?
(What buildings are on the map?)
Student: Restaurant and bookstore.
Interviewer: 학교는 어떤 건물 앞에 있나요?
(What building is the school in front of?)
Student: It's in front of the bookstore.

Note. Participants were encouraged to use English throughout the tests.

Appendix 2. Rating scale

Band	Raw score	Description
5	17-20	▪Pronunciation and intonation correct and natural ▪Response is fluent and communication unhindered ▪Content answers the question ▪Student has a generally good command of basic grammar and vocabulary
4	13-16	▪Pronunciation and intonation generally correct and natural (although there are errors, these do not impede comprehension) ▪Response can still be considered fluent, and communication is unhindered ▪Content generally answers the question ▪Basic grammar and vocabulary are generally correct but still somewhat restrained.
3	9-12	▪Some errors in pronunciation and intonation, but still intelligible ▪Student speaks slowly and sometimes hesitates, but can still communicate ▪Much of the content is difficult to comprehend ▪Frequent grammatical errors and limited vocabulary hinder expression
2	5-8	▪Numerous errors in pronunciation and intonation affect comprehension ▪Slow speech and frequent hesitation impede communication ▪Content difficult to understand ▪Many grammatical errors; speaks largely in phrases; insufficient vocabulary
1	1-4	▪Many errors in pronunciation and intonation ▪Many inappropriate hesitations ▪Responses difficult to understand ▪Almost no sentence structure; vocabulary severely lacking
0	0	▪Unanswered or irrelevant response

Appendix 3. Interaction perception questionnaire (adopted from Lin et al., 2023)

▪Learner-content interactions (seven-point Likert scale)
1. I scheduled watching the video on the flipped learning platform.
2. I included watching the video on the flipped learning platform in my study plan.
3. I finished watching the video in the flipped learning platform without needing reminders or urging from my teacher.

▪Leaner-instructor interactions (seven-point Likert scale)
1. When I didn't understand something in the video, I asked my teacher in the flipped course.
2. When I needed help, my teacher did their best to help me in the flipped course.
3. My teacher actively assisted me in solving my flipped-course-related learning difficulties.
4. My teacher did their best to teach me the course content that I did not understand.
5. My teacher discussed with me to solve the problems I encountered in the flipped course.

▪Learner-learner interactions (seven-point Likert scale)
1. In the flipped course, I participated in course discussions.
2. In the flipped course, I shared ideas and shared information with other classmates.
3. In the flipped course, other classmates shared ideas and information with me.
4. There were many opportunities in the flipped course to interact with other classmates.
5. I actively asked other classmates for help when I had difficulties and problems in the flipped course.

Appendix 4. Semi-structured interview questions (adopted from Jeon, 2024)

1. Can you describe how you studied English before coming to class?
2. Did you find this course helpful for improving your English skills? How?
3. What difficulties did you have when studying English outside the classroom during the course?
4. Can you describe how you used the chatbots during the course?
5. What were the benefits of using chatbots before class?
6. What difficulties did you have when studying with chatbots and how did you overcome the difficulties?