Language Learning & Technology
ISSN 1094-3501 CC BY-NC-ND

June 2024, Volume 28, Issue 2
pp. 62–90

ARTICLE

Using chatbots to support EFL listening decoding skills in a fully online environment

Weijiao Huang, The University of Hong Kong

Chengyuan Jia, Zhejiang University*

Khe Foon Hew, The University of Hong Kong

Jia Guo, Northwest University of Political Science and Law

**Chengyuan Jia is the corresponding author*

Abstract

Aural decoding skill is an important contributor to successful EFL listening comprehension. This paper first described a preliminary study involving a 12-week undergraduate flipped decoding course, based on the flipped SEF-ARCS decoding model. Although the decoding model (N = 44) was significantly more effective in supporting students' decoding performance than a conventional decoding course (N = 36), two main challenges were reported: teacher's excessive workload, and high requirement for the individual teacher's decoding skills. To address these challenges, we developed a chatbot based on the self-determination theory and social presence theory to serve as a 24/7 conversational agent, and adapted the flipped decoding course to a fully online chatbot-supported learning course to reduce the dependence on the teacher. Although results revealed that the chatbot-supported fully online group (N = 46) and the flipped group (N = 43) performed equally well in decoding test, the chatbot-supported fully online approach was more effective in supporting students' behavioral and emotional engagement than the flipped learning approach. Students' perceptions of the chatbot-supported decoding activities were also explored. This study provides a useful pedagogical model involving the innovative use of chatbot to develop undergraduate EFL aural decoding skills in a fully online environment.

Keywords: *Chatbot, EFL Listening, Fully Online Learning, Student Engagement*

Language(s) Learned in This Study: *English*

APA Citation: Huang, W., Jia, C., Hew, K. F., & Guo, J. (2024). Using chatbots to support EFL listening decoding skills in a fully online environment. *Language Learning & Technology*, *28*(2), 62–90. https://hdl.handle.net/10125/73572

Introduction

Listening is the process of "making sense of spoken language, normally accompanied by other sounds and visual input" (Lynch & Mendelsohn, 2010, p. 180). Listening is the primary means by which learners enlarge their knowledge of the spoken forms of the target language (Field, 2008). Yet although listening plays a vital role in daily communication, it is often considered a very difficult skill for EFL learners (Siegel, 2014) due to EFL learners' weak ability to recognize words (Wong et al., 2017), especially in natural speech. Words in natural speech tend to blend into each other compared to words spoken in isolation which can exacerbate the difficulty for EFL learners to recognize words in real-life communication (Jia & Hew, 2021a).

Listening comprehension involves a combination of top-down and bottom-up processing (Vandergrift, 2004). Top-down processing comes into play when listeners utilize prediction, inference, and contextual skills based on their prior knowledge (e.g., topic, genre, culture) to understand the speaker's intended meaning (Vandergrift, 2004). Bottom-up processing is the process of deciphering the sounds in speech and matching them to lexical items in the target language (Field, 2008).

The present study focuses on aural decoding, which is an aspect of bottom-up processing (Leonard, 2019). Broadly speaking, aural decoding is the process of transforming aural input and matching it to the corresponding lexical items in the target language (Field, 2008). Incorporating decoding training approaches is a long-term key to EFL listening comprehension because it allows learners to free up their attention and focus more on the speaker's intended meaning (Field, 2010). Although our current study focuses on aural decoding training, we fully recognize the importance of integrating aural decoding within the top-down approach for EFL learners. Developing aural decoding skills helps EFL student to identify and understand individual sounds, words, and phrases in spoken language, which forms the basis for overall listening comprehension. In other words, aural decoding can help learners to understand enough linguistic elements of what they hear in order to use their top-down skills to listen for main ideas or make the correct inference in an audio input.

The most frequently used strategy for aural decoding training is dictation practice (Jia & Hew, 2021b), which typically requires students to listen to an audio input, transcribe it, and then assess the accuracy of their transcription based on the subsequent answer provided to students. Unlike traditional approaches such as listening to audiobooks, where learners struggled with the pace and complexity of a long-spoken material (Kartal & Simsek, 2017), aural decoding training focuses on helping learners develop the ability to recognize words from spoken language. In recent years, a growing number of studies have demonstrated the positive effect of aural decoding training on learners' listening comprehension (e.g., Jia & Hew, 2021a; Ke & Wang, 2022; Leonard, 2019). For example, Ke and Wang (2022) and Leonard (2019) respectively reported strong positive correlations between aural decoding and L2 listening comprehension ($r = 0.69$, $p < 0.01$), and ($r = 0.837$, $p < 0.001$). A recent meta-analysis reported an overall significant effect in favor of decoding training over non-decoding training in terms of student listening outcomes ($g = 0.553$, $CI = 0.348 – 0.759$, $p < 0.001$) (Jia & Hew, 2021a). Despite its benefits, instructors face challenges in conventional aural decoding training, such as insufficient in-class time for decoding practice, student disengagement with aural decoding practice, and ineffective feedback to support students' competence in decoding skills (Jia & Hew, 2021b).

One way to address these issues is to incorporate chatbots into students' language learning (Hew et al., 2023). Recent reviews (Huang et al., 2022; Zhang et al., 2023) point to the pedagogical value of chatbots in language learning, such as allowing students to practice the target language without time and place constraints, providing instant feedback on students' learning performance, and encouraging students during repeated language practices. As mastering decoding skills and automating the decoding process require substantial engagement with decoding practices (Field, 2008; Jia et al., 2023), the challenge lies in maintaining high levels of student engagement in fully online courses, which, however, are generally associated with lower levels of student engagement than on-campus or blended courses (Bai et al., 2022; Cavinato et al., 2021). Therefore, the effectiveness of chatbots in engaging fully online students remains an unanswered question. In this context, our study contributes to the existing literature by exploring the use of a chatbot to support EFL aural decoding skills and enhance student engagement with decoding practices in a fully online setting.

Literature Review

The use of chatbot in language learning

Bibauw et al. (2022b) used the term dialogue-based computer-assisted language learning (CALL) to refer to any system that can interact dialogically with students via an automated agent for language learning purposes, including intelligent tutoring systems, conversational agents, dialogue systems, and chatbots. Among the various forms of agents, chatbots are typically text-based (Bibauw et al., 2022a).

In this study, the dialogue system is referred to as a chatbot because it fulfils the following essential attributes that comprise a chatbot as outlined by Garcia Brustenga et al. (2018). First, the software uses natural language processing to facilitate interaction with humans. This allows it to interpret and understand human input and then generate appropriate responses. Second, the software's user interface is a conversational interface that is typically integrated into messaging applications on various devices such as smartphones. In educational contexts, chatbots can serve as virtual tutors (Garcia Brustenga et al., 2018) or interactive learning support agents consistent with their pedagogical goals, which can also be considered as the chatbot's personality (Car et al., 2020). Therefore, we used the term chatbot to describe the dialogue system that interacts with students via text during daily decoding training.

Chatbots play an increasingly important role in language learning because they can interact with humans via natural language in textual and auditory ways. In particular, chatbots can interact with students, which helps students practice foreign languages as conversation leads to meaningful use of the target language (Bibauw et al., 2019).

The use of chatbots in language learning can address the lack of practice opportunities, which is the biggest challenge for English language learners in non-English speaking countries (Lin & Mubarok, 2021). Previous chatbots used in language learning primarily focus on speaking skills (see a review by Huang et al., 2022). For example, Hsu et al. (2021) designed a TOEIC Practice Chatbot (TPBOT), which can recognize students' pronunciation in real time and provide appropriate responses, to support students' spoken English. After a four-month intervention, students who participated in the TPBOT activity performed significantly better in oral English skills, compared with those who used textbooks with traditional media (e.g., mp3 and audio CDs).

In addition to speaking, some studies focused on using chatbot to support students' reading skills (Ruan et al., 2019; Xu et al., 2021). For many EFL learners, a lack of vocabulary or grammar can be an obstacle to successful reading comprehension. Several researchers have therefore examined how the use of chatbot can facilitate vocabulary development (e.g., Jia et al., 2012) and grammar (e.g., Kim et al., 2019).

To our best knowledge, very few studies have examined the use of chatbots in supporting listening skills (e.g., Dizon, 2020; Kim, 2018). These previous studies that investigated chatbots in listening instruction were typically conducted in a blended mode where students interacted with the chatbots online and then attended weekly face-to-face classes. Additionally, up till now, no study has examined the use of chatbots specifically on listening aural decoding. The present article contributes to the literature by examining the use of a chatbot in supporting EFL aural decoding skills in a fully online environment.

The flipped SEF-ARCS decoding model

In this study, we used the flipped SEF-ARCS decoding model (see Appendix A) to inform our instructional design for the chatbot-supported decoding training approach. We chose this model because it can improve EFL students' decoding skills (Jia et al., 2023), and has a comprehensive theoretical foundation based on Jia and Hew (2021a) and Keller (1987). The SEF model includes the elements of Self-exploration, Feedback, Generalization & Automation (Jia & Hew, 2021a). In Self-exploration, various types of scaffolding were made available to the students such as breaking the video clip into short segments with few unknown words and allowing repeated listening and slowing down of the clip. In Feedback, students were given not only the correct answers but also detailed explanations on how the words actually sounded in the video clips (i.e., process-level feedback). In Generalization & automation, students were given opportunity to regularly practice decoding skills in different contexts. The ARCS motivation model includes the elements of Attention, Relevance, Confidence, and Satisfaction. To capture and sustain students' attention, one-sentence video clips that were selected from authentic sources such as movies and news report were used as the learning materials. To enhance the relevance of student learning, the teacher explained difficult sounds by establishing connections between new information (blended sounds in connected speech) and what students already knew (isolated sounds in careful speech). To enhance student confidence, students were provided different tasks of easy or hard mode to practice according to their proficiency. To promote student satisfaction, incentives such as additional course grades were given to students who completed the decoding tasks.

In sum, the flipped SEF-ARCS decoding model emphasizes the importance of providing students with authentic materials for decoding training, including movies, cartoons, and real-life songs. Following this model, the content of our chatbot decoding activities was selected from authentic English materials that could provide students with real-life context and language use. Moreover, the interactive nature of chatbots promotes active learning, which has been shown to enhance language acquisition (Warschauer & Liaw, 2011).

Purpose of this research

In the following sections, we will first describe the flipped SEF-ARCS decoding model, along with the results of our preliminary work (Jia et al., 2023). Subsequently, we will discuss the limitations identified in the preliminary study and address these by introducing the main study. Specifically, we developed a chatbot for a fully online decoding course to overcome these limitations. The effectiveness of the chatbot was evaluated by comparing its support of student decoding performance and engagement with the flipped learning approach. It also investigated how students perceived the decoding activities supported by the chatbot. Consequently, the main study was guided by the following research questions:

Research question 1: How does the chatbot-supported online learning approach affect student decoding performance compared to the flipped learning approach?

Research question 2: How does the chatbot-supported online learning approach affect student engagement compared to the flipped learning approach?

Research question 3: How do students in the chatbot-supported online learning approach perceive the decoding activities?

Preliminary Study – Comparing the Effects of Using the Flipped SEF-ARCS Decoding Model and the Conventional Decoding Approach

The preliminary study was a quasi-experiment research that investigated the effectiveness of the flipped SEF-ARCS decoding model in improving EFL learners' decoding skills and listening proficiency (Jia et al., 2023). The study involved 80 first-year university students in China who had an elementary level of English listening proficiency. The students were randomly assigned to an experimental group (N = 44) and a control group (N = 36). The experimental group received the flipped SEF-ARCS decoding model, while the control group received the conventional approach to teaching decoding during a 15-week intervention.

More specifically, in the experimental group, students were asked to do one required task each day in the Moodle, an online learning management system (LMS) (see Appendix B). The required task consisted of a decoding video clip either in easy mode (i.e., answer multiple-choice questions), or a hard mode (i.e., filling in missing words). The content of the video clip was same for the two modes and students could choose to complete either the easy or hard mode based on their language proficiency. After students submitted their answers for the decoding task, the Moodle system automatically provided a pre-recorded teacher explanation video tailored to address students' common mistakes in the decoding task. In these videos, the teacher addressed the phonetic changes of the target words, ensuring that all possible decoding errors that the students encountered during the decoding were addressed. In this way, the system was able to provide students with feedback about the common errors, and promote a deeper understanding of the phonetic changes that occur in each sentence. Afterwards during a face-to-face class lesson, the teacher would review the common difficulties encountered by the students in the pre-class practice and guided students to complete more decoding tasks.

The results of the experiment showed that the decoding test scores of EG ($M_{adjusted}$ = 11.07, SE= 0.59) were significantly higher than those of CG ($M_{adjusted}$ = 8.82, SE = 0.60), F (1, 69) = 7.112, p = .010, partial η^2 = .093, indicating a medium effect size (Cohen, 1988). This indicated that the flipped SEF-ARCS decoding model was more effective in supporting students' decoding performance than the conventional decoding approach. Students also commented positively on the flipped SEF-ARCS decoding model. They appreciated the SEF-ARCS model for decoding learning because it incorporates authentic connected speeches. This allows students to understand the process of decoding words in real-life situations and apply these skills to various listening tasks. In addition, this model guided the teacher to provide feedback on the difficult sounds which involved phonetic modifications. The explanation video, serving as process feedback, helped learners understand how to proceed to the correct answers and thus improve their decoding skills.

Despite the positive results reported in the preliminary study (Jia et al., 2023), two main challenges were identified when using the flipped SEF-ARCS decoding model. First, the biggest challenge was the heavy teacher workload. During the long-term intervention, the teacher had to remind students to complete the tasks and check their answers each day. The heavy workload stressed her out. The other challenge was the high requirement for the individual teacher's decoding skills. The model required the teacher to decipher and explain how the words were sounded in the connected speech, which may be beyond the ability of many non-native EFL teachers (Tsang, 2017).

Main Study – Design and Evaluation of the Chatbot-supported Online Decoding Approach

As mentioned earlier, to address the problem of over-dependence on individual teachers in implementing the flipped SEF-ARCS decoding approach, we designed and developed a chatbot to adapt the flipped learning approach to chatbot-supported online learning approach in the main study. In addition, to overcome students' disengagement and isolation in online learning, we developed the chatbot based on the self-determination theory and social presence theory to serve as a 24/7 conversational agent.

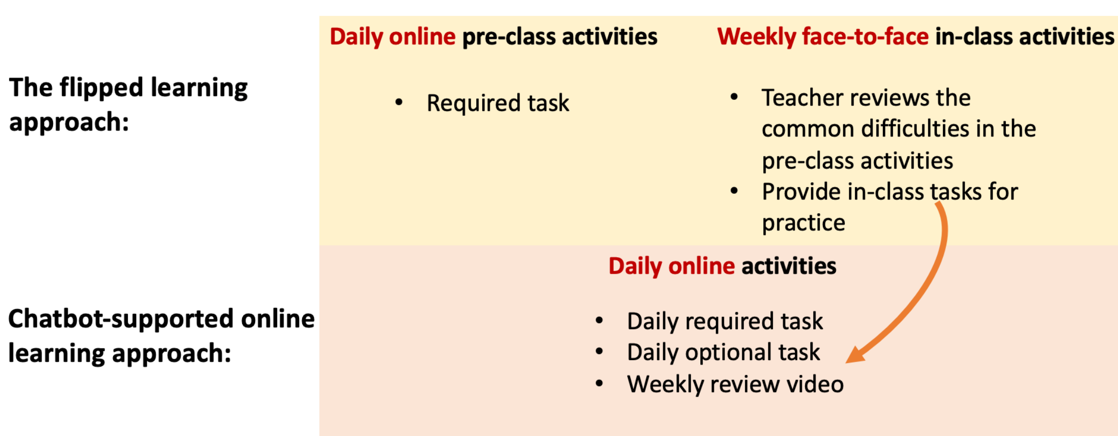
Adapting the flipped learning approach to the chatbot-supported online approach

Activity arrangement

The flipped SEF-ARCS learning approach included daily out-of-class online decoding activities (one required task with two levels of difficulty – hard or easy for students to choose) and weekly face-to-face decoding activities (weekly learning review and in-class tasks). To convert the flipped learning approach to the chatbot-supported online learning approach, the daily required online pre-class activities were used directly in chatbot-supported online learning, whereas the weekly face-to-face activities were assigned as weekly review videos and daily optional tasks in the online learning. See Figure 1 for the description.

Figure 1

Converting the activities from flipped learning to chatbot-supported online learning

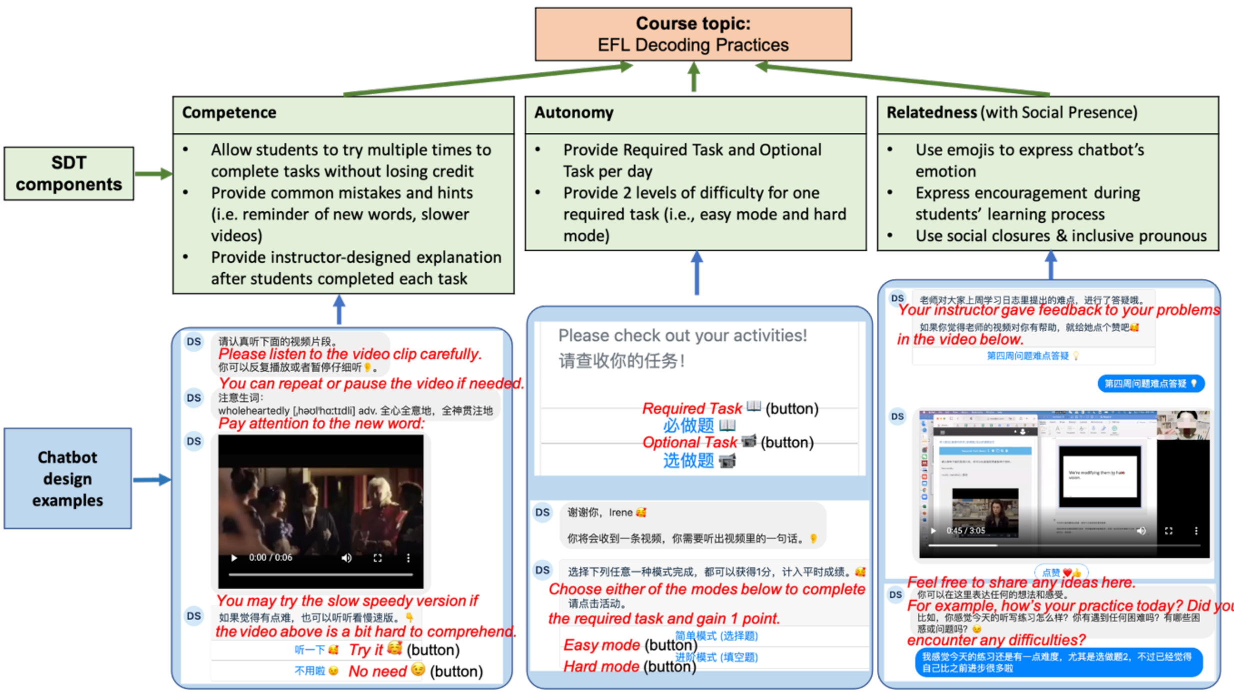


Design and development of the chatbot

The self-determination theory of motivation formed the basis for the chatbot design (see Figure 2) because the theory is often used to support learners' engagement and motivation (e.g., Huang & Hew, 2018). Self-determination theory postulates that an individual is intrinsically motivated when satisfied by three innate psychological needs: the need for competence, the need for autonomy, and the need for relatedness (Deci & Ryan, 2000).

Figure 2

Chatbot design informed by the SDT and Social Presence indicators



To facilitate each student's competence, which is the perceived capability of accomplishing a given task, we incorporated the common decoding mistakes that most students encountered into the chatbot's feedback. The common mistakes were pre-determined by the instructor based on previous students' learning records. When the students' answers activated the pre-set common mistakes, targeted feedback would be given. For example, the chatbot would reply, "There may be two words blended with each other, rather than one word." To further promote each student's competence, students were allowed to attempt the decoding exercises multiple times. For example, the chatbot would remind students that they could listen to the video clip for each decoding task multiple times if necessary. We also provided hints throughout the learning process. For instance, the chatbot highlighted new words that students would encounter in the decoding task before they listened to the video clip. In addition, if students found the video clip difficult to understand, the chatbot provided an option for students to choose a slower version to help students complete their daily tasks.

Autonomy refers to learners' sense of free choice when participating in an activity (Deci & Ryan, 2000). The chatbot offered two difficulty levels for each required task: (a) easy mode (i.e., multiple-choice questions), and (b) hard mode (i.e., filling in missing words). Students could choose to complete either the easy or hard required decoding task and earn a point. In addition to the required tasks, the chatbot provided one optional task each day for students who wanted to practice more. The optional tasks were in the form of fill-in-the-blank questions. The optional tasks did not count toward students' course credit. The free choice offered by chatbot is expected to motivate students to spend more time on daily decoding practice.

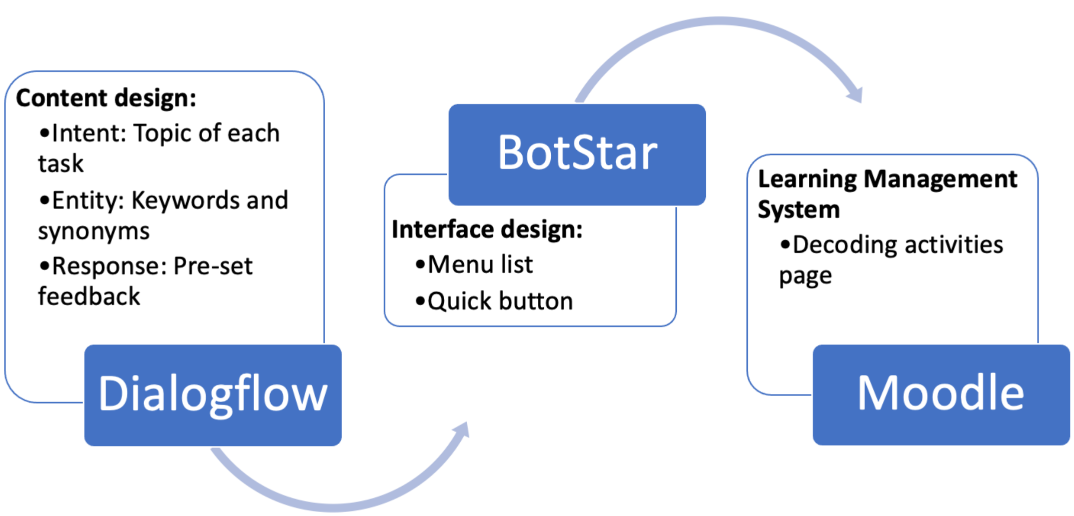
Relatedness can be defined as students' desire to feel associated with others and to own a sense of belonging (Deci & Ryan, 2000). In teaching and learning, students have an imperative demand for interactions with their peers and instructors. To meet students' need for relatedness, that is interaction with others, we incorporated the indicators of social presence (Hew et al., 2023) into the chatbot conversational design. Social presence refers to the extent to which participants consider themselves as real persons socially and emotionally in an online learning environment (Garrison et al., 1999). In this study, the chatbot as a learning partner would greet students, mention students' names, and use social closures and inclusive pronouns. For example, "Good day, [student's name]! Let us start today's exercises." We used emojis to express the chatbot's emotional reactions to students' performance, by which students may picture the learning partner with vivid and lively characteristics. The chatbot would also encourage students when they made mistakes. For example, the chatbot would respond to student's mistakes by stating, "Don't worry. You're not alone. Many of your peers did not recognize this word, either." In this way, students were more likely to develop a sense of social connection with other students who had the same decoding problems in the online learning environment.

The chatbot's dialogue design was created using Google Dialogflow, a visual chatbot development platform. Designers can build chatbot's conversation in this platform using three components, namely, intent, entity, and response. The topic of each conversation can be created in one intent. To help the chatbot understand students' inputs, we included keywords and synonyms in entities. Pre-set feedback were added in responses (see Appendix C for the design of how the chatbot detects a student's answers).

After designing all conversational contents in Dialogflow, we structured the chatbot's interface using the chatbot platform BotStar (https://app.botstar.com/bots), which allowed us to create user-friendly features such as menu lists and quick buttons. Once the interface was developed, an embedded code was generated which could be inserted into Moodle. Students can access the chatbot in Moodle via laptop, tablet, or mobile phone, depending on their preference. Figure 3 shows the chatbot design process.

Figure 3

Chatbot design process



Method

Participants

This study was conducted from September to December 2021 at a public university in western China. Participants were first-year undergraduate students in two pre-assigned classes who enrolled in a compulsory English course. Since the two groups were pre-assigned before the semester started, we employed a quasi-experimental design to randomly assigned the two groups to either a chatbot-supported online learning class (N = 46) or a flipped learning class (N = 43).

The students' listening proficiency was also evaluated using the listening section of Cambridge English: Preliminary, which corresponds to the B1 level of the Common European Framework of Reference (CEFR), indicating that learners can communicate in English in practical, everyday situations (Cambridge Assessment English, 2021). Ethical approval for data collection was obtained from the authors' university Institutional Review Board. Informed consent was obtained from the students to participate in the study.

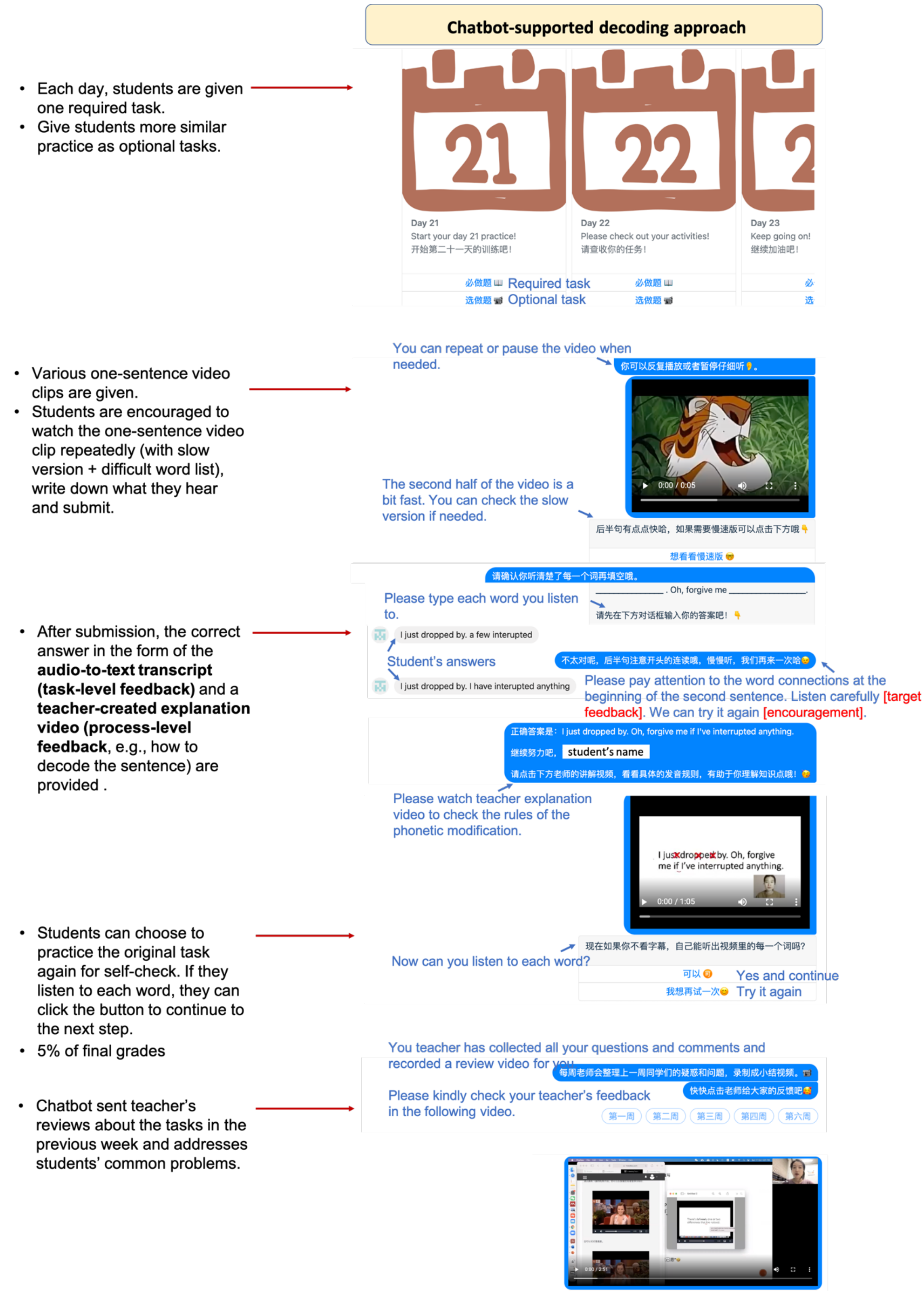
Treatments

In the flipped learning group, students were asked to do one required task each day in the Moodle LMS. We used the same SEF-ARCS decoding model as the treatment group in our preliminary study to design the flipped learning group (see preliminary study section for a detailed description).

The chatbot-supported online learning group used the same curriculum and the same online LMS (the chatbot was embedded in Moodle). Contrary to the flipped learning group, all the learning activities were completed online with the help of the chatbot. The face-to-face learning activities in the flipped learning group were moved to online learning in the chatbot-supported online learning (as shown in Figure 1). Participants in the two groups were given the same decoding tasks. Each task consisted of the following components: (1) a one-sentence video clip that students could play repeatedly to reconstruct as many missing words as possible (see Appendix D for more detailed description of the video content), (2) chatbot's targeted hints, which reminded students of their mistakes and provided clues to the correct answers, (3) explanatory videos recorded by the teacher on the phonetic changes of the missing words in the video clip. Similar to the flipped learning group, the chatbot could provide the teacher explanation video immediately after students submitted their answers to the decoding task. Students could spend 5 to 10 minutes on each decoding task, depending on their own needs (i.e., whether they needed the slow version of the video clip or played the video clip and the teacher explanation video repeatedly). Throughout the experiment, the decoding tasks were updated daily. Students in both groups were able to access the decoding tasks on their convenience once the tasks were posted on Moodle. Therefore, the amount of time students spent on the decoding tasks was roughly equivalent for the two groups. Serving as a learning partner, the chatbot was expected to encourage and guide students' learning. Figure 4 shows an example of the interface of the chatbot.

Figure 4

The Interface of the Chatbot-Supported Decoding Tasks

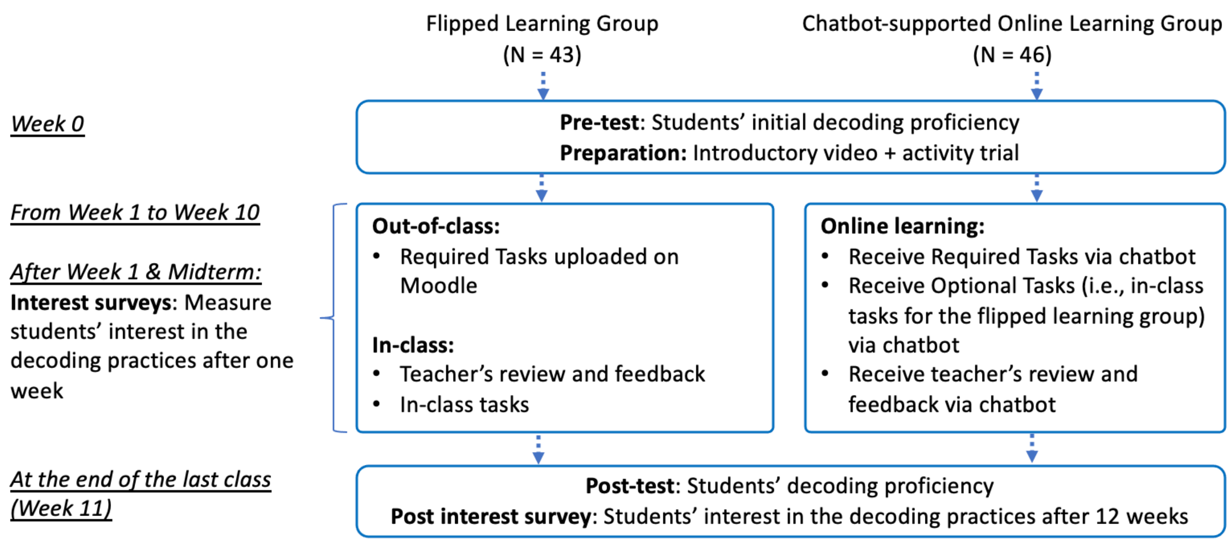

Chatbot-supported decoding approach
• Each day, students are given one required task.
• Give students more similar practice as optional tasks.
21
22
Day 21
Start your day 21 practice!
开始第二十一天的训练吧！
Day 22
Please check out your activities!
请查收你的任务！
Day 23
Keep going on!
继续加油吧！
必做题 Required task
选做题 Optional task
• Various one-sentence video clips are given.
• Students are encouraged to watch the one-sentence video clip repeatedly (with slow version + difficult word list), write down what they hear and submit.
You can repeat or pause the video when needed.
你可以反复播放或者暂停仔细听。
0:00 / 0:05
The second half of the video is a bit fast. You can check the slow version if needed.
后半句有点点快哈，如果需要慢速版可以点击下方哦
想看看慢速版
请确认你听清楚了每一个词再填空哦。
. Oh, forgive me
Please type each word you listen to.
请先在下方对话框输入你的答案吧！
• After submission, the correct answer in the form of the **audio-to-text transcript (task-level feedback)** and a **teacher-created explanation video (process-level feedback**, e.g., how to decode the sentence) are provided .
I just dropped by. a few interupted
Student's answers
不太对呢，后半句注意开头的连读哦，慢慢听，我们再来一次哈
I just dropped by. I have interupted anything
Please pay attention to the word connections at the beginning of the second sentence. Listen carefully [target feedback]. We can try it again [encouragement].
正确答案是：I just dropped by. Oh, forgive me if I've interrupted anything.
继续努力吧， student's name
请点击下方老师的讲解视频，看看具体的发音规则，有助于你理解知识点哦！
Please watch teacher explanation video to check the rules of the phonetic modification.
I just dropped by. Oh, forgive me if I've interrupted anything.
0:00 / 1:05
• Students can choose to practice the original task again for self-check. If they listen to each word, they can click the button to continue to the next step.
• 5% of final grades
现在如果你不看字幕，自己能听出视频里的每一个词吗？
Now can you listen to each word?
可以 Yes and continue
我想再试一次 Try it again
• Chatbot sent teacher's reviews about the tasks in the previous week and addresses students' common problems.
You teacher has collected all your questions and comments and recorded a review video for you.
每周老师会整理上一周同学们的疑惑和问题，录制成小结视频。
Please kindly check your teacher's feedback in the following video.
快快点击老师给大家的反馈吧
第一周 第二周 第三周 第四周 第六周
0:00 / 2:51

Data collection and analysis

The quasi-experimental procedure is shown in Figure 5. The experimental design lasted 12 weeks. In the first lesson (Week 0), students in both groups were invited to take a pre-test to measure their initial decoding skills. They were instructed to watch an introductory video about the decoding activities and the learning platform used in their group. We also provided a trial decoding activity to students in both groups to familiarise them with the technology used (i.e., Moodle for the flipped learning group and chatbot for the fully online learning group). A technical assistant was assigned to solve students' technical problems throughout the experiment.

Figure 5

Experimental Procedure of the Study



To answer the first research question, we compared students' post-test as their decoding performance between the chatbot-supported online learning group and the flipped learning group. Students were invited to complete the post-test to assess their decoding performance in the last class. Both the pre- and post-tests were administered in the form of a partial dictation test (see Appendix E for an introduction of the tests), in which students were given an incomplete transcript and were required to fill in the missing words (Jia & Hew, 2021a). Of the 89 participants, 3 students were absent for either the pre-test or post-test. Therefore, the sample size for quantitative analysis of decoding performance was 86, with 45 students in the chatbot learning group and 41 students in the flipped learning group. We used the independent samples t-test to evaluate the differences of decoding skills for two groups in both pre- and post-tests.

To answer the second research question, we measured students' engagement with decoding practices from the following two aspects, behavioral engagement and emotional engagement. First, the core concept of behavioral engagement refers to participation (Fredricks et al., 2004), such as completing the required learning activities (Lo & Hew, 2021). We collected students' online learning records in the Moodle LMS and chatbot systems, and compared their required task completion each week. Since students were provided with five required tasks each week, the weekly completion scores for students' behavioral engagement were 5 in total. We performed a statistical analysis using Fisher's exact test to examine the difference between the two groups in required task completion. We chose to use Fisher's exact tests rather than Chi-square tests of independence because the expected counts of multiple cells is less than 5 and Fisher's exact test is more appropriate for smaller values (Bolboacă et al., 2011).

Second, emotional engagement refers to students' affective reaction, and interest is considered an important indicator of emotional engagement (Fredricks et al., 2004). We hence employed the interest/enjoyment subscale of the Intrinsic Motivation Inventory (IMI) questionnaire to measure students' level of interest in the decoding practices in two different learning approaches. The IMI questionnaire is used to assess participants' interest and enjoyment in an activity (Monteiro et al., 2015). We measured students' interest levels at three time points (i.e., after Week 1, at midterm, and at the end of the semester) to determine whether using the chatbot had a significant effect on students' emotional engagement over the time, compared to the flipped learning group. The questionnaire consists of seven items with a seven-point scale (1 for "strongly disagree" and 7 for "strongly agree"). An example item is "This activity was fun to do." Cronbach's alpha results for the "interest" subscale were high at three time points for both groups (flipped learning group: 0.952 at Time 1, 0.927 at Time 2, and 0.913 at Time 3; chatbot-supported learning group: 0.884 at Time 1, 0.929 at Time 2, and 0.930 at Time 3). We used the one-way ANOVA to compare the differences in students' interest levels between the two groups at each time point. The results of all statistical assumptions checks are provided in Appendix F.

To address the third research question, an open-ended survey was administrated to students in the chatbot-supported learning group at the end of semester to identify: (a) which parts of the decoding activities engaged them during the learning process, and (b) which parts of the activities discouraged them from the decoding practices. Students' responses were analyzed using thematic analysis (Braun & Clarke, 2006). Specifically, we read each sentence of students' responses to identify the initial themes inductively, rather than applying a pre-determined framework to code the qualitative data. Two independent coders undertook the thematic analysis procedure and then compared the identified themes with each other, which yielded 96% inter-rater agreement. Disagreement was solved by careful discussion among all authors. The results of the third research question were used to explain the quantitative analysis for the first two research questions.

Results

RQ1: How does the chatbot-supported online learning approach affect student decoding performance compared to the flipped learning approach?

First, we conducted an independent-samples *t*-test to determine the initial difference in students' decoding skills between the chatbot-supported online learning group and the flipped learning group. The result showed no significant difference in students' initial decoding skills between the two groups, $t\ (84) = 0.429$, $p = 0.669$. Then, we used an independent-samples *t*-test to compare students' final decoding performance between the two groups. Figure 6 shows the boxplots of the post-test decoding scores. The results showed that the students' post-decoding performance in the chatbot-supported online learning group ($M = 17.33$, $SD = 7.35$) had no significant difference from those in the flipped learning group ($M = 17.59$, $SD = 6.51$), $t\ (84) = 0.168$, $p = 0.298$ (Table 1). This suggests that students who practiced decoding skills online with the chatbot achieved similar decoding performance as participants in the flipped learning group.

Figure 6

Boxplots of Students' Post-Test Scores in Two Groups

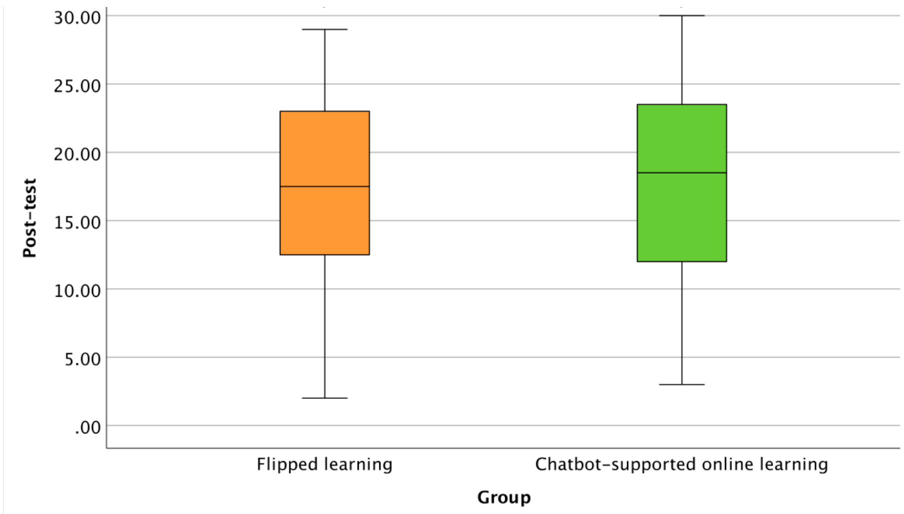


Table 1

Independent-Samples t-test on Students' Post-Test Decoding Performance Scores

Groups	*N*	*Mean (SD)*	*t*-value	*p*-value
Chatbot-supported online learning	45	17.33 (7.35)	.168	.298
Flipped learning	41	17.59 (6.51)		

RQ2: How does the chatbot-supported online learning approach affect student engagement compared to the flipped learning approach?

As mentioned in data collection and analysis, we examined student engagement from both behavioral and emotional aspects. Students' behavioral engagement was measured by calculating students' weekly completion scores of the required tasks. We assigned one point when students completed one day's required task. Since there were five required tasks each week, the full task completion score for each week was 5. The descriptive data presented in Figure 7 showed that the completion scores of the chatbot-supported learning group were higher than those of the flipped learning group all the time. Although both groups experienced a decrease in completion scores over time, the decrease was less severe in the chatbot-supported online learning group compared to the flipped learning group.

The results of the Fisher's exact tests (Table 2) showed significant differences in completion scores between the two groups from week 3 to week 10. To address the concern of Type I errors, we applied a Bonferroni correction by dividing the usual alpha level of 0.05 by the number of tests performed, which resulted in an adjusted alpha level of 0.005. When comparing the *p*-values from the Fisher's exact tests to this adjusted alpha level, we found that student completion scores between the two groups at Week 8 no longer reached statistical significance. Nonetheless, it is important to note that there is a clear trend in the data indicating that the chatbot group performed better in completing the required tasks over time. For example, at week 10, students in the chatbot-supported learning group (87.0%) were more likely to complete all five required tasks than students in the flipped learning group (53.5%).

Figure 7

Descriptive Data of the Weekly Required Task Completion Scores

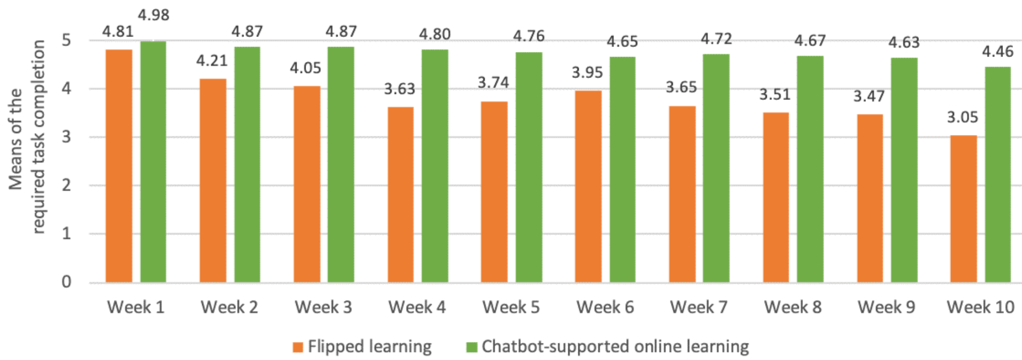


Table 2

Comparison of the Required Tasks Completion

		Completion scores						
Time	**Groups**	**5**	**4**	**3**	**2**	**1**	**0**	***p*-value**
Week 1	Flipped	97.8%	2.2%	0.0%	0.0%	0.0%	0.0%	.061
	Chatbot	86.0%	11.6%	0.0%	2.3%	0.0%	0.0%	
Week 2	Flipped	74.4%	9.3%	0.0%	4.7%	2.3%	9.3%	.053
	Chatbot	87.0%	13.0%	0.0%	0.0%	0.0%	0.0%	
Week 3	Flipped	79.1%	0.0%	0.0%	0.0%	4.7%	16.3%	<.001*
	Chatbot	87.0%	13.0%	0.0%	0.0%	0.0%	0.0%	
Week 4	Flipped	53.5%	11.6%	14.0%	2.3%	2.3%	16.3%	.001*
	Chatbot	89.1%	8.7%	0.0%	0.0%	0.0%	2.2%	
Week 5	Flipped	51.2%	20.9%	11.6%	0.0%	0.0%	16.3%	<.001*
	Chatbot	89.1%	6.5%	0.0%	0.0%	4.3%	0.0%	
Week 6	Flipped	67.4%	9.3%	0.0%	0.0%	20.9%	2.3%	.004*
	Chatbot	84.8%	8.7%	2.2%	0.0%	0.0%	4.3%	
Week 7	Flipped	67.4%	0.0%	7.0%	2.3%	2.3%	20.9%	.002*
	Chatbot	89.1%	6.5%	0.0%	0.0%	0.0%	4.3%	
Week 8	Flipped	65.1%	2.3%	2.3%	0.0%	9.3%	20.9%	.006
	Chatbot	87.0%	6.5%	2.2%	0.0%	0.0%	4.3%	
Week 9	Flipped	62.8%	7.0%	0.0%	2.3%	0.0%	27.9%	.002*
	Chatbot	91.3%	0.0%	2.2%	0.0%	0.0%	6.5%	
Week 10	Flipped	53.5%	7.0%	0.0%	2.3%	4.7%	32.6%	.003*
	Chatbot	87.0%	2.2%	0.0%	0.0%	2.2%	8.7%	

Note. Flipped represents the flipped learning group (n = 43); Chatbot represents the chatbot-supported online learning group (n = 46).

*significant using a Bonferroni correction of $p < 0.005$.

Second, we compared students' interest levels to identify the difference in the emotional engagement between the two groups. The three interest questionnaires were completed by 40 students in the flipped learning group and 42 students in the chatbot-supported learning group. Descriptive data (Figure 8) showed students in the chatbot-supported learning group showed high levels of interest in the decoding practices from week 1 (M_{T1} = 5.73, SD_{T1} = .91) through midterm (M_{T2} = 5.90, SD_{T2} = 1.12) to the end of the intervention (M_{T3} = 6.06, SD_{T3} = 1.19). Although students in the flipped learning group showed increased interest in the decoding practices (M_{T1} = 5.09, SD_{T1} = 1.21; M_{T2} = 5.28, SD_{T2} = 1.16; M_{T3} = 5.63, SD_{T3} = 1.21), their interest was lower than that of the chatbot-supported learning group at all time points. The results of the one-way ANOVA test (Table 3) indicated that differences in students' interest levels between the two groups were statistically significant at Time 1 ($F(1, 80) = 7.421$, $p = 0.008$), and Time 2 ($F(1, 80) = 6.225$, $p = 0.015$), but not at Time 3 ($F(1, 80) = 2.679$, $p = 0.106$).

Figure 8

Students' Interest Level at Three Time Points in Two Groups

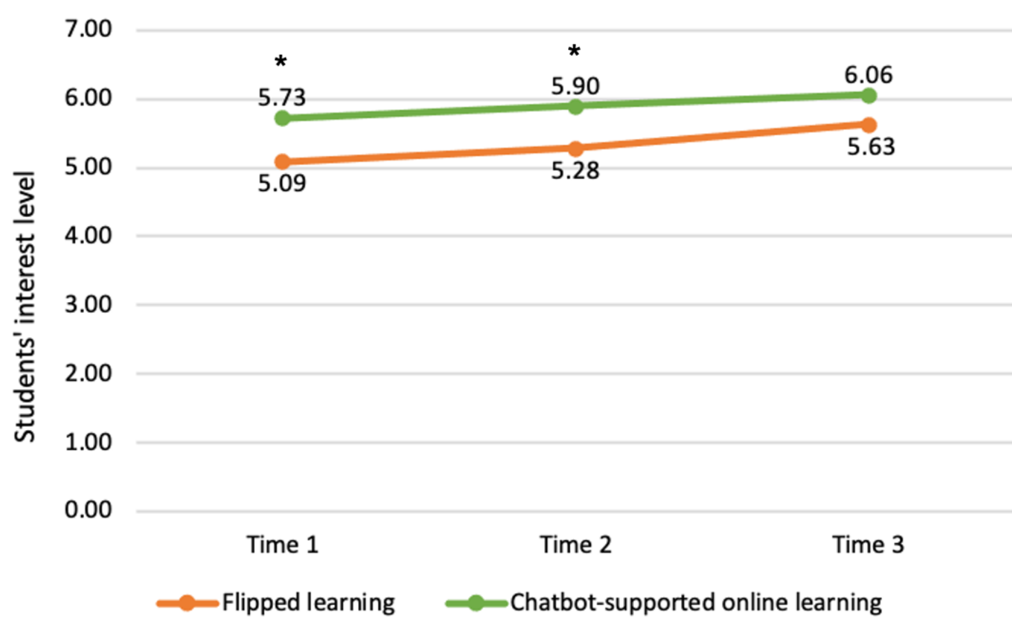


Table 3

Comparison of students' interest between the two group at three time points

Time points	Groups	*n*	*Mean (SD)*	*p*-value
Time 1	Flipped	40	5.09(1.21)	.008*
	Chatbot	41	5.73(.91)	
Time 2	Flipped	40	5.28(1.16)	.015*
	Chatbot	41	5.90(1.12)	
Time 3	Flipped	40	5.63(1.21)	.106
	Chatbot	41	6.06(1.21)	

Note. *significant using a Bonferroni correction of $p < 0.017$.

Research question 3: How do students in the chatbot-supported online learning approach perceive the decoding activities?

We received 41 responses (89% response rate) from students in the chatbot-supported learning group. Thematic analysis revealed 5 factors that students valued during the decoding practices: 1) personalized supports from the chatbot, 2) various real-life decoding materials, 3) teacher's explanatory videos, 4) interactive learning process using chatbot, and 5) encouraging learning climate. Each factor is explained in detail below. See Appendix G for details regarding the coded data.

First, students highly appreciated the personalized supports offered by the chatbot to help them complete the decoding practices (*n* = 25). One of the highly mentioned personalized support was the target feedback. If the students' answer did not contain the correct phonetic modifications, the target feedback would be activated and sent to students via the chatbot, saying "Please pay attention to the word connections at the beginning of the second sentence" (see Figure 4). Another personalized support was the different levels of difficulty. The chatbot provided students with the required tasks in easy or hard modes. If they felt it challenging to understand the video clip in normal speed, they could click on the "need slow version" button. In this way, students were provided with a "personalized learning environment" (Student 35).

Second, students frequently reported that the various real-life decoding materials increased their engagement (*n* = 21). The learning materials were selected by the teacher from current movies, TV programs, and music sounds and structured in the form of short video clips. Using these "interesting and eye-catching" (Student 36) materials helped "increase students' curiosity to the target culture" and they "enjoyed these authentic materials more than traditional listening exercises from textbooks" (Student 29).

Third, students found it useful to watch explanatory videos after completing each task to understand the related decoding knowledge (*n* = 21). In each video, the teacher explained the rules of a particular phonetic modification within one minute to engage students in bite-sized learning. Although all of the explanatory videos were recorded by the teacher and pre-set in the chatbot's feedback, students indicated that their learning problems were addressed clearly and promptly through the explanatory videos. In addition, when students received teachers' explanations from the chatbot, they could gain "a strong sense of teacher-student bonding" (Student 24).

Fourth, completing decoding practices in an interactive way was another attractive factor for students (*n* = 5). During the learning process, the chatbot provided step-by-step instructions and learning content, which helped students focus on their tasks. Students reported, "The way we answered the decoding tasks with the chatbot was fascinating" (Student 11). Unlike doing the exercises passively, students received the decoding tasks during their conversation with the chatbot, which helped reduce their negative attitudes (e.g., anxiety and boredom) when practicing listening online by themselves (Student 32). In addition, they perceived interacting with the chatbot as "talking to a real person, accompanying them throughout the semester" (Student 37).

Finally, the chatbot created an encouraging and pressure-free learning environment (*n* = 4). For example, when students entered incorrect words, the chatbot prompted them, "Almost there! I believe you can do it! Let's try again." In this situation, students reported that "the chatbot was always patient with me and never blamed me when I made mistakes" (Student 36). Once students got each word correctly, the chatbot responded to them with positive feedback, such as "You did it", which gave students "a sense of accomplishment" (Student 34). These "encouraging, warm, and friendly" (Student 41) expressions helped engage students to participate in the decoding activities.

In summary, the students appreciated the personalized learning experience, teachers' detailed explanations, various authentic listening materials that catered to their different needs and the interaction with the chatbot.

Discussion and Implication

In this study, we tested the use of a chatbot, developed based on the self-determination theory and social presence theory, to serve as a 24/7 conversational agent to support student aural decoding practice in a fully online course. We developed the chatbot-student interaction using self-determination theory because this theory helps facilitate student sense of competence, autonomy, and relatedness. Social presence can help alleviate the sense of online isolation especially in a fully online learning environment. In this section, we discuss the major findings based on our research questions in employing chatbots for listening decoding purposes.

The results revealed that the chatbot-supported fully online group and the flipped group performed equally well in the decoding test. One possible explanation for the similar performance of the two groups is the use of the SEF-ARCS decoding model in designing the learning content. Both chatbot-supported learning and flipped learning approaches in this study supported students' decoding learning by directing their limited cognitive load through one-sentence materials to the decoding process. The decoding model allowed students to focus on specific connected speech features and improve their decoding skills more efficiently. In addition, the model emphasizes automation and encourages students to practice until their decoding skills becomes automatic. During the 12-week intervention, both groups received sufficient opportunities for practice, various decoding tasks and materials in different contexts, and explanatory videos from the teacher addressing common errors. This contributed to their ability to automatically decode real-life English which involves blended sounds. Given the equally satisfactory performance in both groups, chatbot-supported online learning could be an alternative for institutions lacking sufficient staff to teach decoding skills. We recommend teachers to design their own chatbots with targeted learning materials to help individual students practice decoding in an interactive way. It is also important to note that while this study focused on decoding training within both the flipped and chatbot-supported learning approaches, we acknowledge the potential benefits of integrating our decoding training into the broader context of listening instruction. For instance, in a typical overall listening instruction that consists of pre-listening, intensive listening and post-listening for details, we may incorporate decoding training at different stages. During the pre-listening phase, the teacher may select a few challenging or crucial sentences for students to practice decoding, followed by the main listening activity. In the post-listening phase, students could share the sentences they struggled to understand. The teacher could then offer some guidance and encourage the students to practice decoding these sentences again, while providing further explanations on the phonetic features within those sentences.

Moreover, the chatbot-supported online learning approach was more effective than the flipped learning approach to promote students' behavioral and emotional engagement over a 12-week period. In terms of behavioral engagement, weekly task completion of the chatbot-supported online learning group was similar with that of the flipped learning group for the first two weeks. This similarity within the first two weeks can be explained by students' boosted engagement when a new learning approach was introduced due to the novelty effect (Clark, 1983). However, as time went on, students in the chatbot-supported online learning group completed significantly more required tasks than those in the flipped learning group. In addition, the chatbot-supported online learning group showed a stable and relatively high interest level in the decoding practices over the whole semester. Although students might be initially interested to use the chatbot at the beginning of the semester due to possible novelty effect (Huang et al., 2022), we were encouraged to note that student interest levels were sustained throughout the entire 12 weeks.

The higher and sustained behavioral and emotional engagement in the chatbot-supported online learning can be explained with students' responses of open-ended survey. First, students' perceived competence of decoding the sentences were increased by chatbot's personalized support. One of the highly mentioned personalized support was the target feedback. With the common mistakes pre-set in the chatbot, the chatbot provided target hints for each student's specific mistakes. The students found the feedback very helpful, because it scaffolded them through the learning process. In contrast, the flipped learning group received only the pre-recorded teacher explanation videos, which discussed general common mistakes without specific guidance for individual learners. Feedback at the process level is typically more effective as it provides students with guidance on how to approach tasks and apply the acquired knowledge to other similar listening tasks (Hattie & Timperley, 2007). The positive outcomes in our study thus suggested that teachers can consider implementing chatbots to provide instant and personalized feedback to strengthen students' competence and sustain their behavioral engagement in online learning.

Second, students' increased and sustained engagement may be explained by their perceived high relatedness with the chatbot as a learning partner. In the open-ended survey, students indicated that the chatbot interacted with them like "a real person accompanying them throughout the semester." Unlike a human being (e.g., a teacher or a course mate), a chatbot is available 24/7 to converse with the students (Hew et al., 2022). In this study, the chatbot's human-like interaction was facilitated by the indicators of social presence, such as encouraging students and expressing emotions with emojis. According to the social response theory, people are inclined to treat computers as social beings (Nass & Moon, 2000), especially when the chatbot exhibits human-like behaviors (Holzwarth et al., 2006). We therefore advise future research to investigate the usefulness of social presence indicators in fostering students' need for relatedness in a chatbot-supported EFL learning.

To sum, these findings are surprising and encouraging given the fact that fully online courses are generally associated with lower levels of student engagement than on-campus or blended courses (Bai et al., 2022; Cavinato et al., 2021). Although a recent survey of 1,027 internet users reported that some 64% of respondents think that chatbots can replace teachers in the future (Rajnerowicz, 2024), we do not expect chatbot to completely replace traditional classroom-based education because it cannot understand the human context around students' learning struggles and replicate a real human connection (Kupperstein, 2023).

Limitations

This study is limited in the following ways. First, our participants were 89 first-year students from a Chinese university. Future studies can include students with different backgrounds (e.g., country or region, language proficiency, and age) to test the generalizability of our chatbot-supported fully online approach. Another concern is that this study specifically focused on EFL students' listening decoding ability. Future research is needed to investigate the use of chatbots for other English listening skills, which can help teachers and researchers make informed decisions about integrating chatbots into EFL teaching practices. Furthermore, although students acknowledged the benefits of chatbot-supported online learning, some (n = 9) found the daily task instruction tedious, especially later in the semester. This was because students had to manually select the learning tasks by clicking buttons instead of receiving the tasks directly from the chatbot. To improve the interaction between students and chatbot, we plan to customize the instructions to individual preferences and needs in the future.

Conclusion

Although the number of studies on chatbot-supported language learning has increased recently, little attention has been paid to its effect on students' listening skills. The present study extends the previous literature by using a well-designed chatbot to facilitate EFL students' listening decoding development. Self-determination theory and social presence were used as theoretical framework for our chatbot design. The results showed that students in the chatbot-supported online learning group performed equally well in decoding skills, and exhibited higher behavioral engagement, and sustained interest in the decoding practices, compared to the flipped learning group with teacher involvement. These encouraging results suggested that the chatbot-supported online learning approach can help overcome the challenges faced by teachers in our preliminary study (i.e., excessive workload and high demands on individual teachers' decoding skills). We therefore conclude that the chatbot, if carefully developed based on an appropriate theoretical framework, can provide language teachers with an alternative to free themselves from the excessive workload and better support students' online learning engagement.

This study also contributes to the growing body of research on the use of chatbots to facilitate student language learning in the following aspects. First, it compared a chatbot-supported learning approach to another learning approach (i.e., flipped learning) instead of a control group without equivalent support (e.g., Dizon, 2020; Kim, 2018). This comparison allows teachers to recognize the unique advantages of chatbot-supported learning over other active learning methods, such as flipped learning, and make a more informed decision when choosing an appropriate approach to facilitate students' language learning. Second, this study examined students' weekly participation in decoding activities and their emotional engagement at various time points, rather than solely assessing their performance at the end of the intervention. Such a research design allows educators to better understand the dynamics of student engagement in chatbot-supported learning environments and to design their instruction to support students at different stages of their language learning process.

References

Bai, S., Hew, K. F., Gonda, D. E., Huang, B., & Liang, X. (2022). Incorporating fantasy into gamification promotes student learning and quality of online interaction. *International Journal of Educational Technology in Higher Education*, *19*(1), 29. https://doi.org/10.1186/s41239-022-00335-9

Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, *32*(8), 827–877.

Bibauw, S., François, T., & Desmet, P. (2022a). Dialogue systems for language learning: Chatbots and beyond. In N. Ziegler & M. González-Lloret (Eds.), *The Routledge Handbook of Second Language Acquisition and Technology* (pp. 121–134). Routledge.

Bibauw, S., Van den Noortgate, W., François, T., & Desmet, P. (2022b). Dialogue systems for language learning: A meta-analysis. *Language Learning & Technology*, *26*(1), 1–24.

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2018). Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit? *Behavior Research Methods*, *50*(3), 937–962. https://doi.org/10.3758/s13428-017-0918-2

Bolboacă, S. D., Jäntschi, L., Sestraş, A. F., Sestraş, R. E., & Pamfil, D. C. (2011). Pearson-Fisher chi-square statistic revisited. *Information*, *2*(3), 528–545.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Psychiatric Quarterly, 3*(2), 77–101. https:// doi.org/10.1111/j.1460-2466.1978.tb01621.x

Cambridge Assessment English. (2021). *Cambridge English qualifications*. https://www.cambridgeenglish.org/exams-and-tests/qualifications/

Car, L.T., Dhinagaran, D.A., Kyaw, B.M., Kowatsch, T., Joty, S., Theng, Y.-L., & Atun, R. (2020). Conversational agents in health care: scoping review and conceptual analysis. *Journal of Medical Internet Research*, *22*(8), e17158.

Cavinato, A. G., Hunter, R. A., Ott, L. S., & Robinson, J. K. (2021). Promoting student interaction, engagement, and success in an online environment. *Analytical and Bioanalytical Chemistry*, *413*, 1513–1520. https://doi. org/10.1007/s00216-021-03178-x

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, *53*(4), 445–459.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268.

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

Field, J. (2008). *Listening in the Language Classroom*. Cambridge University Press.

Field, J. (2010). Listening in the language classroom. *ELT Journal, 64*(3), 331–333. https://doi.org/10.1093/elt/ccq026

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engage- ment: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109. https://doi.org/10.3102/00346543074001059

Garcia Brustenga, G., Fuertes-Alpiste, M., & Molas-Castells, N. (2018). *Briefing paper: Los chatbots en educación*. eLearn Center. Universitat Oberta de Catalunya.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2–3), 87–105.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Holzwarth, M., Janiszewski, C., & Neumann, M. M. (2006). The influence of avatars on online consumer shopping behavior. *Journal of Marketing, 70*(4), 19–36. https://doi.org/10.1509/jmkg.70.4.019

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 Higher Education*, 1–29.

Hsu, M. H., Chen, P. S., & Yu, C. S. (2021). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*, 1–12.

Huang, B., & Hew, K. F. (2018). Implementing a theory-driven gamification model in higher education flipped courses: Effects on out-of-class activity completion and quality of artifacts*. Computers & Education*, 125, 254–272.

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.

Jia, C., & Hew, K. F. (2021a). Toward a set of design principles for decoding training: A systematic review of studies of English as a foreign/second language listening education. *Educational Research Review*, 33, 100392. https://doi.org/10.1016/j.edurev.2021.100392

Jia, C., & Hew, K. F. (2021b). Meeting the challenges of decoding training in English as a foreign/second language listening education: Current status and opportunities for technology-assisted decoding training. *Computer Assisted Language Learning*, 1–30. https://doi.org/10.1080/09588221.2021.1974051

Jia, C., & Hew, K. F. (2022). Supporting lower-level processes in EFL listening: The effect on learners’ listening proficiency of a dictation program supported by a mobile instant messaging app. *Computer Assisted Language Learning*, *35*(1–2), 141–168. https://doi.org/10.1080/09588221.2019.1671462

Jia, C., Hew, K. F., & Li, M. (2023). Towards a flipped SEF-ARCS decoding model to improve foreign language listening proficiency. *Computer Assisted Language Learning*, 1–28.

Jia, J., Chen, Y., Ding, Z., & Ruan, M. (2012). Effects of a vocabulary acquisition and assessment system on students' performance in a blended learning class for English subject. *Computers & Education*, *58*(1), 63–76.

Kartal, G., & Simsek, H. (2017). The effects of audiobooks on EFL students' listening comprehension. *The Reading Matrix: An International Online Journal*, *17*(1), 112–123.

Ke, Z., & Wang, Y. (2022). Exploring the relationship between aural decoding and listening comprehension among L2 learners of English. *System*, 104, 102688. https://doi.org/10.1016/j.system.2021.102688

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, *10*(3), 2–10. https://doi.org/10.1007/ BF02905780

Kim, N.-Y. (2018). A study on chatbots for developing Korean college students' English listening and reading skills. *Journal of Digital Convergence*, *16*(8), 19–26.

Kim, N.-Y., Cha, Y., & Kim, H.-S. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning*, *22*(3), 32–53.

Kupperstein, J. (2023, June 15). AI can't replace high-quality teaching: Using the technology as a tool. *EDTech Digest*. https://www.edtechdigest.com/2023/06/15/ai-cant-replace-high-quality-teaching-using-the-technology-as-a-tool/

Leonard, K. R. (2019). Examining the relationship between decoding and comprehension in L2 listening. *System*, *87*, 102150. https://doi.org/10.1016/j.system.2019.102150

Lin, C. J., & 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.

Lo, C. K., & Hew, K. F. (2021). Developing a flipped learning approach to support student engagement: A design-based research of secondary school mathematics teaching. *Journal of Computer Assisted Learning*, *37*(1), 142–157.

Lynch, T., & Mendelsohn, D. (2010). Listening. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 180–196). Hodder & Stoughton Ltd.

Monteiro, V., Mata, L., & Peixoto, F. (2015). Intrinsic motivation inventory: Psychometric properties in the context of first language and mathematics learning. *Psicologia: Reflexão e Crítica*, *28*, 434–443.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Rajnerowicz, K. (2024, February, 13). AI in education: Will tech destroy education or save it. https://www.tidio.com/blog/ai-in-education/

Ruan, S., Willis, A., Xu, Q., Davis, G. M., Jiang, L., Brunskill, E., & Landay, J. A. (2019). Bookbuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. *Proceedings of the Sixth (2019) ACM Conference on Learning Scale* (pp. 1–4). ACM. https://doi.org/10.1145/3330430.3333643

Siegel, J. (2014). Exploring L2 listening instruction: Examinations of practice. *ELT Journal, 68(*1), 22–30. https://doi.org/10.1093/elt/cct058

Tsang, A. (2017). EFL/ESL Teachers' General Language Proficiency and Learners' Engagement. *RELC Journal*, *48*(1), 99–113. https://doi.org/10.1177/0033688217690060

Vandergrift, L. (2004). 1. Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24*, 3–25. doi:10.1017/S0267190504000017

Warschauer, M., & Liaw, M. L. (2011). Emerging technologies for autonomous language learning. *Studies in Self-Access Learning Journal*, *2*(3).

Wong, S. W. L., Mok, P. P. K., Chung, K. K., Leung, V. W. H., Bishop, D. V. M., & Chow, B. W. (2017). Perception of native English reduced forms in Chinese learners: Its role in listening comprehension and its phonological correlates. *TESOL Quarterly, 51*(1), 7–31. https://doi.org/10.1002/tesq.273

Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers & Education*, 161, 4059.

Zhang, S., Shan, C., Lee, J. S. Y., Che, S., & Kim, J. H. (2023). Effect of chatbot-assisted language learning: A meta-analysis. *Education and Information Technologies*, 1–21. https://doi.org/10.1007/s10639-023-11805-6

Appendix A. The Flipped SEF-ARCS Decoding Model (Jia et al., 2023)

Design Principles of the Flipped SEF-ARCS Decoding Model

Pre-Class Online Learning

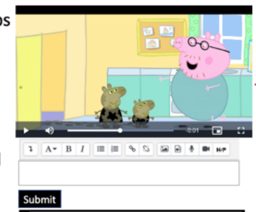
- **Engage:** engage students with various daily one-sentence video clip
- **Decode:** scaffold students to decode words from speech by allowing repeated listening, providing slow version and difficult word list.
- **Feedback:** explain the sentence by addressing its phonetic modifications and summarize the rules of phonetic modifications, using teacher-created short videos
- **Self-check:** self-check for successful decoding of each word in the clip without answer
- **Material selection:** the material for decoding training should be one-sentence interesting video clips, selected from various real-life sources (e.g., news report)
- **Incentives:** offer incentives (additional course grades) to motivate out-of-class learning
- **Differential learning:** allow students to choose differential practice according to their proficiency

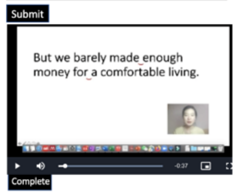
In-Class Face-to-Face Learning

- **Review:** Teacher reviews the common difficulties in the out-of-class decoding tasks at the beginning of in-class learning
- **Practice:** Provide more decoding practices in class.
- **Discuss:** discuss the answers with peers.
- **Feedback:** peers explain how they perceived the difficult sounds in the tasks; teacher summarizes and stresses the difficult features in the practice

Appendix B. The Treatments for Experimental and Control Groups in the Preliminary Study

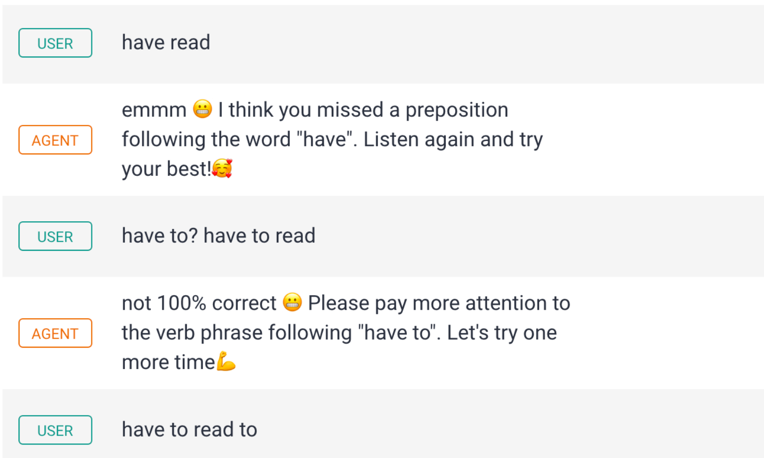
	EG using the flipped SEF ARCS decoding model		CG using the conventional decoding approach
out of class	• Each day, students are given one required task with two difficulty modes		• One required task, with no different mode.
	• Various one-sentence video clips are given. • Students are encouraged to watch the one-sentence video clip repeatedly (with slow version + difficult word list), write down what they hear and submit.	Engage Decode Material selection	• Same, but no scaffolds (e.g., slow video or difficult word list)
	• After submission, the correct answer in the form of the **audio-to-text transcript (task-level feedback)** and a **teacher-created explanation video (process-level feedback**, e.g., how to decode the sentence) are provided .	Feedback	• After submission, the correct answer in the form of the **audio-to-text transcript (task-level feedback)** is provided.
	• Students are directed to the original video again for self-check. If they listen out each word, click the button to finish.	Self-Check	• N/A
	• 5% of final grades	Incentives	• N/A
in- class	• Teacher reviews the tasks in the previous week and addresses students' common problems.	Review	• Teacher explains how to decode the sentences in the previous week's tasks **(process-level feedback),** as in the teacher-created explanation videos in EG.
	• Give students more similar practice • Ask them to work in pairs • Invite students to present answers and explain how they reach the answers • Teacher summarizes	Practice Discuss Feedback	• Same





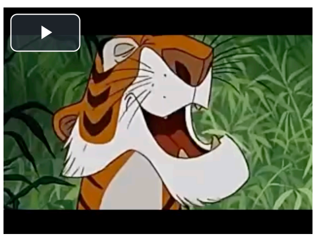
Appendix C. The Chatbot Detecting Student’s Input in Google Dialogflow

The entity contains the parameters that help the chatbot to understand students’ inputs. A fuzzy matching function in Google Dialogflow was used to help the chatbot to recognize students’ misspelled words or answers including parts of missing words. For example, students may enter verb phrases that are close to “have to”, such as "have”, “had to” or “has to”. These phrases were added to an entity called “have_to”, by which the chatbot was able to identify the meaning of students’ inputs using synonymous phrases. Before the chatbot was released for students’ online decoding learning, the teacher and the chatbot developer tested and improved it until no error was found. The figure below shows the chatbot detecting a student’s input regarding the required phase “have to read at”.



Appendix D. Materials of the Decoding Video Clips

Students were exposed to a variety of authentic materials from movies (e.g., “The Jungle Book”), public presentations (e.g., Ted Talks, or TV shows), and songs. These selections were made to ensure diverse themes, language styles, and cultural backgrounds to maximize the EFL students’ learning experience. The sections of audio used for the decoding practice contained one to two sentences from the authentic materials. These sentences were selected for their relevance to the phonetic modifications the students were expected to achieve in this compulsory English course. Examples of phonetic modifications include assimilation (e.g., ten people → tem people), elision (e.g., lea(ve) school → [li:’sku:l]), and weak forms (e.g., a → /ə/). For each decoding task, the missing words were approximate 10 words that contained 2 to 3 phonetic features of connected speech. The instructions of the decoding tasks were in Mandarin. Out of a total of 100 decoding videos provided over a period of fifty days (with one required and one optional task daily), three videos contained one new word, individually. The three new words were preselected by the teacher according to the curriculum. All students received the same new words in the three decoding tasks. We provided two examples of the decoding tasks in the table below.

Decoding video clips	**Decoding tasks (with missing words underlined)**
Video clip of a movie: 	<u>I just dropped by</u>. Oh, forgive me <u>if I’ve interrupted anything</u>.
Video clip of a song: 	Look at me. You may think <u>you see who I really am, but you never know me.</u>

Appendix E. Dictation Tests Used in the Study

The dictation materials were all authentic English news report from Voice of America (e.g., https://editorials.voa.gov/a/safe-water-haiti/4175506.html). The average speech rate was 140 wpm. Each partial dictation test consisted of five sentences, with a total of 25 missing words, and each word was counted as one point. The maximum score for each student's pre-test and post-test was 25 points. The missing words were selected from the Oxford 3000 key words, which, according to the teacher, were already familiar to most students. Although the missing words were different with each other, they shared common features of spoken connected English. An example of the transcripts provided is as follows: "<u>It is the start of</u> (iti the startof, [iti ðə sta:təv]) a longer-term, more comprehensive approach to water, sanitation, and hygiene that will strengthen <u>local and national</u> (localan national, [ləukə lən næʃənl]) Haitian institutions working in this sector." Students were expected to fill in the missing words, which are underlined in the example. Teacher played the audio of each sentence at normal speed three times while students wrote down the missing words. We provided a sample of the dictation test below.

Dictation Test

Name:

Student Number:

1.The United States ________________________ to support access to clean water for Haiti's citizens and ________________________ cholera.

2.The U.S. __________________________________, or USAID, Water and Sanitation project supports the United States' and Haiti's shared goal of________________________ and sanitation access to vulnerable communities.

3.The USAID Water and Sanitation ___________________________a 41.8 million dollar investment aligned with the priorities of Haiti's Ministry of _____________ and the water and sanitation directorate.

4. _____________________________________ a longer term, more comprehensive approach to water, sanitation, and hygiene that will strengthen ______________________ Haitian institutions working in th is sector.

5.The Project ___________________ the U.S. Government's __________________________ for a water-secure world.

Appendix F. Assumptions Checks for Statistical Analyses

We used the independent samples *t*-test to evaluate the differences of decoding skills for two groups in both pre- and post-tests. In terms of the pre-test, the results of Shapiro-Wilk's test showed students' decoding pre-test scores were normally distributed for both groups: W_pre (41) = 0.965, p = 0.243 for the flipped learning group and W_pre (45) = 0.961, p = 0.132 for the chatbot-supported online learning group. There was homogeneity of variances for students' pre-test in two groups, as assessed by Levene's test for equality of variances (p = 0.317). The post-test scores in both groups also showed a normal distribution: W_post (41) = 0.972, p = 0.408 for the flipped learning group and W_post (45) = 0.976, p = 0.388 for the chatbot-supported online learning group. Students' post-test scores had homogeneity of variances in both groups, as indicated by Levene's test for equality of variances (p = 0.298).

We used the one-way ANOVA to compare the differences in students' interest levels between the two groups at each time point. There was homogeneity of variances, as tested by Levene's test of homogeneity of variances (Time 1: p = 0.130; Time 2: p = 0.892; Time 3: p = 0.443). The Shapiro-Wilk test for normality is provided below. Although, the p-values for two data sets are greater than 0.05, and the remaining data sets are less than 0.05, the ANOVA test is robust to non-normal data and remains a valid statistical test under the condition of non-normality (Blanca et al., 2018).

	Shapiro-Wilk		
Group (Time)	**Statistic**	**df**	**Sig.**
Flipped (Time 1)	.900	40	.002
Chatbot (Time 2)	.950	42	.062
Flipped (Time 2)	.963	40	.207
Chatbot (Time 2)	.857	42	.000
Flipped (Time 3)	.886	40	.001
Chatbot (Time 3)	.777	42	.000

Appendix G. Summary of Chatbot Group Survey Findings

Theme	Example quotation
Engaging factor	
Personalized supports (n = 25)	"I preferred the chatbot's provision of two difficulty modes in each required task," "The in-progress feedback aided my comprehension of the sentences".
Various real-life materials (n = 21)	"The short videos are interesting and eye-catching," "The video contents cover various real-life conversation".
Teacher's explanatory videos (n = 21)	"The prompt teacher feedback in the explanation video helped my understand the phonetic modifications".
Interactive learning process (n = 5)	"The way we answered the decoding tasks with the chatbot was fascinating," "The interaction was like talking to a real person".
Encouraging learning climate (n = 4)	"The chatbot was always patient with me and never blamed me when I made mistakes".
Disengaging factor	
Tedious task instruction (n = 9)	"I found it slightly bothersome to have to review the same instructions for the required task every day".

Note. Forty-one students responded to the survey. *n* = number of student comments in the category.

About the Authors

Weijiao Huang holds a Degree of Philosophy from the University of Hong Kong. Her research focuses on student engagement, instructional design in blended and online learning, and technology-enhanced learning, especially using chatbots in learning environments.

E-mail: wjhuang1@connect.hku.hk

ORCiD: https://orcid.org/0000-0002-2095-1204

Chengyuan Jia is an assistant professor under the "100 Young Professor" scheme at the College of Education, Zhejiang University. Her research interests lie in technology assisted language learning, EFL listening, and student engagement in blended and online learning. Chengyuan Jia is the corresponding author.

E-mail: jiacy@zju.edu.cn

ORCiD: https://orcid.org/0000-0003-4214-2229

Khe Foon Hew is a Professor in the Faculty of Education at the University of Hong Kong. His research interests focus on understanding what engages people to learn in blended- and online-learning environments.

E-mail: kfhew@hku.hk

ORCiD: https://orcid.org/0000-0003-4149-533X

Jia Guo is a lecturer at the School of Foreign Languages, Northwest University of Political Science and Law. Her research interests lie in college students' creativity and college English language teaching.

E-mail: guojia528@163.com