ReCALL (2022), 34: 3, 327–343
doi:10.1017/S0958344022000039

RESEARCH ARTICLE

Implementation of an AI chatbot as an English conversation partner in EFL speaking classes

Hyejin Yang
Chung-Ang University, Republic of Korea (hjyang1112@gmail.com)

Heyoung Kim
Chung-Ang University, Republic of Korea (englishnet@cau.ac.kr)

Jang Ho Lee
Chung-Ang University, Republic of Korea (jangholee@cau.ac.kr)

Dongkwang Shin
Gwangju National University of Education, Republic of Korea (sdhera@gmail.com)

Abstract

With the growth of chatbots, concerns about implementing artificial intelligence (AI) chatbots in educational settings have consistently arisen, especially for the purpose of language learning. This study introduced a task-based voice chatbot called "Ellie", newly developed by the researchers, and examined the appropriateness of its task design and performance as an English conversation partner and students' perceptions on using it in EFL class. Korean EFL learners ($N = 314$) aged 10–15 years performed three speaking tasks with Ellie in their school classroom. The participants took 9.63 turns per session on average using the first 1,000-word band, indicating that the chatbot highly encouraged students to engage in conversation, which rarely occurs in general EFL classes in Korea. The high task success rates (88.3%) showed the design appropriateness of both L2 tasks and operational intents in terms of users' successful understanding and completeness of the given chatbot tasks. The participants' responses to the survey not only supported the positive potential of the chatbot in EFL settings but also revealed limitations to be resolved. Future suggestions for advancing and implementing AI chatbots in EFL classrooms are discussed.

Keywords: artificial intelligence; English as a foreign language; voice chatbot; speaking tasks; second language learning chatbot

1. Introduction

Chatbots are software applications that simulate human-like communication (Berns, Mota, Ruiz-Rube & Dodero, 2018; Fryer, Ainley, Thompson, Gibson & Sherlock, 2017). They have evolved dramatically with recent advances in artificial intelligence (AI) technologies, including machine learning algorithms, natural language processing, and speech synthesis (Janarthanam, 2017). Voice chatbots are now very close to enabling natural spoken conversation. As chatbots are able to run on such devices as mobile phones, smart speakers, or other Internet of Things (IoT) devices, their roles are becoming more concrete and contextualized. From personal assistants to customer service in a wide range of areas for business, medical, and other various organizations (Janarthanam, 2017), chatbots carry out *tasks* that fulfill users' needs (e.g. setting an alarm),

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

© The Author(s), 2022. Published by Cambridge University Press on behalf of European Association for Computer Assisted Language Learning.

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

and these tasks make chatbots more usable and beneficial to humans (Hu, 2019). The idea of task-based chatbots can also be applied to second language (L2) learning.

For several decades, chatbots have been considered potential conversation partners or language tutors for language learning (Atwell, 1999; Wang & Petrina, 2013). Some studies have reported increases in student interest and motivation in learning languages as benefits of implementing chatbots in English language teaching classrooms (Fryer & Carpenter, 2006; Kanda & Ishiguro, 2005). However, examples like Artificial Linguistic Internet Computer Entity (ALICE) and Kuki, widely recognized existing chatbots, are all conversational chatbots designed for first language (L1) users' general use. Despite previous studies on L1 chatbots, almost none have reported attempts to develop L2 chatbots that are designed solely for L2 learning and with L2 learner data, or to implement them in a language curriculum for a substantial amount of time (e.g. Jia & Ruan, 2008). More purposeful and interlanguage-aware L2 chatbots should be developed for the sake of creating positive effects on language learning and for sustained use. In particular, chatbot *tasks* need to be designed by identifying L2 users' needs and language proficiency, which also fit into language curricula.

Well-designed tasks can foster L2 learners' cognition processes and meaning negotiations that are essential for L2 acquisition (Ellis, 2003; Robinson, 2001). Effective pedagogic tasks need to be designed based on identified language needs in target domains. In particular, task type and complexity affect non-native speakers' oral performance in terms of the measure of communicative success (De Jong, Steinel, Florijn, Schoonen & Hulstijn, 2012). To promote L2 learners' successful task completion and language acquisition, tasks should be carefully designed considering topic familiarity, cognitive demand, functional adequacy, fluency, and lexical diversity. Similarly, in developing L2 chatbots, agents should be defined by selecting appropriate task types, and chatbot intent should be generated by analyzing interlanguage discourse patterns and other measures of task performance.

For this purpose, a task-based voice chatbot named "Ellie" had been under development for years by researchers in this study to determine whether adapting the idea of L2 pedagogic tasks into voice chatbot can work successfully in creating better speaking opportunities in English as a foreign language (EFL) classes in Korea. As an initial step, this study investigated the appropriateness of the chatbot task design and the effectiveness of the integration of task-based chatbots in English classrooms by measuring users' conversation turns, task success rates, and administering surveys about EFL students' perceptions of using chatbots for speaking practice. Finally, this study intended to provide insights into the future advancement of and directions for AI chatbots in the field of language education.

2. Literature review

2.1 Emerging AI chatbots and their application for language learning

The first versions of intelligent chatbots began with ELIZA, created by Joseph Weizenbaum in the 1960s, and Parry, developed by Kenneth Colby in the 1970s, which were programs that simulated human-like conversation in the form of texts (Janarthanam, 2017; Shum, He & Li, 2018). These early chatbots recognized keywords or phrases in users' input and provided pre-programmed, and thus limited, responses corresponding to those key expressions (Weizenbaum, 1966). In the 1990s, ALICE, developed by Richard Wallace, was able to maintain a more sophisticated conversation. This chatbot adopted Artificial Intelligence Markup Language (AIML), which identified topics, categories, and key patterns in users' input and utilized saved information to provide corresponding responses (Wang & Petrina, 2013).

Chatbots can be classified into text-based chatbots and voice-enabled chatbots by their mode of communication. Text-based chatbots interact with users via texts, and the aforementioned chatbots are representative text-based chatbots. In recent years, Cleverbot, created by Rollo Carpenter, is one of the most well-known chatbot applications. Cleverbot is unique in its ability to learn from users' input and to provide human-like responses that are not pre-programmed

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

(Torrey, Johnson, Sondergard, Ponce & Desmond, 2016). It automatically saves users' entire input and uses it to provide adequate responses to other users by determining how a user responds to the chatbot's input. Kuki, previously known as Mitsuku, developed by Steve Worswick, is another representative text chatbot adopting AIML files. Kuki simulates the most human-like conversation in the world, shown by its being a five-time Loebner Prize winner from 2013 to 2019. It is capable of providing human-like responses to users' input and even understanding mood in users' typed language.

Voice chatbots mainly interact with users via voice commands and responses. They have been widely implemented in personal devices like laptops and IoT devices that are connected with appliances and smart home systems. Voice chatbots are equipped in smartphone devices as stand-alone programs like Siri by Apple or applications like Google Assistant by Google, and Echo by Amazon. In response to users' commands or queries, voice chatbots in smartphones simply perform phone actions, check a user's schedule, or search the internet for news. Voice chatbots are also available in the form of mobile applications. For example, Lyra Virtual Assistant sets alarms and searches users' contacts or information on the web as a personal AI assistant. It allows users to chat with her by answering simple queries or even presenting her opinions on certain topics. Talk to Eve simulates natural human-like conversation by exchanging voice queries. However, it is restricted in its capability to process lengthy input from users (Kim, Cha & Kim, 2019). The aforementioned chatbots have served as general conversation chatbots that would be more suitable for native speakers of English to carry out daily life inquiries. Such limitations call for the necessity of further developing chatbots, especially for L2 learning.

With the exponential growth of chatbots, researchers have integrated them into language classrooms. In Fryer and Carpenter's study (2006), 211 university students using ALICE and Jabberwacky for learning English reported that the students felt comfortable and enjoyed using the chatbots. Coniam (2008) evaluated the ability of five chatbots, including Cybelle, Dave, George, Jenny and Lucy. Their capabilities as language learning tools, especially focusing on their linguistic accuracy, were evaluated based upon the following criteria: word-, sentence-, and text-level language structures. He found that these chatbots, chosen as currently the most capable chatbots, held the most human-like conversation and could be suitable for advanced language learners. However, he concluded that the chatbots were still insufficient to serve as ESL conversation partners from a language learning perspective.

Kim (2017) examined the effects of a commercial voice-based chatbot, called Indigo, on EFL students' negotiation of meaning according to their proficiency levels. This study compared student–student communication to student–voice chatbot communication in spoken interactions for 16 weeks. It examined different communication strategies used by the students depending upon their language proficiency. The results showed that the students who interacted with a voice chatbot exhibited more negotiation strategies than those in student–student communication groups.

Kim, Shin, Yang and Lee (2019) explored the potential use of commercial AI chatbots as conversation partners in EFL English classrooms. The participants interacted with Google Assistant and Amazon Alexa to perform three different tasks, namely exchanging small talk, asking for information, and solving problems. They found that both chatbots could serve as effective L2 conversation partners if well-designed language learning tasks are supplied. These studies pointed out that the chatbots provided students with language learning experiences to some extent, but this activity would result in meaningful learning only if well-designed tasks were integrated together with them. Coniam (2008) also highlighted the necessity of developing L2 learner-centered chatbots for English learners.

2.2 Recent development of AI chatbots for L2 learning

Concerns about implementing AI chatbots into L2 learning have consistently arisen and relevant chatbots have been studied. A virtual talking chatbot, called Computer Simulation in Educational

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

Communication (CSIEC), was developed and used for English learners (Jia & Chen, 2008; Jia & Ruan, 2008). After a six-month deployment of the chatbot in a high school, the students reported that the chatbot helped them feel more confident in using English and improved their listening skills, subsequently increasing their interest in language learning. Although the system allowed English learners to converse with the chatbot to some extent, the system focused more on providing automatic scores on restricted activities like gap-filling exercises, listening practices, or pattern drills in given scenarios.

Wang and Petrina (2013) presented a commercial chatbot called Lucy for English learners that offered language learners the chance to practice over 1,000 sentences on certain topics in terms of travel, hotels, and restaurants, to name a few. The chatbot provided learners with feedback on the pronunciation or grammatical accuracy of their utterances. Although this enabled learners to practice the target language to some extent, it was very close to pre-scripted pattern drills on given topics, which was far from being free conversation.

Chang, Lee, Chao, Wang and Chen (2010) also investigated the perspectives of instructors and students in an elementary school toward a robot equipped with five different modes, namely a storytelling mode, an oral reading mode, a cheerleader mode, an action command mode, and a question-and-answer mode. The findings emphasized the necessity of teacher training and well-prepared content that is suitable for the target learners' needs.

GenieTutor, developed by the Electronics and Telecommunications Research Institute in South Korea, is another AI chatbot developed for English learners (Choi, Kwon, Lee, Roh, Huang & Kim, 2017). The chatbot enabled English learners to practice diverse expressions on given topics at a restaurant or a shopping mall, among others. Although it provided feedback on English learners' grammar and pronunciation, the user–chatbot conversation was close to form-focused dialogues.

The implementation of AI chatbots into language learning has consistently arisen in research, but previous studies have revealed significant research gaps that need to be further investigated. Above all, the existing chatbots designed for language learning, like CSIEC (Jia & Chen, 2008; Jia & Ruan, 2008) or Lucy (Wang & Petrina, 2013), have centered more on pattern drills in certain situations where language learners have limited opportunities to practice diverse meaningful exchanges in target languages. Next, existing chatbots are still considered insufficient as L2 conversation partners, especially in EFL settings. Gallacher, Thompson and Howarth (2018) asserted that current chatbots were not fully able to generate follow-up questions or relevant, natural responses to develop continuous conversations.

The present study aimed to investigate the extent to which this task-based EFL chatbot, Ellie, could be effectively implemented as a conversation partner in EFL speaking classes and, further, to gather empirical data for improving chatbot design and performance. Three research questions were addressed:

RQ1: How did the task-based chatbot assist L2 learners to develop their conversation?
RQ2: How successfully were the three tasks completed by the participants?
RQ3: How was using Ellie in EFL class perceived by the participants?

3. Methodology

3.1 Participants

The participant were 314 English learners in South Korea, consisting of 177 students (fifth or sixth graders from 10 to 11 years old) from three elementary schools, and 137 first-year high school students (15 years old). The elementary school students included 77 male and 95 female students (five students did not answer the question), and the high school participants were male. Their English proficiency levels varied across different regions of South Korea, approximately ranging from beginners to intermediate-low levels. Individual students had diverse durations of English

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

learning experiences, but they officially began to learn English from the third grade of elementary school. They were selected for the convenience of sampling because the teachers voluntarily decided to implement Ellie in their classes. Since this study focused on the chatbot, and especially the appropriateness of its design and performance, user factors such as proficiency, age, gender, or language learning background were not strictly identified or filtered out.

3.2 Materials

3.2.1 An AI chatbot for English learning

The L2 learning voice chatbot named "Ellie" is an ongoing project funded by the National Research Foundation of Korea and was under development for three years by the authors. The chatbot was developed using the chatbot builder API platform *Dialogflow*™ by Google, whose fulfillment library is written in JS nodes. This API platform enabled developers to create and keep training chatbot agents by accumulating and analyzing user data through its own machine learning algorithms. The name of the chatbot was selected by researchers seeking English names both easily pronounced by Korean people and accurately recognized by speech-to-text technology of the chatbot. Since its first pilot test in spring 2019, the chatbot has been undergoing continuous advancements with accumulated user data and machine learning training. For more human-like conversation, the chatbot character was predefined as a female native English speaker and EFL teacher living in South Korea. The chatbot supports multiplatforms and web and mobile apps, using both iOS and Android, so that students can use laptops, tablet PCs or mobile phones in any place where a Wi-Fi connection is available. Prior to this study, each task for the chatbot was tested at least 10 times and cross-checked by eight people in the research team in order to ensure its function and capability. In the web platform "Talk with Ellie", 71 tasks are currently displayed under the name of "Mission Challenge". As seen in Figure 1, they are sorted into task complexity levels, from the easiest (Level 1) to the most difficult (Level 3), and task types in which users are expected to exchange conversation turns at least three to eight times for task completion depending on task complexity.

The target tasks were identified based on the K–12 English curriculum in South Korea. According to Jeon, Lee and Kim (2018), the Korean elementary-level national curriculum of English was in line with A1 to A2 levels of the Common European Framework of Reference for Languages (CEFR). When comparing the CEFR levels and the elementary-level (Years 3–6) achievement standards embedded in the national curriculum by utilizing a survey (related to communicative functions) and a performance test (related to performance levels), the elementary-level national curriculum showed a high correlation with Levels A1 and A2 of CEFR. Pedagogic tasks were designed considering target users' proficiency, task types (Willis, 1996), and the computer-assisted language learning (CALL) design principle (Chapelle, 2001).

Based on task-based language teaching (e.g. Skehan, 1998; Willis & Willis, 2001), chatbot tasks suggested goal-oriented and problem-solving tasks, all of which require cognition process and meaning negotiation (Ellis, 2003). Topics and contexts covered real-life situations that young EFL students are likely to encounter in an English speaking context (e.g. a telephone conversation or finding places). No fixed dialogue patterns are given, but rather specific missions that refer to task goals so the users can freely exchange their own meaning with the chatbot.

Three similarly designed tasks of Level 2 were selected from the tasks in the chatbot (see Table 1). The level was defined by complexity of decision trees (DT; Lobo, 2017) and the number of expected conversation turns to complete a task. In developing a chatbot task, the use of DT is effective in establishing task completion scenarios by predicting possible users' questions during the conversations with the chatbot (Kamphaug, Granmo, Goodwin & Zadorozhny, 2018; Shah, Jain, Agrawal, Jain & Shim, 2018). The target tasks included six to eight *intents*, which describe users' intention for one conversation turn with the chatbot, and a similar conversation flow structure consisting of seven to eight conversation turns at a minimum to fully

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

Table 1. Three tasks employed in this study

Tasks	Goals	Missions	Task complexity
Task 1	Taking orders at a restaurant	• To answer questions from a customer • To recommend dishes from the menu • To take an order for a drink	Level 2
Task 2	Shopping at a shoe store	• To give a size for shoes • To ask for a specific shoe color • To ask the price	Level 2
Task 3	Reserving a party room	• To select a room • To select a day and a time • To give the user's name and phone number	Level 2

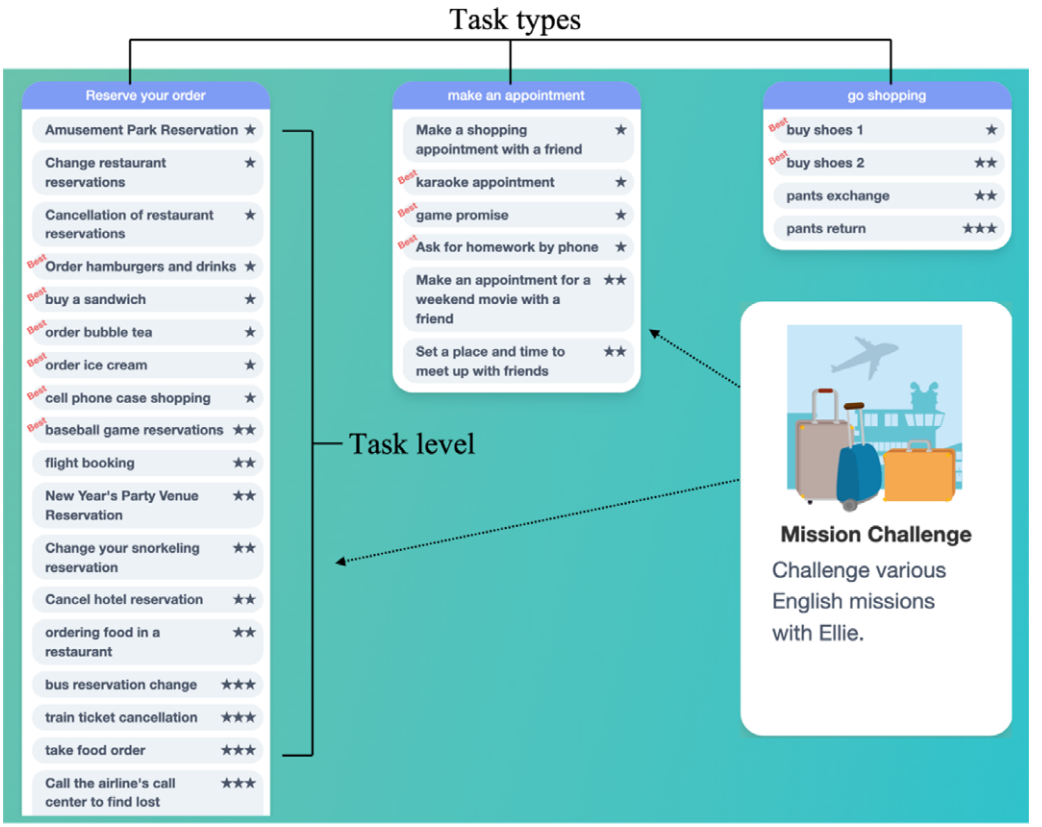


Figure 1. The first three menus of *Mission Challenge*

complete the three specific missions. In the conversation of Task 1 (see Figure 2), for example, students, as a waiter or waitress at a restaurant, were asked to recommend dishes to the chatbot. If a student's suggestion did not match the chatbot's preferences, the chatbot would request that the student recommend other dishes.

To talk to the chatbot, users choose a menu, click on the "Click and speak to Ellie" button at the bottom of the screen, and speak to the device. The student–chatbot conversations are displayed on the screen. Figure 3 presents examples of user–chatbot conversations for Task 1 and Task 2. It shows how a student negotiated meanings as the student recommended "pepperoni pizza" in

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

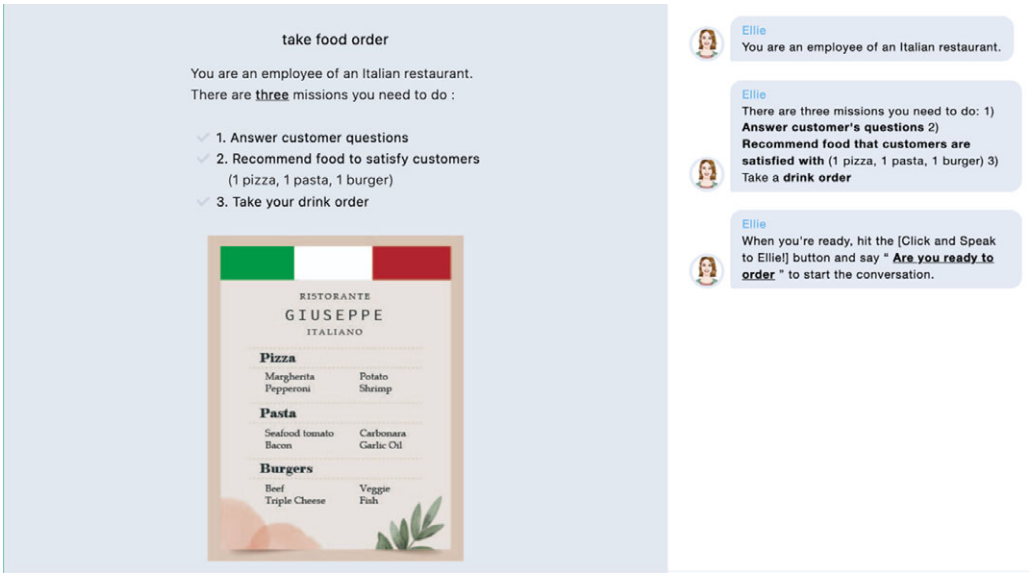


Figure 2. The screenshot of missions and directions of Task 1

response to the chatbot's talk ("Well, I don't prefer potato. Anything else?"). To complete all the tasks, the students needed to take multiple turns until they achieved the specific missions, consequently fostering students' negotiation strategies. The teachers used these tasks in their classes, and user data logs were recorded in all the classes.

3.2.2 Questionnaire

The questionnaire was administered to examine the students' perceptions of the chatbot as an English conversation partner (see supplementary materials for detailed information about the questionnaire). It included three multiple-choice questions about the students' background information (e.g. grades, whether they own chatbots or not, and previous experience using chatbots, if any), six 5-point Likert scale statements (1 = *disagree*, 5 = *agree*), and two open-ended questions about their perceptions of the chatbot.

3.3 Data collection procedures

Data were obtained from conversation logs of the student–chatbot conversations and the students' responses to the questionnaires. To collect the conversation logs, the students had a 10-minute orientation session where the teachers demonstrated how to talk with the chatbot on the first day of their chatbot use. Any questions from the students were resolved to confirm that the students comprehended how to use the chatbot. The frequency and duration of the chatbot use varied across schools. The students ($n = 132$) from two of the elementary schools used the chatbot only one time in an after-school English class, whereas those of the other elementary school ($n = 45$) and the high school students had two or three sessions with the chatbot in their regular classes over three weeks. The students then talked to the chatbot in groups of three or four using laptops, tablet PCs, or smartphones. The students took turns to talk with the chatbot because each group had one device. Thus, the conversation was one on one with the chatbot, not among peers. On average, they took approximately 10 to 15 minutes to carry out each task. The conversations were automatically converted to text logs and saved in the database of the chatbot. The text logs were then extracted from the database as Excel files for analysis.

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

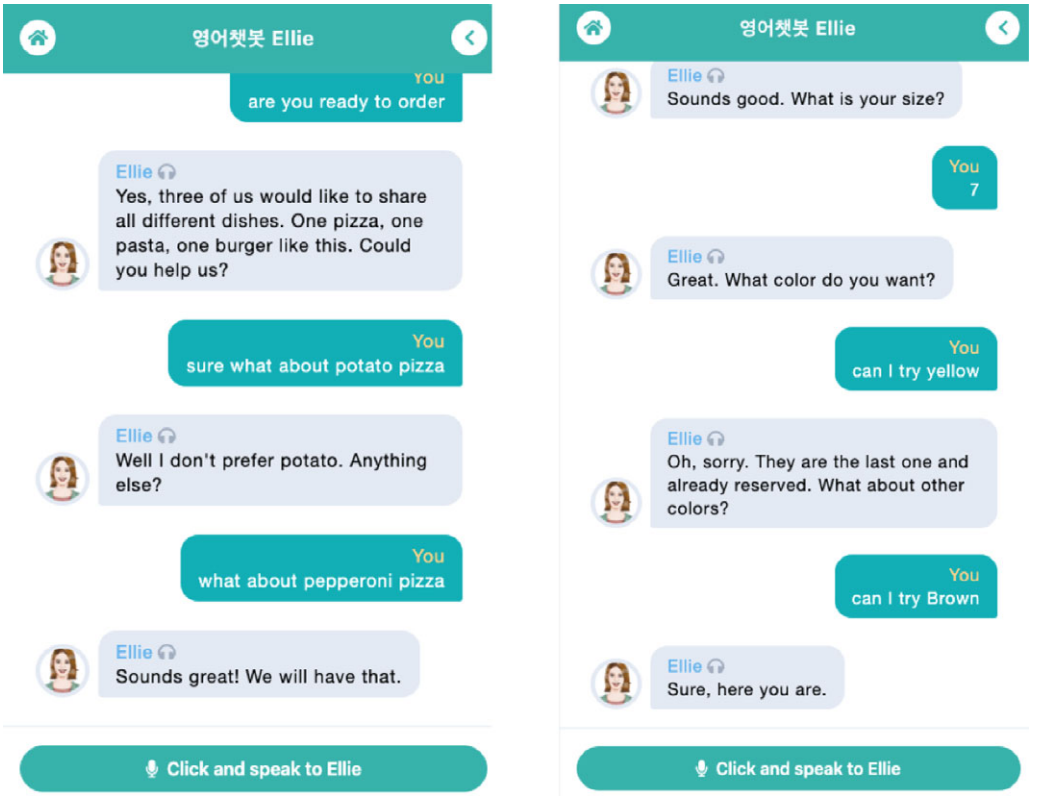


Figure 3. The screenshot of a student conversation with the chatbot for Task 1 and 2

The students' responses to the questionnaires were collected on the last day of their use of the chatbot. Paper-based questionnaires were distributed to the elementary school students, whereas the high school students took the same questionnaire online. Their responses were prepared as Excel files for subsequent analysis.

3.4 Data analysis

A mixed-methods approach (Creswell & Plano Clark, 2011) was employed by triangulating the data collected from different sources, including the conversation logs between the students and the chatbot, and the students' responses to the questionnaires. For the first research question, three elements were examined: conversation session, conversation-turns per session (CPS), and vocabulary level. Conversation session refers to the period from start to end of a user–chatbot conversation. CPS is the average number of dialogue exchanges between the student and the chatbot in a conversation session, and a higher CPS indicates a user's higher engagement in conversation with a chatbot (Shum *et al.*, 2018). CPS is often used as a success metric that determines the extent to which a chatbot is successful in maintaining conversations with users.

Students' vocabulary levels were examined to determine the extent to which the student–chatbot conversations were linguistically appropriate by employing a corpus analysis program: BNC-COCA 25,000 Range Program (Nation, 2012). This software can provide lexical profiles by using BNC-COCA 25,000, containing 25,000-word bands extracted from a 600-million-word compiled corpus with the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). In this study, only the first four graded word lists were applied to

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

the analysis because the first 4,000 words were considered to be adequate for measuring Korean high school students' competency as non-native English speakers (Joo, 2008).

The second research question evaluated the appropriateness of task designs. The three tasks were quantitatively and qualitatively reviewed by measuring task success rates for each task. Task success rate is another metric used to measure chatbots' performance (Shum *et al.*, 2018). Adopting this concept, the task success rates of this study were calculated using the following formula:

$$task\ success\ rates = \frac{\text{successful conversation sessions}}{\text{total conversation sessions} - \text{invalid conversation sessions}}$$

Herein, the term total conversation sessions refers to the sum of all the conversation sessions recorded in user logs for each task. Invalid conversation sessions included the sessions in which the users' voices were not successfully recognized or in which the users terminated the conversation abruptly for unidentified reasons. The invalid conversation sessions, taking up 45.6% of the total conversation sessions recorded in the user logs, were excluded to calculate the task success rates because such errors were unpredictable and hard to control in classroom settings. Successful conversation sessions indicated the sum of sessions where the users fully completed the three specific missions of each task.

The users' task success was further examined to determine the reasons for the failure of the users' successful task completion. Conversation sessions that qualified as unsuccessful task completion were initially extracted from the total conversation sessions. Then, the authors identified a coding scheme that could explain the reasons for unsuccessful task completion. The coded results were cross-checked and confirmed through peer discussion among the two coders.

The third research question about the students' perceptions of the chatbot was examined through descriptive statistics and content analysis of the questionnaires. Descriptive statistics were computed to analyze the students' responses to 5-point Likert-scale statements. Content analysis was conducted to analyze students' written responses to two open-ended questions to obtain more in-depth information about the students' perception of the chatbot. In the first cycle coding, initial coding was carried out to explore students' perceptions by assigning categories to each segment (Saldaña, 2016). The categories emerging from the students' responses were then classified to create the coding scheme. For the second cycle coding, focused coding was adopted, as the other coder cross-checked the segments in the categories that were defined in the initial coding procedure. Any disagreements between the two coders were resolved through discussion.

4. Results

4.1 The students' interaction with the AI chatbot, Ellie

The first research question examined how well the chatbot assisted L2 learners to develop their conversation. The number of conversation sessions, CPS, and vocabulary levels were obtained for each task.

4.1.1 Conversation sessions and conversation-turns per session

A different number of conversation sessions were held for each task. Table 2 shows the number of conversation sessions, the average of the CPS for each task and overall averages. The students engaged in the highest number of conversation sessions to complete Task 2 (152 sessions), followed by Task 1 (83 sessions) and Task 3 (48 sessions). As for the CPS, each conversation session lasted 9.63 turns on average to carry out the tasks between students and the chatbot. To specify, students displayed the highest CPS in completing Task 1 (11.5 CPS), suggesting that students were more engaged in performing Task 1 than Tasks 2 and 3, given the fact that a higher CPS number refers to the students' engagement in conversation. It is also noteworthy that the

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

Table 2. A summary of conversation sessions and conversation-turns per session (CPS) for each task, with averages

Tasks	Conversation sessions	CPS
Task 1	83	11.5
Task 2	152	8.2
Task 3	48	9.2
Average	94	9.63

Table 3. Lexical profiles of students' utterances

	Task 1 (%)	Task 2 (%)	Task 3 (%)	Average (%)
1st 1,000	69.27	88.31	71.2	76.26
2nd 1,000	6.56	2.87	14.79	8.07
3rd 1,000	.36	.97	.79	.71
4th 1,000	1.45	.58	.96	1.00
Subtotal	77.64	92.73	87.74	86.04
5th and above	22.36	7.27	12.26	13.96
Total tokens	**3,232**	**3,952**	**1,784**	**2,989**
Total tokens/session	**38.94**	**26**	**37.17**	**33.45**

students took between 8.2 and 11.5 turns to hold the conversations required for each task, which rarely happens in normal EFL classrooms, which lack opportunities to speak English.

4.1.2 Vocabulary levels in the students' utterances

The students' lexical usage was analyzed to explore how the chatbot facilitated students' conversation development. The words making up the students' utterances were examined at the levels of the first to fourth 1,000-word bands in BNC-COCA 25,000, as shown in Table 3.

Overall, students produced 2,989 tokens on average to carry out the three tasks and 33.45 tokens per session. As for the vocabulary level of the students' utterances, Nation (2006) claimed that the 1st 1,000 words account for 81%–84% of spoken data, including conversations in everyday life, and the 2nd 1,000 words represent 5%–6%. Given this, 76.26% of the students' utterances belonged to the 1st 1,000 words, meaning that the chatbot assisted the participants in using expressions in daily life situations.

The students showed the largest number of utterances in Task 2 with 3,952 tokens, followed by Task 1 with 3,232 tokens, and Task 3 with 1,784 tokens. Looking at the students' lexical profiles, the coverage of words beyond the first 4,000 words in Task 1 (22.36%) was much higher than the 4%–6% of normal conversation, indicating that Task 1 might require users to use more topical or context-specific vocabulary.

In brief, the use of the chatbot encouraged the students to develop their conversation via the three tasks, given the fact that the students took approximately 9.63 turns per session on average. It seems very likely that this average CPS was higher than that of L2 learners at the same levels in typical English classes. In addition, the students' predominant use of the 1st 1,000-word band supported that the chatbot fostered the students' appropriate use of English in daily situations that were covered in the national English education curriculum of South Korea.

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

Table 4. The task success rate for each task

Tasks	Task success rate (%)	Valid total conversation sessions	Successful conversation sessions
Task 1	81.4	43	35
Task 2	90.9	77	70
Task 3	91.2	34	31
Average	88.3	154	136

Table 5. Major issues that affected task success rates

Major issues	Task 1 (%)	Task 2 (%)	Task 3 (%)	Average (%)
Inappropriate responses ($n = 12$)	62.5	85.7	33.3	60.5
Technological flaws ($n = 6$)	37.5	14.3	66.7	39.5

4.2 Students' task success rates

The second research question examined the extent to which the target chatbot tasks were completed successfully. Task success rates were obtained by calculating the percentages of successful conversation sessions among the valid total conversation sessions, as shown in Table 4. Overall, high task success rates were witnessed, approximately 88.3% on average. The students showed the highest task completion rate for Task 3 (91.2%), followed by Task 2 (90.9%) and Task 1 (81.4%). This indicates that Task 3 appeared to be the easiest, whereas Task 1 was the most difficult task.

The task success rates were further examined to identify major issues that hindered successful task completion. Two major issues were found through qualitative text analysis of the student–chatbot conversations, as shown in Table 5.

The major issue affecting the successful completion of conversations was inappropriate responses on the part of students, which made up 60.5% on average. Inappropriate responses were most often observed in Task 2 (85.7%), followed by Task 1 (62.5%) and Task 3 (33.3%). To specify, some students misunderstood the tasks, so they selected options that were not listed in the tasks or skipped to perform specific missions. In Excerpt 1 from Task 2, for example, the students were required to select a color of shoes based on a given list in Task 2. However, it turned out that some students picked the name of a color ("White") that was not on the list, as shown in line 5. Therefore, the chatbot was unable to respond in line 6.

Excerpt 1. Examples of inappropriate responses (Task 2)

1 [student] my shoe size is 20 and 19 in
2 [chatbot] Great. What color do you want?
3 [student] height
4 [chatbot] What was that?
5 [student] White
6 [chatbot] I missed that, say that again?

Students' responses, including offensive words, also contributed to unsuccessful task completion in student–chatbot conversations. This result sheds light on the necessity for appropriate and continuous guidance that could lead students to have successful conversations with the chatbot.

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

The other issue was some flaws in the design of the DT for the tasks, which accounted for 39.5% of task non-completion on average. The design flaws were observed the most frequently in Task 3 (66.7%), followed by Task 1 (37.5%) and Task 2 (14.3%). A typical flaw was that the chatbot sometimes repeated the same questions to the students although the questions had already been answered. In Excerpt 2, for example, the chatbot asked the students to select the room type twice, as shown in lines 2 and 8 ("Ok, here are four rooms you can choose"). However, the student selected a sunflower room in line 3. The student ended up closing the talk when the chatbot asked the same question in line 8, which could have affected the students' willingness to maintain the conversation.

Excerpt 2. Example of the inappropriate decision tree design for Task 3

1	[students]	so is it possible to make a reservation in a 24th of December right
2	[chatbot]	Ok, here are four rooms you can choose.
3	[students]	so I want sunflower
4	[chatbot]	Great. Please tell me the day and the time that you want.
5	[students]	I choose 18th on December
6	[chatbot]	I am sorry, the room is already booked. Any other preference?
7	[students]	so are you make a reservation 23rd December
8	[chatbot]	Ok, here are four rooms you can choose.

This finding could be explained by the fact that each task contained three missions to be completed in order. To accomplish each mission, a student must negotiate with the chatbot to find the appropriate day and time, as shown in line 4. In the process of the student's negotiation, the chatbot sometimes asked the student to perform a mission that was already completed, as in lines 2 and 8. This result provides a future direction regarding DT in that missions should be accomplished regardless of certain pre-planned sequences in the chatbot.

4.3 Students' perception

The third research question explored the students' perception of the chatbot as an English conversation partner. Six statements were scored on 5-point Likert scales. In addition to summarizing the participants' responses to each scale point, we further calculated the mean and standard deviation of participants' responses to each statement, which could reveal the participants' overall perception of the chatbot, as shown in Table 6. Overall, the students had a neutral to positive attitude toward the chatbot ($M = 3.40$, $SD = .60$). The students appeared to feel comfortable when talking with the chatbot, given the mean score on the statement ($M = 3.66$, $SD = 1.15$), the highest score among the remaining statements. The perceived area of lowest positive perception was whether talking with the chatbot was the same as talking with a human being ($M = 3.29$, $SD = 1.02$).

The students' responses to the open-ended questions provided more comprehensive insights to explain the students' perceptions of the chatbot. The majority of the students appeared to express a positive impression of the chatbot. The students frequently mentioned the strengths of the chatbot, indicating "The chatbot's ability to understand and respond to their talk" (36.0% out of 314 survey responses) and that it was "fun and interesting" (28.0%):

I can talk with the chatbot like a friend. (student 50)
I felt joy and good when the chatbot understood my talk and replied. (student 154)
It is fun to talk in English with the chatbot. (student 266)
It understood and answered to my talk. (student 137)

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

Table 6. Average and percentage of the students' perception of the chatbot

Statements	*M* (*SD*)	Strongly disagree (%)	Disagree (%)	Neutral (%)	Agree (%)	Strongly agree (%)
I feel comfortable when talking with the chatbot ($n = 311$)	3.66 (1.15)	4.2	10.0	34.4	19.0	32.5
I think talking with the chatbot is the same as talking with human beings ($n = 312$)	3.29 (1.02)	6.0	16.4	41.3	25.3	11.2
It is easy to understand the chatbot's talk ($n = 309$)	3.30 (1.17)	9.1	12.6	35.3	24.9	18.1
It is fun to talk with the chatbot ($n = 310$)	3.40 (1.08)	5.5	11.9	38.1	26.5	18.1
I want to talk more with the chatbot ($n = 310$)	3.42 (1.08)	5.2	11.9	37.4	26.8	18.7
Talking with the chatbot helps me improve my English skills ($n = 310$)	3.63 (.99)	3.5	5.2	37.7	31.6	21.9

Note. 1 = *strongly disagree*; 5 = *strongly agree*.

They also reported that the chatbot helped them improve their English speaking skills (11.1%). This finding was in line with their opinions about the usefulness of the chatbot in improving their English skills ($M = 3.63$, $SD = .99$), as shown in Table 6. The students appreciated having chances to talk in English and to expand the scope of their vocabulary knowledge. This result suggested that the chatbot had assisted the students to practice useful vocabulary and expressions, consequently helping their conversation development:

> Talking with the chatbot helps me enhance English speaking skills. (student 250)
> I could learn new expressions from the chatbot. (student 210)
> I learned vocabulary while talking with the chatbot. (student 262)
> I had to think about vocabulary to talk with the chatbot. I could improve vocabulary. (student 275)

On the other hand, the students shared some difficulties in using the chatbot. The difficulties can be categorized into three aspects, namely (1) the chatbot could not understand the students' utterances (58.0% out of 314 responses), (2) the chatbot's speech was sometimes too fast or too long for them to understand (6.7%), and (3) it was difficult to understand some expressions in the chatbot's speech (4.5%):

> The chatbot did not understand my talk. (student 127)
> The chatbot talks too fast and difficult. (student 45)
> It was difficult to understand when the chatbot talked too much. (student 168)
> Some words are difficult to understand. I hope they are translated into Korean. (student 143)

More than half of the students (58.0%) mentioned that their utterances were sometimes not recognized successfully. Factors including noise in the classrooms or the students' unclear pronunciation could explain the students' difficulties.

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

5. Discussion and conclusion

Overall, the task-based AI chatbot seemed to build a positive learning environment for EFL learners. First, the students developed authentic conversations in English with the chatbot by exchanging 9.6 conversational turns on average, which they would be unlikely to experience in EFL classes in Korea. Meaning negotiation must have been involved while exchanging messages for social purposes. Meanwhile, the students tended to employ vocabulary that is frequently used in daily life situations: 76.26% of the vocabulary in students' conversations with the chatbot occurred within the first 1,000 words of BNC-COCA 25,000, suggesting that use of the chatbot encouraged students to use expressions that were linked to the vocabulary list in the national English curriculum of South Korea.

Next, high task success rates were found across the three tasks (88.3% on average), meaning that the chatbot offered opportunities for the students to practice the target language via achievable task goals. Among the three tasks, Task 3 turned out to be the easiest (91.2%) whereas Task 1 was the most difficult (81.4%), which had been expected from the task complexity. The qualitative analysis of the students' unsuccessful conversation sessions revealed two significant issues that interrupted successful task completion: students' inappropriate or offensive responses to the chatbot's inquiries and technical flaws in the DT for the tasks. The observed issues shed light on the necessity of appropriate and ongoing guidance from instructors and future enhancement of task design in the chatbot.

Finally, students showed a neutral to positive perception of the chatbot as an English conversation partner, generally believing that the chatbot helped them improve their speaking skills and they were comfortable chatting with it in English. This result was in line with early discussions on the benefits of AI chatbots on language leaners' learning processes. For example, Chiu, Liou and Yeh (2007) maintained that voice chatbots could enrich English learners' language learning experiences by expanding their opportunities to practice speaking in English.

The present study exhibits several limitations. Above all, approximately 45.6% of the conversation sessions recorded in the chatbot were excluded to obtain the task success rates. The invalid conversation sessions involved cases in which the user's voice was not recognized successfully and those in which the conversation terminated for unidentified reasons. This result was associated with the survey question in which 58.0% of the students pointed out the limitations in the chatbot's voice recognition. In future research, this issue could be resolved by assigning students to work in relatively small groups in the classes or by adopting platforms with more advanced voice recognition capability. Moreover, only three tasks were utilized to investigate the effectiveness of the chatbot as an English conversation partner. A longitudinal study utilizing diverse tasks and participants could offer new insights to teachers and researchers by revealing the long-term effects of the application of an AI chatbot in English classrooms. Next, this study focused on the appropriateness of the chatbot task design by examining students' overall performance and their perceptions. However, previous studies presented divergent conclusions about the effectiveness of chatbots depending upon the learners' proficiency levels, age groups, or gender difference. For example, some studies reported that voice chatbots were beneficial for lower-level students (Lee, 2001; Nakahama, Tyler & Van Lier, 2001; Rosell-Aguilar, 2005), whereas others believed they were good for higher-level students (Kötter, 2001; Stockwell, 2004). Future research thus needs to take students' different proficiency levels, age groups, or gender into account. Furthermore, the students in this study completed the tasks in groups and responded individually to the questionnaires. Although the group work could lead students to lessen their anxiety when they talked in English with the chatbot, students' individual interaction with the chatbot could also be further examined in future research. Lastly, flaws in the DT for the tasks were the other weakness that need to be resolved. In particular, the DT for Task 3 should be checked carefully to determine what caused the chatbot to ask the same questions repeatedly during the conversation, which led to a breakdown of communication.

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

Despite the limitations, the findings were fairly positive. The tasks appeared to help the students to continue meaningful and lengthy conversations more easily and comfortably in English, regardless of their proficiency level. Moreover, the participants negotiated meanings with their limited vocabulary and successfully achieved non-linguistic task goals. From the aspect of future development in AI chatbots for English learning, this study proposes guidance to CALL researchers and developers for designing task types that could promote L2 learners' speaking practice using chatbots. With the EFL context generally being an input-poor one, we believe that chatbots have significant potential to provide L2 learners with greater opportunities to receive English input as well as to use English through chatbot-based tasks. In addition, our chatbot is expected to cater to the needs of the learners with diverse English proficiency levels by virtue of offering L2 tasks with different difficulty levels. This characteristic of our chatbot appears to allow us to overcome the limitations of existing chatbots, which have been suggested to be effective only for learners at particular levels of English proficiency. Constant attention and research centering on the development and the application of AI chatbots are called for.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0958344022000039.

Acknowledgements. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A2A03037255).

Ethical statement and competing interests. We testify that the content of this paper presents an accurate account of the work performed as well as an objective discussion of its significance. This research project was conducted in accordance with and the approval of the institutional review board of the institution of this study. Participants were volunteers. There are no conflicts of interest.

References

Atwell, E. (1999) *The language machine: The impact of speech and language technologies on English language teaching*. London: The British Council.

Berns, A., Mota, J. M., Ruiz-Rube, I. & Dodero, J. M. (2018) Exploring the potential of a 360° video application for foreign language learning. In García-Peñalvo, F. J. (ed.), *TEEM'18: Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality*. New York: Association for Computing Machinery, 776–780. https://doi.org/10.1145/3284179.3284309

Chang, C.-W., Lee, J.-H., Chao, P.-Y., Wang, C.-Y. & Chen, G.-D. (2010) Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educational Technology & Society*, 13(2): 13–24. http://www.jstor.org/stable/jeductechsoci.13.2.13

Chapelle, C. (2001) *Computer applications in second language acquisition*. Cambridge: Cambridge University Press.

Chiu, T.-L., Liou, H.-C. & Yeh, Y. (2007) A study of web-based oral activities enhanced by automatic speech recognition for EFL college learning. *Computer Assisted Language Learning*, 20(3): 209–234. https://doi.org/10.1080/09588220701489374

Choi, S.-K., Kwon, O.-W., Lee, K., Roh Y.-H., Huang, J.-X. & Kim, Y.-G. (2017) English tutoring system using chatbot and dialog system. In Korea Information Processing Society (ed.), *Proceedings of the Korea Information Processing Society Conference*. Seoul National University of Science and Technology, 3–4 November, 958–959. https://doi.org/10.3745/PKIPS.Y2017M04A.958

Coniam, D. (2008) Evaluating the language resources of chatbots for their potential in English as a second Language. *ReCALL*, 20(1): 98–116. https://doi.org/10.1017/S0958344008000815

Creswell, J. W. & Plano Clark, V. L. (2011) *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks: SAGE.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R. & Hulstijn, J. H. (2012) The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In Housen, A., Kuiken, F. & Vedder, I. (eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins, 121–142. https://doi.org/10.1075/lllt.32.06jon

Ellis, R. (2003) *Task-based language learning and teaching*. Oxford: Oxford University Press.

Fryer, L. & Carpenter, R. (2006) Bots as language learning tools. *Language Learning & Technology*, 10(3): 8–14. https://doi.org/10125/44068

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

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

Gallacher, A., Thompson, A. & Howarth, M. (2018) "My robot is an idiot!" – Students' perceptions of AI in the L2 classroom. In Taalas, P., Jalkanen, J., Bradley, L. & Thouësny, S. (eds.), *Future-proof CALL: Language learning as exploration and encounters: Short papers from EUROCALL 2018*. Voillans: Research-publishing.net, 70–76. https://doi.org/10.14705/rpnet.2018.26.815

Hu, Y. (2019) *Do people want to message chatbots? Developing and comparing the usability of a conversational vs. menu-based chatbot in context of new hire onboarding*. Aalto University, unpublished master's thesis.

Janarthanam, S. (2017) *Hands-on chatbots and conversational UI development: Build chatbots and voice user interfaces with Chatfuel, Dialogflow, Microsoft Bot Framework, Twilio, and Alexa Skills*. Birmingham: Packt.

Jeon, J.-H., Lee, J.-Y. & Kim, J.-R. (2018) Development of survey to inquire learners' awareness of language competence based on CEFR basic user level. *Asia-Pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 8(3): 199–210.

Jia, J. & Chen, W. (2008) Motivate the learners to practice English through playing with chatbot CSIEC. In Pan, Z., Zhang, X., El Rhalibi, A., Woo, W. & Li, Y. (eds.), *Technologies for e-learning and digital entertainment: Edutainment 2008: Vol. 5093: Lecture notes in computer science*. Berlin: Springer, 180–191. https://doi.org/10.1007/978-3-540-69736-7_20

Jia, J. & Ruan, M. (2008) Use chatbot CSIEC to facilitate the individual learning in English instruction: A case study. In Woolf, B. P., Aïmeur, E., Nkambou, R. & Lajoie, S. (eds.), *Intelligent tutoring systems: ITS 2008: Vol. 5091: Lecture notes in computer science*. Berlin: Springer, 706–708. https://doi.org/10.1007/978-3-540-69132-7_84

Joo, H. W. (2008) *A corpus-based analysis of vocabulary in the BEWL and the CSAT*. Seoul: Korea University, unpublished master's thesis.

Kamphaug, Å., Granmo, O.-C., Goodwin, M. & Zadorozhny, V. I. (2018) Towards open domain chatbots—A GRU architecture for data driven conversations. In Diplaris, S., Satsiou, A., Følstad, A., Vafopoulos, M. & Vilarinho, T. (eds.), *Internet science: INSCI 2017: Vol. 10750: Lecture notes in computer science*. Cham: Springer, 213–222. https://doi.org/10.1007/978-3-319-77547-0_16

Kanda, T. & Ishiguro, H. (2005) Communication robots for elementary schools. In *AISB'05: Proceedings of the Symposium on Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction*. Brighton: The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, 54–63.

Kim, H., Shin, D. K., Yang, H. & Lee, J. H. (2019) A study of AI chatbot as an assistant tool for school English curriculum. *Journal of Learner-Centered Curriculum and Instruction*, 19(1): 89–110. https://doi.org/10.22251/jlcci.2019.19.1.89

Kim, N. (2017) Effects of types of voice-based chat on EFL students' negotiation of meaning according to proficiency levels. *English Teaching*, 72(1): 159–181. https://doi.org/10.18095/meeso.2017.18.1.03

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

Kötter, M. (2001) Developing distance language learners' interactive competence—Can synchronous audio do the trick? *International Journal of Educational Telecommunications*, 7(4): 327–353.

Lee, L. (2001) Online interaction: Negotiation of meaning and strategies used among learners of Spanish. *ReCALL*, 13(2): 232–244. https://doi.org/10.1017/S0958344001000829a

Lobo, J. (2017) *What is a decision tree and why should my chatbot use it?* https://www.inbenta.com/en/blog/decition-tree-chatbot/

Nakahama, Y., Tyler, A. & Van Lier, L. (2001) Negotiation of meaning in conversational and information gap activities: A comparative discourse analysis. *TESOL Quarterly*, 35(3): 377–405. https://doi.org/10.2307/3588028

Nation, I. S. P. (2006) How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1): 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2012) *The BNC-COCA word family lists*. http://www.victoria.ac.nz/lals/about/staff/publications/BNC_COCA_25000.zip

Robinson, P. (2001) Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1): 27–57. https://doi.org/10.1093/applin/22.1.27

Rosell-Aguilar, F. (2005) Task design for audiographic conferencing: Promoting beginner oral interaction in distance language learning. *Computer Assisted Language Learning*, 18(5): 417–442. https://doi.org/10.1080/09588220500442772

Saldaña, J. (2016) *The coding manual for qualitative researchers*. Thousand Oaks: SAGE.

Shah, A., Jain, B., Agrawal, B., Jain, S. & Shim, S. (2018) Problem solving chatbot for data structures. In *Proceedings of 2018 IEEE 8th Annual Computing and Communication Workshop and Conference*. Las Vegas, 8–10 January, 184–189. https://doi.org/10.1109/CCWC.2018.8301734

Shum, H., He, X. & Li, D. (2018) From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1): 10–26. https://doi.org/10.1631/FITEE.1700826

Skehan, P. (1998) *A cognitive approach to language learning*. Oxford: Oxford University Press.

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press

Stockwell, G. (2004) Communication breakdown in asynchronous computer-mediated communication (CMC). *Australian Language and Literacy Matters*, 1(3): 7–10, 31.

Torrey, L., Johnson, K., Sondergard, S., Ponce, P. & Desmond, L. (2016) The turing test in the classroom. In Leake, D. (ed.), *Proceedings of the Sixth Symposium on Educational Advances in Artificial Intelligence*. Palo Alto: AAAI Press, 4113–4118.

Wang, Y. F. & Petrina, S. (2013) Using learning analytics to understand the design of an intelligent language tutor–chatbot Lucy. *International Journal of Advanced Computer Science and Applications*, 4(11): 124–131. https://doi.org/10.14569/IJACSA.2013.041117

Weizenbaum, J. (1966) ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery*, 9(1): 36–45. https://doi.org/10.1145/365153.365168

Willis, D. & Willis, J. (2001) Task-based language learning. In Carter, R. & Nunan, D. (eds.), *The Cambridge guide to teaching English to speakers of other languages*. Cambridge: Cambridge University Press, 173–179. https://doi.org/10.1017/CBO9780511667206.026

Willis, J. (1996) *The framework for task-based learning*. Harlow: Longman.

About the authors

Hyejin Yang received her PhD in applied linguistics and technology from Iowa State University and is currently a full-time researcher at Chung-Ang University, South Korea. Her research interests include CALL and language testing.

Heyoung Kim received her PhD in second and foreign language education from the State University of New York at Buffalo and is currently a professor at Chung-Ang University, South Korea. Her research interests include CALL, digital literacy, and task-based learning. All correspondence regarding this publication should be addressed to her.

Jang Ho Lee received his DPhil in education from the University of Oxford and is presently an associate professor at Chung-Ang University. His work has been published in *Applied Linguistics*, *The Modern Language Journal*, *ELT Journal*, *TESOL Quarterly*, *Language Teaching Research*, *Language Learning & Technology*, *System*, among others.

Dongkwang Shin received his PhD in applied linguistics from Victoria University of Wellington and is currently an associate professor at Gwangju National University of Education, South Korea. His research interests include corpus linguistics, CALL, and chatbot-based language learning.

Author ORCiD. Hyejin Yang, https://orcid.org/0000-0002-7819-7945
Author ORCiD. Heyoung Kim, https://orcid.org/0000-0001-7428-3945
Author ORCiD. Jang Ho Lee, https://orcid.org/0000-0003-2767-3881
Author ORCiD. Dongkwang Shin, https://orcid.org/0000-0002-5583-0189

https://doi.org/10.1017/S0958344022000039 Published online by Cambridge University Press