

ARTICLE



## The application of chatbot as an L2 writing practice tool

*Suh Keong Kwon, Chinju National University of Education*

*Dongkwang Shin, Gwangju National University of Education*

*Yongsang Lee, Inha University*

### Abstract

*This study investigates the effect of chatbot-based writing practices on second language learners' writing performance and perceptions of using the chatbot in L2 writing practices. A total of 75 Korean elementary school students were randomly allocated to two groups. While the control group received traditional teacher-led writing instruction, the experimental group used a chatbot for individual writing practices for 15 weeks. The chatbot was developed using Google's Dialogflow machine-learning AI platform by encoding expressions from an elementary school English textbook. A pretest was carried out prior to the experiment to examine the initial writing performance, and a posttest was carried out 15 weeks later with a different writing topic. The participants in the experimental group also responded to a short survey to report their perceptions and opinions about the chatbot. The results showed that the two groups generally showed a similar writing proficiency in the pretest scores, but the experimental group performed significantly better in the posttest than the control group, suggesting that the chatbot-based writing practice had a facilitating effect on their test performance. The participants of the experimental group also found the chatbot useful in improving their language skills and made them feel comfortable when learning a foreign language.*

**Keywords:** *Chatbot, Artificial intelligence in language learning, Machine-learning language learning platform, Pedagogical agent*

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

**APA Citation:** Kwon, S. K., Shin, D., & Lee, Y. (2023). The application of chatbot as an L2 writing practice tool. *Language Learning & Technology*, 27(1), 1–19. <https://doi.org/10.125/73541>

### Introduction

The increasing reliance on artificial intelligence (AI) technology today has drastically changed our ways of life. Until recently, robots as personal assistants had only existed either in our imaginations or in science fiction films (e.g., TARS from *Interstellar*, Joi from *Blade Runner 2049*). It should be noted that most AI bots in such films are fluent multilinguals, which implies that they are capable of learning multiple languages more quickly and accurately than humans can. Likewise, AI technology is replacing humans in various fields of our lives, particularly because machine-human interactions are becoming more feasible and realistic. The interactive and communicative functions of AI are featured in robots that are currently being used in customer service (Shum et al., 2018), as mental health therapists (Vaidyam et al., 2019), or as personal assistants (e.g., Google Assistant or Siri).

Chatbot is a robot that can converse with human beings. Dehn and van Mulken (2000) defined a chatbot as a character on the computer screen that produces realistic speech, emotions, locomotion, and body movements. While chatbot technology has been pervasively used in various fields for both commercial and academic purposes, its use in language education is yet at a pioneering stage. Early studies on this topic, for example Lu et al. (2006), proposed the development of a chatbot and demonstrated a number of concepts

for a computer application performing an online conversation with a human via text or voice for language-learning purposes. Lu et al. (2006) insisted that second-language (L2) learners could better understand the material when interpersonally conversing with a partner instead of receiving information from a human instructor. In addition, the instant messaging feature of a chatbot was expected to provide L2 learners with more opportunities to use and practice a target language without time and place constraints. Recently, Fryer et al. (2020) systematically reviewed overall history, current stage, and future potentials of developing chatbots for language learning with two major conventional chatbots, Cleverbot and Mondly, as examples. They claimed that both Cleverbot and Mondly have not yet made enough impact as a foreign-language-learning tool despite its usefulness. They also argued that chatbots for a foreign-language-learning purpose is at a “revolutionary stage” that will continue to develop, and there are still “untapped areas” for the chatbot developers to make “small adjustments” to improve the technology (Fryer et al., 2020, p. 13, 17). Despite continuous efforts and advancement of features, chatbots are still a weak supplementary language-learning tool at the present stage (Fryer et al., 2019).

With the rapid development of Internet of Things technology, intelligent personal assistants (IPAs) are receiving substantial attention. IPAs are task-performing chatbots that make phone calls, manage personal schedules, or search for and report information upon the user’s request. However, to date, the degree to which IPAs perform tasks and provide services is yet limited, mostly because most of them cannot engage in conversation with a human at a discourse level. Exceptionally, XiaoIce, developed by Microsoft, is a socializable chatbot that can perform dialogue-attempt functions, such as prompting a new topic for a conversation with the user. Although there are a number of other socializing-focused chatbots available, they have mainly been developed for conversing in a user’s first language (L1) rather than for L2 learning. Therefore, chatbots for L2 learning, particularly for writing practice, are not sufficiently available in the market yet. The present study examined how much the use of a chatbot may contribute to improving the English writing skills of L2 learners. In the past, the social chatbot was developed with a greater focus on the role of a conversation partner but not under a specific purpose such as foreign-language learning. Nevertheless, Lu et al. (2006) devised a plan to incorporate a social chatbot into education. To this end, Lu et al. (2006) proposed a design of a chatbot for foreign language education using instant messaging. In an English-as-a-foreign-language (EFL) environment where exposure to the target language or opportunities to use the language are limited, the idea that any L2 learner can have an AI native speaker as a conversation partner and practice the target language without time and space constraints may spark a huge change.

## Literature Review

### Development of Chatbots as Pedagogical Agents

Before discussing the present use and potential of chatbots in language learning, it is essential to cover a brief history of how such technology has evolved so far. The root of AI goes back to 70 years ago, when Alan Turing (1950) published the paper *Computing Machinery and Intelligence*, in which he asked the question *can machines think?* What this question asks is whether a computer can trick human beings into believing that they are conversing with another human interlocutor, not a computer. This question is still being used to test the degree to which AI programs and human beings are alike, known as the *imitation game* or *Turing test*. Since then, the key aim of chatbot research and development has been to develop a chatbot that possesses “human tendencies such as attitudes, language, reactions, and mannerisms” (Savin-Baden et al., 2015, p. 296). A number of attempts have been made to develop various chatbots that can imitate human beings as closely as possible. The notion of *animism* was proposed by Frude (1983), who suggested that human beings and computers would form beneficial relationships if combined effectively (Frude & Jandrić, 2015). Although many recent chatbot studies are still optimistic about Frude’s (1983) notion of animism (e.g., Savin-Baden et al., 2015), a number of studies also argue that Turing’s ultimate goal of a human computer is difficult to achieve because we have not established a consensus on the definition of humanity (Kiran & Verbeek, 2010; Veletsianos & Russell, 2014).

Theoretically, chatbots as pedagogical agents have been referred to as tools that have humanistic qualities

and support learners' cognitive development. In fact, the benefits of a chatbot go beyond its efficiency or cost-effectiveness. Fryer and Carpenter (2006) illustrated a total of six advantages of using chatbots in language learning. They pointed out that chatbots are useful because (a) learners feel more relaxed and comfortable talking to them compared to another human being; (b) the same materials and drills can be repeated (see also Fryer et al., 2019; Fryer et al., 2020); (c) both listening and reading skills can be practiced; (d) they are novel and facilitate learners' motivation; (e) they provide a variety of language expressions; and (f) they provide prompt feedback, particularly on mechanical errors. Among these six benefits, the relaxed and comfortable conversation exchange with a chatbot is often attributed to the emotional connection between the user and the chatbot. This benefit is also mentioned as one of the two reasons why students prefer practicing with a chatbot instead of a human partner (Fryer et al., 2020). With regard to this, Savin-Baden et al. (2013) claimed that trustworthiness is an essential component in the formation of a connection between the two. In a later study, Savin-Baden et al. (2015) argued that the emotional connection formed between users and pedagogical agents is affected by the extent to which the users can personalize or relate to the agent. In other words, the degree to which learners trust the agent and disclose information is closely associated with how much the agent is personalized. Savin-Baden et al. (2015) also found that the length of engagement with the pedagogical agent and the topic of conversation are significant factors that have an effect on the level of emotional connection formed between users and agents. For example, they revealed that the users of the chatbot disclosed more information when they had longer engagements and on sensitive topics (e.g., sexual health, drugs). The qualitative findings of this study further indicated that the participants disclosed more information to an agent because the agent was "almost" like a human being, but not "judgemental of responses to sensitive questions" (Savin-Baden et al., 2015, p. 308). Questions remain whether such a chatbot can be considered one of the valid examples of AI. We believe that the decision should be made based on the degree to which machine learning (ML) is involved in the software algorithm. For example, there are a number of chatbots that work under fixed scenarios only, but in the case of free conversational chatbots (e.g., Mitsuku [now called Kuki] or Cleverbot), technologies related to natural language processing (NLP) and ML are applied. We argue that the latter can be considered an AI, but not the former. However, recently developed scenario-based chatbots can now cover a wider range of natural language because NLP and ML have already been applied to a default database embedded in the chatbot builder (e.g., Dialogflow). Therefore, the present study considers the scenario-based chatbot with NLP and ML technology as one of the AI programs.

With the rapid development of AI technology, chatbots are becoming more and more like human beings. Nass and Yen (2010) argued that we exchange conversation with computers in the same way as we do with other human beings, expecting the same conversational interaction as in our previous experiences with humans. After a decade since this claim was made, we now exchange more realistic conversations with computers with the advancement of technology. The realistic nature of chatbot technology has shifted the perceptions and attitude of today's learners and has become a part of their lives, including in the educational arena.

### **Chatbot Studies in Language Learning**

Existing studies that have looked into the advantages of using chatbots for language-learning purposes are largely categorized into two types. One type explores the feasibility of using existing commercialized chatbots in language education, and the other type examines the effectiveness of chatbots developed for language learning.

In terms of the feasibility of using a commercial chatbot, Kim (2017) examined elementary school students' interaction with Amazon's AI speaker, Echo. For this, Kim (2017) developed three levels of tasks: (a) a controlled task (e.g., to say "How is the weather in London?") where the question presented by the teacher is repeatedly used, (b) a guided task to solve with some given information (e.g., find out the meaning of "workaholic" using the pattern "What is ~?"), and (c) an independent task where the learner tries to independently solve the given request (e.g., find out what the capital of the United States is). Participants in this study were allocated into two groups and performed these three levels of tasks in different conditions

(teacher-directed speaking vs. chatbot speaking). The results showed that the group that used the AI chatbot made more speech attempts over time, while showing lower levels of anxiety. Shin (2019) compared the feasibility of using two of the most advanced free-conversation chatbots, namely Mitsuku and Cleverbot, for teaching and learning English in an EFL context. A total of 27 college students exchanged conversation with the two chatbots according to the given tasks, and 90% or more utterances produced by the two chatbots during these conversations were within the 3,000 most frequently used words. The participants of this study preferred Mitsuku to Cleverbot mainly because Mitsuku offered more opportunities for, and exposure to, English use while reducing their anxiety about foreign-language learning at the same time. However, Kim et al. (2019) criticized that conventional chatbots, such as Mitsuku or Cleverbot, are mainly developed to handle short conversations and that they are not sufficiently ideal for language learning. On the other hand, Shin (2019) argued that chatbot technology at the current level is sufficient to be used as an effective language-learning tool, although it has not reached the human-conversation level. Further investigation on the use of chatbots for language-learning purposes are required to establish a consensus on the feasibility of applying the technology to the field of language learning.

With regard to the discussion, a number of studies have attempted to develop new chatbot systems. Jia and Chen (2009) tentatively applied and evaluated the interactive web-based human-computer dialogue system CSIEC (Computer Simulation in Educational Communication). Key features of the CSIEC include (a) the availability of both text and voice chat, (b) spelling and grammar checks upon the user's request, and (c) an automatic topic-generator for interactive chatting. After applying the chatbot in an actual learning environment, Jia and Chen (2009) found that the users preferred not to use the spelling and grammar check feature, but used the chatting mode with a topic given by the chatbot instead. The participants believed that using the CSIEC's application in English teaching and learning could motivate the learners to use English and facilitate their learning process at the same time. Kwon et al. (2015) developed a chatbot called Genie Tutor and conducted small-scale experiments to examine the accuracy of feedback on meaning and grammar. Findings showed that the accuracy of Genie Tutor's feedback on meaning was 87.8% and on grammar was 79.2%. However, since the topics used in Genie Tutor were limited in diversity and the structure of the conversation was designed in advance mainly for vocabulary or syntax-based learning, the chatbot was more like a conversation practice program rather than an authentic conversation partner. Both CSIEC and Genie Tutor were developed as chatbots for language learning, and they focus on providing plenty of language input and feedback appropriate to the learner's language proficiency level rather than free conversation. On the other hand, chatbots such as Mitsuku and Cleverbot mentioned above are social chatbots that were developed with a focus on free conversation as a conversation partner rather than language learning, and the data accumulated through conversations with users were implemented to enable more sophisticated conversations through ML.

A number of recent studies have attempted to develop or investigate the effect of different features of chatbots in L2 learning context. Shawar (2017) discussed previous approaches to develop natural dialogue systems under a computer-assisted language-learning (CALL) environment. Shawar (2017) argued that the CALL system needs to be integrated with the chatbot as a conversational partner to enable L2 learners to practice language. As a model, the study introduced a conversation agent chatbot using ALICE, an AIML (Artificial Intelligence Markup Language) technique, and argued that the chatbot has great potential as a future language-learning domain. Pham et al. (2018) introduced a chatbot that was developed particularly for English-learning purposes. The key features of the mobile-app chatbot, called English Practice, were reminding learners to study or assisting the learners to learn new vocabulary and answer multiple-choice questions. The chatbot developed by Pham (2018), however, does not involve any ML algorithms and thus cannot be considered as an AI-based chatbot. In terms of language learners' perceptions on replacing a real conversation partner with a chatbot, Jia (2009) distributed the ALICE chatbot to Chinese college students. When the experiment began, the participants of the study were not informed that they were chatting with a chatbot, but they soon realized that the conversation partner was a machine after exchanging a few messages. Most of them (88%) chatted with the chatbot only once for a short period of time, and 24% of them expressed a negative perception on the use of ALICE as a language-learning device. It should be noted that

the chatbot technology (ALICE and many other new ones) has improved since Jia (2009) conducted the experiment. Chatbot-based learning is also claimed to improve learners' memory retention. Abbasi and Kazi (2014) compared college students' learning outcomes and memory retention through a chatbot and Google Search Engine. The experimental group used the chatbot to find the solution of the given problem while the control group used Google Search Engine. It turned out that even though both groups made a significant improvement between the pre- and posttests, the experimental group performed significantly better and showed longer memory retention than the control group. This shows that the contextual understanding through conversations with a chatbot could contribute to learners' long-term memory.

In terms of recent approaches to developing a chatbot using Dialogflow, Sung (2019) developed a flowchart-based English-speaking chatbot focusing on the role of negotiation-of-meaning functions. The chatbot used in Sung's (2019) study was a work-in-progress prototype, and the study did not empirically investigate its use with actual L2 learners.

In terms of the effectiveness of chatbots, Coniam (2014) evaluated five representative chatbots, comparing the accuracy of their lexical and grammatical uses and the amount of output produced by each chatbot. He found that 69–78% of the words generated by the five chatbots belong to the most frequent 2,000 words with only a few mechanical errors (e.g., spelling, upper/lowercase). The grammatical accuracy of the chatbots was also very high (77–93%), but with regard to meaning fit, there was a large variability across the different chatbots (47–71%). In terms of word meaning and grammar fit, the appropriateness rate was even lower at 44–66%. Finally, the total number of tokens of the chatbots' utterances was substantially different according to chatbot, ranging from 912 to 3,258 words. Taken together, the findings of Coniam (2014) provided an insight into the development of criteria for examining the effectiveness of chatbots for language-learning purposes. Yang et al. (2019) examined the potential of the AI chatbot, Ellie, as a language practice agent for young learners in an EFL context. In the study, 177 elementary school students in Korea were asked to perform two types of conversation tasks with Ellie. The results showed that the students mostly made successful conversation with Ellie by employing several types of conversation strategies. However, there were some technical difficulties associated with the voice recognition feature of the chatbot and its lengthy utterances. Nevertheless, most of the students responded positively toward Ellie as a conversation partner for English-speaking practice.

In summary, existing studies have examined the potential and effectiveness of chatbots for language learning, but little is known about the degree to which chatbot learning might affect L2 learners' improvements in language skills. Previous studies have employed conventional chatbot programs such as Alexa and Cleverbot, but these were neither technically developed for language-learning purposes nor are they targeted for low-proficiency EFL learners such as the participants in this current study (Kim, 2017; Shin, 2019). In addition, there are still only a few empirical studies that have examined the effectiveness of using chatbots over a long-term period. To this end, this present study developed a chatbot that was systematically designed based on lessons in an English textbook and conducted an experiment with elementary school students over the course of a semester to examine the degree to which chatting with the chatbot might contribute to the improvement of the learners' English writing skills. With the findings of previous studies in mind, we hypothesize that (a) chatbot writing practice will improve the participants' writing performance and (b) the chatbot will be perceived as a useful, non-face-threatening language-learning tool by L2 learners.

The research questions guiding this study are as follows:

Research Question 1: To what extent does chatbot writing practice have an impact on L2 learners' writing performance?

Research Question 2: To what extent do L2 learners of English perceive chatbot writing as a language-learning tool?

## Research Methods

### Participants

The participants of this study were fifth-grade elementary school students in South Korea aged between 10 and 11. Of the total sample, 38 were male and 37 were female. A total of 75 students were sampled from two intact classes that were randomly selected. Both classes were taught by the same teacher and used the same English textbook. All participants' L1 was Korean, except for one student who had a multicultural background in which English is also a foreign language. In Korea, English is taught as a mandatory subject at school starting in the third-grade during elementary school. Therefore, at the start of this study, it was the students' third year of learning English through public education. Elementary school students receive 40-minute English lessons three times per week, but most of them also take extra English lessons outside of school such as at private cram schools. Most participants also received English learning opportunities outside of school, and 70% of them have experienced using a chatbot for a short period of time.

The two sampled intact classes were assigned to control and experimental groups. A total of 38 students were in the control group and 25 of them (65.8%) attended English cram schools. On the other hand, there were 37 students in the experimental group, and 26 of them (70%) attended cram schools for extra English classes. During the 15-week intervention period, both groups practiced traditional writing activities (e.g., filling in the blank, translating a sentence from their L1 into the L2), but the experimental group was given a chatbot-based writing task as an extra task to practice the expressions they learned. The experiment was conducted in compliance with research ethics with the consent of all participants and their parents.

### Instruments

#### *The English Writing Chatbot*

Unlike existing studies that have employed a variety of language-use scenarios and tasks, the materials used in the present study targeted comparably low-level young EFL learners. Therefore, developing an original chatbot program was crucial to implement specific language-learning tasks. The chatbot developed for this present study was specifically designed to function as a conversational partner that interacted with participants based on what they learned in their English classes at school. Thus, the target expressions were extracted from the English textbooks that the participants actually used in class. The chatbot used for this present study was developed by the researchers using Google's (2020) Dialogflow, an open-source chatbot-builder that has ML capabilities. The researchers of this study manually encoded the target language expressions that were to be learned in English classes in school and designed algorithms for effective interactions between the participants and the chatbot. In encoding the language expressions for the chatbot-based learning, a Grade-5 English textbook (Lee et al., 2018) was used, which is what the target participants were using when this study was conducted.

#### *The Writing Task*

The English textbook used in this study was one of five government-authorized English textbooks for elementary schools in Korea, and each chapter consisted of listening, reading, speaking, and writing lessons focused on a particular topic and target expressions. Each chapter was organized into six to seven lessons, with each lesson dealing with one or two specific language skills. The present study specifically used the writing tasks of the textbook, which were the fifth and sixth lessons of each chapter—the reading and writing lessons—and the expressions and simple tasks extracted from a total of five chapters were encoded to the chatbot. It should be noted that the type of language the students were taught and had practiced in class was mostly spoken language, such as writing dialogue responses, while the pre- and posttests focused on paragraph writing. The aim of elementary English education in Korea is to encourage students to familiarize themselves with the language and acquire communicative language skills. For this reason, the reading and writing expressions that are present in the English textbooks are also taught mainly through spoken language. That is, regardless of the genre of the given task, students are likely to show a tendency to use spoken language in their writing.

The writing tasks developed for this study were fully controlled scenario-based dialogues based on key expressions learned in class. Table 1 presents the five chapters that were selected for chatbot learning in this study.

**Table 1**

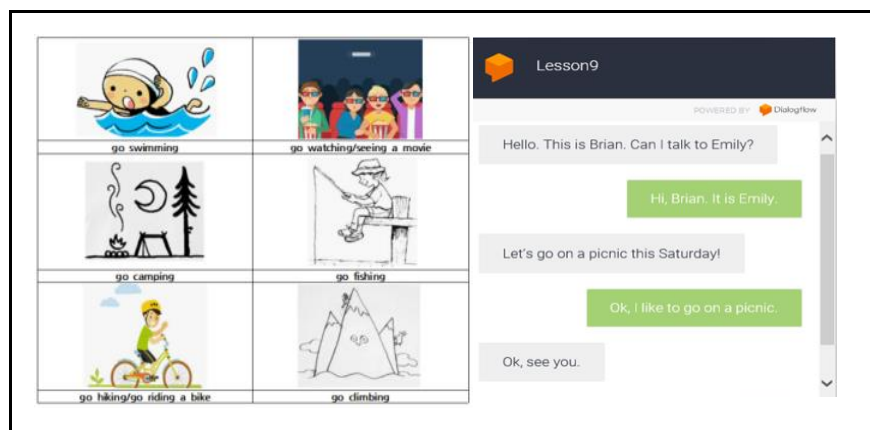
*Writing Tasks and Key Expressions Employed in the Chatbot Writing System*

Chapter	Writing task scenario	Key expressions
8	Descriptions of appearances	colors, types of clothes, hairstyles
9	Suggesting leisure activities over the phone	This is NAME, go + gerund(-ing)
10	Giving directions	go straight, turn right, turn left, next to
11	Different types of job	I want to be + a(an) JOB,
12	Future plans for a vacation	I will + infinitive

The five target chapters respectively addressed expressions related to descriptions of appearances, suggesting leisure activities over the phone, giving directions, different types of job, and future vacation plans. Each task was based on a scenario consisting of three conversation turns on average, and consisting of six to eight subtasks with different topics (e.g., giving directions: bakery, post office, zoo, bank). The tasks were given to participants during in-class English lessons prior to the data collection phase. The tasks and key expressions were carefully selected to investigate the ways in which participants individually practiced the target expressions they had learned. A sample task from Chapter 9 with an actual conversation between the chatbot and a participant in the experimental group is presented in Figure 1.

**Figure 1**

*Screenshot of a Sample Conversation Between the Chatbot and a Participant*



Chatbot writing tasks that use the English expressions identical to the traditional writing lessons were given to participants in the control group as well, but they were given the traditional *writing words and phrases*, *completing sentences*, and *writing sentences on my own* activities that were presented in the textbook. These passive activities were oriented toward practicing the target expressions on a sentence level but not on a discourse level. To measure the test-takers' writing proficiency, a total of two short writing tests were developed. The first task was used as a pretest, and the second task was used as a posttest. Both tests were

carefully controlled in terms of the writing type, difficulty, and administration in order to minimize other variables intervening, except for the treatment—chatbot practice. The topic for the pretest was *the most difficult moment of my life*, and the topic for the posttest was *the happiest moment of my life*. Specifically, the participants were asked to write (a) about the most difficult/happiest time of their lives, (b) what happened at that time, and (c) how they felt at that time. All instructions for both writing tasks were given in Korean, the L1 of the participants, and the tests in both conditions were conducted in a timed condition (25 minutes). To score the participants' writings submitted for the two tests, a holistic rating was conducted using Shin et al.'s (2012) holistic 3-point scoring rubric. The scoring rubric was initially developed to diagnose the English proficiency of elementary school students in Korea. The rubric measures content, vocabulary, grammar, and mechanics (e.g., spelling, punctuation), while organization of logical development was excluded from the scoring factor because it required a relatively short answer (see [Appendix](#) for specific guidelines).

### Survey

A survey on the use of chatbot writing was given to the experimental group's participants in order to investigate their general perceptions of the usage of the technology. Participants in the experimental group were asked to reflect on their own experiences using the chatbot writing program and their perceptions and opinions about using it for L2 learning. The control group was only asked demographic information, which included their previous experience using chatbots and the number of hours they spent studying English outside the classroom. All the questions were given in Korean, the L1 of the participants. [Table 2](#) presents the questions used in this study.

**Table 2**

#### *Structure of Survey Items*

Questions	No. of items	Type of items
Demographic information	6	Multiple choice
Chatbot text-chat behavior	2	Open-ended; multiple choice
Perceptions of chatbot	7	5-point Likert scale
Reflection on the chatbot learning experience	2	Multiple choice
Reflection on the chatbot learning experience	3	Open-ended

### Data Collection Procedure

The participants in both groups were given a pretest at the beginning of the semester before they were taught the five lessons that were used to practice with the chatbot in the experiment. Based on the pretest results, we ensured that there was no significant difference between the two conditions at the initial stage of this study. Throughout the semester, participants in both groups studied the five chapters using the same English textbook and were taught by the same teacher. Only the experimental group participants were given the chatbot writing lessons as an additional task and exchanged text-chat conversations with the chatbot for about 10–15 minutes as an individual activity at the end of the classes. This chatbot-based writing activity was carried out once or twice every two weeks, which was usually the fifth or sixth lesson for each chapter of the textbook. Specifically, the experimental group participated in conversations with the chatbot 10 times as part of the regular classroom curriculum during a period of 15 weeks. Some participants used the chatbot to practice writing outside the classroom. While analyzing the conversation log, it was found that about 20% of the students in the experimental group attempted to converse with the chatbot after school even though it was not mandatory. During the text-chat activities, the teachers' intervention was controlled to be at a



minimum with a few exceptions for technical difficulties that occurred.

The same teacher taught the control group, teaching the same expressions based on the same textbook and asked the students to practice them by completing sentences or writing sentences without any context. In other words, both groups received the same English writing lessons with activities such as filling in blanks and completing short sentences in regular English lessons. Therefore, the only distinct difference between the two groups was the use of the chatbot-based writing tool as a treatment, which was given to the experimental group only at the end of each lesson. After 15 weeks (i.e., three and a half months after the pretest), the participants in each condition were given the posttest in the same manner as the pretest but with a different writing topic. Immediately after the posttest was finished, the experimental group's participants completed a short survey that asked about their perceptions of the use of chatbots as a language-learning tool.

## Data Analysis

The submitted writings for the pre- and posttests were scored independently by the first and corresponding authors of this study, and a final score (between 0 and 3) was given to each participant. Both raters were language-testing specialists who have sufficient experience in developing language tests, rating L2 writing, and researching English education at the elementary-school level. If there was a gap of one point or more between the raters, a third experienced rater reviewed the writing and confirmed the final score. An interrater reliability analysis using Cohen's kappa statistic was computed on the initial scores that were given by the two raters. The interrater reliability was found to be  $Kappa = .74$  with  $p < .001$ , which indicates substantial agreement on the scores given by the two raters (Landis & Koch, 1977).

The scores of both pre- and posttests were analyzed to examine whether there was a significant difference in writing performance caused by the treatment—the chatbot writing practice. Both descriptive and inferential statistics were computed to determine whether the differences found between the scores of the two groups were statistically significant. Since the topics of the pre- and posttests were not technically identical and were not counterbalanced, we also conducted an analysis of covariance (ANCOVA) to compare the posttest scores while holding the effect of the pretest score constant. The pretest scores were used as references to establish a baseline writing proficiency between the two respective groups.

The responses to the survey were also coded into IBM's SPSS 26 for quantitative analysis, and descriptive statistics were computed to show an overview of the extent to which participants found the chatbots helpful as a writing practice tool. In particular, perceptions of the use of chatbots as a writing practice tool were compared between the two groups to determine whether the perceptions varied by actual experience using the chatbot.

## Results

### Descriptive Statistics

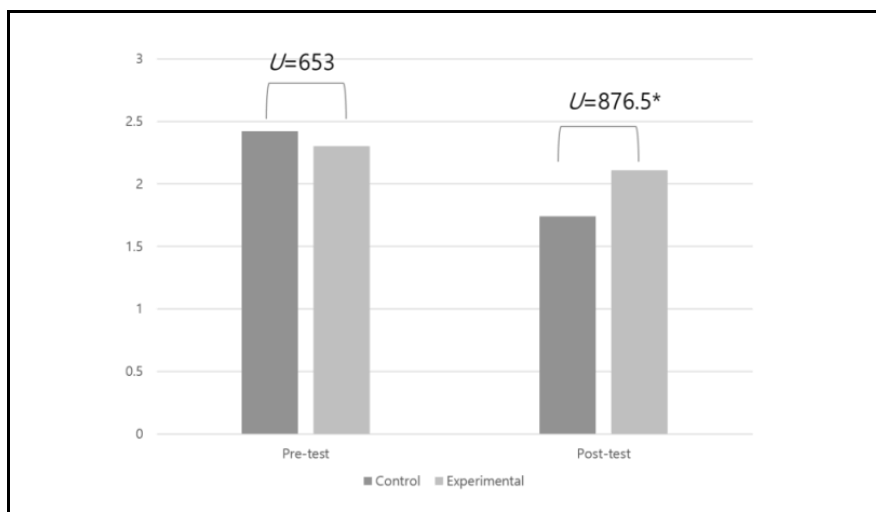
Descriptive statistics of the participants' writing scores are presented in [Table 3](#).

**Table 3***Descriptive Statistics of Test Scores*

Test	Group	<i>n</i>	<i>M</i>	<i>SD</i>	K-S test sig.
Pretest	Control	38	2.42	0.60	<.001
	Experimental	37	2.30	0.74	<.001
Posttest	Control	38	1.74	0.76	<.001
	Experimental	37	2.11	0.61	<.001

Note. *n* = number of participants; *M* = mean scores; *SD* = standard deviation; K-S test sig. = Significance level obtained from Kolmogorov-Smirnov normality test.

The results of the Kolmogorov-Smirnov test showed that all mean test scores had *p*-values lower than .05, which rejects the null hypothesis of normal distribution of data. Thus, nonparametric tests were conducted instead for all pairwise comparisons. The results of the pretest showed that the control group scored higher than the experimental group by 0.12, and the Mann-Whitney *U* test results showed that the difference found between the two groups in the pretest scores was not statistically significant ( $U = 653.00$ ,  $p = .56$ ). The nonsignificant difference found in the pretest scores indicates that there was no significant difference between the two groups in writing proficiency at the initial stage. The results of the posttest show that the experimental group scored 0.37 higher than the control group, and this was statistically significant ( $U = 876.50$ ,  $p < .05$ ). This finding suggests that the students in the experimental group outperformed the control-group students after the treatment of the chatbot writing practice (see Figure 2). It should also be noted that the total mean score for the pretest was higher than the score for the posttest for both groups, which requires careful interpretation, as this might be attributed to an effect of the topic, since the topic given in the pretest was relatively easier than the topic of the posttest.

**Figure 2***Writing Test Score Comparison Between the Two Groups*

Note. *U* = *U* score obtained from Mann-Whitney *U* test.

\* $p < .05$

### Analysis of Covariance (ANCOVA) Results

To examine the effect of the chatbot, an ANCOVA was computed. The ANCOVA was used to precisely measure the main effect of the treatment—the chatbot practice—while holding the pretest score constant. Since the distribution of data was symmetrical, Levene’s test was computed first (see Table 4).

**Table 4**

#### *Levene’s Test Results*

<b>F</b>	<b>df1</b>	<b>df2</b>	<b>Sig.</b>
1.022	1	73	0.315

*Note.* df1 = No. of different groups – 1; df2 = No. of observations – No. of different groups.

According to the Levene’s test results, we found that the null hypothesis could not be rejected because the  $p$ -value was .315, which indicates homogeneity of variance in the two groups. Table 5 presents the results of the ANCOVA.

**Table 5**

#### *ANCOVA Results*

<b>Source</b>	<b>Type I Sum of Squares</b>	<b>df</b>	<b>Mean Square</b>	<b>F</b>	<b>Sig.</b>
Corrected Model	5.628a	2	2.814	6.353	0.003
Intercept	276.480	1	276.480	624.185	0.000
Pretest	2.521	1	2.521	5.692	0.020
Treatment	3.107	1	3.107	7.014	0.010
Error	31.892	72	0.443		
Total	314.000	75			
Corrected Total	37.520	74			

a.  $R^2 = .150$  (adjusted  $R^2 = .126$ )

*Note.* df = degree of freedom; F = F statistic; Sig. = significance level.

The findings show that the effect of the pretest score on the posttest score was significant ( $p < .05$ ). When the effect of the pretest score was held constant, the effect of treatment was also statistically significant ( $p < .01$ ). Hence, the significant difference found in the posttest scores of the control and the experimental groups was still significant while holding the effect of the pretest score constant. Therefore, we can interpret this to mean that the main effect of the treatment, the chatbot practice, had a significant positive effect on test-takers’ writing performance.

**Table 6***Adjusted Mean Posttest Scores*

Category	Mean	SE	95% C.I.	
			Lower	Upper
Control	1.72	0.11	1.503	1.934
Experimental	2.13	0.11	1.909	2.346
Total mean	1.92	0.08	1.770	2.076

Note. SE = standard error; 95% C.I. = confidence interval at the 95% level.

According to Table 6, the estimated adjusted mean post-hoc test scores were 1.72 for the control group and 2.13 for the experimental group. From this result, we can confirm that the students in the experimental group, who used the chatbot to practice writing, showed better writing performance than the control group, which received traditional writing lessons.

**Survey Results**

The first part of the survey conducted in this study was analyzed to collect demographic information on the participants and provide background information about their English-learning experience and writing proficiency. Participants in the control group responded that they studied English at school (92.1%), while a substantial number of them also studied at home (68.4%) or at private cram schools (65.8%). In the experimental group, most of the students reported that they practiced English writing at school or at cram schools (each 70.3%) and at home (59.5%). Only one participant in the experimental group responded that they did not study English at all. However, as aforementioned in the pretest results, the writing proficiency gap between the two groups was not statistically significant, despite the difference found in the amount of English learning experiences.

The participants in the experimental group were asked more questions regarding their perceptions of using the chatbot. The following seven questions in Table 7 were provided to the students to answer using 5-point Likert scales (1 = *strongly disagree*, 5 = *strongly agree*).

**Table 7***Perceptions of Chatbot Writing Practice (Experimental Group)*

Questions	N	Mean	SD
The chatbot was helpful in reviewing what I learned at school.	37	3.54	0.96
I felt comfortable having a conversation with the chatbot.	37	3.19	1.20
I felt nervous when I talked with the chatbot.	37	2.14	1.34
The chatbot was like a native English-speaking friend.	37	2.81	1.22
The chatbot was like a native English-speaking teacher.	37	2.81	1.08
Conversations with the chatbot were more fun than traditional English drills.	36	3.06	1.09
The chatbot helped me improve my English writing.	37	3.08	1.19

Note. Mean = mean of participants' responses to the 5-point Likert scale survey questions (1 = *strongly disagree*, 5 = *strongly agree*); SD = standard deviation.

The results showed that the students generally found the chatbot helpful because they could review and

practice writing again at end of the lesson (3.54). However, they showed somewhat neutral perceptions on the degree of comfortableness with the chatbot (3.19) and their expectation about the chatbot having a facilitating effect on their English writing skills (3.08). Most students did not feel nervous when having conversations with the chatbot (2.1) but at the same time did not strongly agree that the chatbot was like a native English speaker (2.8).

More questions were asked to the experimental group regarding their behaviors while using the chatbot. On the question that asked what they did when the chatbot failed to comprehend their writing, 43.2% of students responded that they stopped the activity after attempting to rephrase the writing, 51.4% of them started over from the beginning after attempting to rephrase the writing, and 24.3% of them quit without attempting to rephrase their writing. The participants of the experimental group also responded about why and when they wanted to quit using the chatbot. Findings showed that 37.8% of the participants wanted to quit when the chatbot could not understand what they said, 24.3% wanted to quit when they had to start a lesson all over again after failing a task, and 27% wanted to quit when they perceived the chatbot as study material.

To answer Research Question 2, the experimental-group participants' intentions to use a chatbot for L2 writing practice in the future were examined. Table 8 presents rationales for chatbot writing based on test-takers' perceptions. The question in the survey was as follows: "If I had a chatbot on which I could practice English writing, I would use it because..."

**Table 8**

*Rationales for Using a Chatbot in the Future*

	I can practice what I learned	I can exchange conversation	I can use English repeatedly	It's not embarrassing even if I say something incorrectly
Frequency	15	11	4	14
Ratio (%)	40.54	29.73	10.81	37.84

Findings show that most students who experienced chatbot writing in this study were willing to use the chatbot again in the future mostly because they could practice what they learned at school (about 40%) while not feeling embarrassed about making errors in their writing when chatting with the chatbot (about 38%). This finding adds more plausible evidence to the aforementioned results on the lower level of nervousness that the participants felt when using the chatbot. About 30% of the participants also mentioned that they would use the chatbot to exchange conversation. Only a small number of participants reported that they would use the chatbot to practice English repeatedly (about 11%).

## Discussion

### Answers to Research Question 1: To What Extent Does Chatbot Writing Practice Have an Impact on L2 Learners' Writing Performance?

The present study examined whether L2 learners who received chatbot writing practice opportunities performed differently compared to another group of L2 learners who received traditional writing lessons from human teachers. After 15-weeks' treatment, we found that the group of learners who used the chatbot for writing practice performed significantly better than the learners who did not. The pretest scores were not significantly different between the two subgroups, which indicates that the improved posttest score in the experimental group is largely attributed to the effect of chatbot learning. To verify the effect of chatbot learning more precisely, we further examined the effect of the chatbot as found in the posttest score while holding the pretest score constant. The results of an ANCOVA confirmed that the positive effect of chatbot

learning was found when the effect of the pretest score was held constant.

The improved writing performance after 15-weeks' chatbot practice can be explained in several ways. First, the chatbot writing provided more hands-on writing practice to individual learners, which offered more effective and focused writing activities to the students. In contrast, the students in the control group received offline English writing activities that practiced the same expressions through traditional top-down writing drills. Although the control-group students also completed read-and-write tasks from the textbook, they had relatively limited opportunities to practice in a more accuracy-focused way because one teacher cannot look after all students as carefully as a chatbot can within a limited time. Second, the learners in the experimental group, presumably the less-proficient ones who are more anxious in using foreign languages, benefitted from chatbot learning particularly in affective domains. More specifically, the use of the chatbot was not face-threatening because they did not need to worry about making mistakes or being judged by other pupils or teachers. Just as the participants in Savin-Baden et al.'s (2015) study disclosed more information to the pedagogical agents than other human beings on sensitive topics, students who practiced writing with the chatbot in the present study also felt more comfortable and confident when completing the given writing tasks. Such a benefit of chatbots from an affective-domain perspective is one of the six main advantages proposed by Fryer and Carpenter (2006).

Although the scores on the pretest were higher than the scores of the posttest in both conditions, this does not mean that the participants' writing performances deteriorated. Careful interpretation is required since the reduced posttest scores might be attributed to an effect of the topic, since the topic given in the pretest (*the most difficult moment of my life*) was relatively easier than the topic of the posttest (*the happiest moment of my life*). According to an existing study, writing about a negative experience is relatively easier than writing about a positive experience because it enhances the reactivation and storage of sensory details and thus leads to a richer subjective quality of the memory as well as to some objectively quantifiable aspects of memory (Bowen et al., 2018).

### **Answers to Research Question 2: To What Extent do L2 Learners of English Perceive Chatbot Writing as a Language-learning Tool?**

To investigate how participants perceived the chatbot as an L2 writing practice tool and how it might be different compared to the classroom practice, participants in the experimental group responded to a list of questions that asked about their perceptions and opinions about the chatbot writing activities. The key questions that were asked to the participants were as follows: (a) whether the chatbot writing practice was helpful in reviewing and practicing what they learned, (b) whether they felt comfortable when practicing writing, and (c) whether they thought their English writing skills had improved. The findings showed that participants in the experimental group who had experienced the chatbot writing generally found the learning method useful and beneficial in improving their writing skills. While a large number of participants responded that the chatbot was helpful in reviewing what they learned, they showed somewhat neutral perceptions on comfortableness and effectiveness of the chatbot. Other responses also showed that participants in the experimental group did not fear making mistakes. Studies have shown that anxiety in L2 learning is strongly affected by fear of negative evaluation, which comes not only from teachers but also from classmates (Horwitz et al., 1986; Shamas, 2006). It can be assumed that practicing L2 writing through a chatbot prevents L2 learners from feeling anxious or stressed about making mistakes or being evaluated by others. Such a permissive learning environment provides a significant benefit to L2 learners, as they could feel more welcomed and reassured when taking risks while practicing writing. Such language skills confidence was suggested as one of the two major benefits of using a chatbot for language learning (Fryer & Carpenter, 2006; Fryer et al., 2020).

The reduced fear and increased confidence are established by forming a beneficial relationship between the learners and the chatbot, as suggested in Frude's (1983) notion of animism. This phenomenon can also be explained through the participants' responses to the other survey questions, which asked whether they felt the chatbot was similar to a real native English-speaking teacher or friend. The findings showed that not many participants felt the chatbot was similar to a real human being but that they still felt comfortable when

using it. This might be because the tasks involved in conversing with the chatbot were fully controlled scenarios instead of free conversation. Different perceptions could have been found if the participants were given free-conversation chatting experiences. Taken together, we can extrapolate that the chatbot-mediated writing provided a positive impact on participants' affective domains because they created an emotional connection with the chatbot while acknowledging that it is not a real human being. This is particularly effective in learning a new language for young learners, like the participants of this study, more than adult learners who would easily lose interest from simple question-and-answer exchanges (Tewari et al., 2013). As Savin-Baden et al. (2015) suggested, an emotional connection between learners and a pedagogical agent (chatbot) is established based on the extent to which they personalize the agent. Over the 15 weeks of chatbot-use opportunities, the learners in this present study might have naturally become familiar with this way of exchanging conversation with the chatbot and personalized it, forming emotional connections. From these findings, it can be argued that the chatbot users in this present study successfully accomplished Frude's (1983) goal of the formation of beneficial relationships between humans and computers. We argue that such beneficial relationships can be established better in areas where human-to-human interaction is relatively uncomfortable, such as in L2 learning for less-proficient learners.

It was also found that the participants who experienced the chatbot-mediated writing activity were willing to use the chatbot for practicing what they had learned in class. They believed that the new method was useful because they could actively exchange conversation without feeling embarrassed. It should be noted that approximately 70% of the entire sample of participants had experience with using chatbots for various other purposes in the past. This proves that AI technology is now prevailing in our lives, particularly among young learners today. Thus, the participants' positive opinion toward the use of chatbots in L2 learning was not a surprise, and many of them might have already been quite familiar with the idea of learning a foreign language through this technology.

### **Limitations**

The limitations of this study are twofold. First, the two writing prompts that were given to the participants were not counterbalanced in each group. For this reason, the topics of the pre- and posttest may have had an influence on writing performance as an intervening variable. In fact, the two groups' posttest scores were lower than their pretest scores, though this does not indicate that their writing skills deteriorated. Instead, we examined the effect of treatment by holding the effect of the pretest score constant using an ANCOVA. Despite this, we acknowledge that different results might have been found if we had counterbalanced the topics at the research design stage. Also, although there was no significant difference in the participants' writing proficiency between the two groups in the pretest results, the experimental-group participants attended English cram schools slightly more than the control group (by five percentage points). However, most English cram schools in Korea, particularly at the elementary-school level, mainly focus on teaching reading and speaking skills and very little writing. Therefore, the larger ratio of the English cram school experience may have not substantially improved the participants' writing skills.

Secondly, as previously mentioned, languages presented in the chatbot writing practices were mostly spoken languages in a conversation style, while the pre- and posttests required written language composition. The target participants, elementary school students in an EFL country, had not sufficiently learned written language in their English classes due to the structure of the national curriculum. Also, it was suggested that learners can learn better when the language is presented in a conversational style (Mayer, 2017). We suggest that future studies replicating this present study should use written-language stimuli for a more precise investigation into the effects of chatbot writing, but the target sample should be those who are already familiar with paragraph writing.

### **Conclusion**

The present study has shown that chatbot-mediated writing practice offers a number of positive benefits to young L2 learners in terms of their writing performance and affective perspectives. Such findings are

important since new generations are becoming more familiar with AI technology, and the use of chatbots for various purposes has become part of our day-to-day lives. However, most commercial chatbots are speaking chatbots that mainly function as personal assistants, while their use in language education has not received much attention yet. Therefore, expanding the use of the AI chatbot in L2 learning, particularly with regard to productive skills like speaking and writing, is strongly recommended. As Fryer et al. (2020) argued, the “golden age” of chatbot learning in the area of L2 learning is “on the horizon” and more active collaborations with researchers in digital multimedia learning is urgently needed (p. 8). The development of future chatbots may focus on social chatbots such as Kuki (formerly Mitsuku), which is able to communicate on any topic, and the language level used by the chatbot becomes adjustable. Thus, it is going to be a tool that supports language learning as a conversation partner without time and space constraints. We strongly believe that chatbots could also be a promising language-learning tool in the near future that can aid learners in practicing all four skills. For instance, having text-chat conversations with a chatbot could not only help learners practice writing, but also could lead them to read the chatbots’ messages, negotiate meaning, and correct mistakes in a natural way. Furthermore, most AI chatbot systems provide error-correction features such as spell check, grammar correction, or translation, which are already being widely used among students today. Studies have argued that these features should be included in future L2 writing activities and assessments (Lee, 2020; Oh, 2020; Tsai, 2019). Follow-up studies are required to look further into what more can be done in an L2 classroom using a chatbot as a pedagogical agent. For this, more chatbot-based language tasks need to be developed and examined empirically to see what specific tasks are effective and how those might vary by different language skills. For example, a chatbot can offer excellent listening-speaking integrated practices by making meaningful conversations with learners. Yet speech-recognition technology is still an obstacle to implement such aural-based language tasks. An excessive amount of data comprising different voices and dialects need to be accumulated over time for better speech-recognition performance. Also, future studies may implement chatbots as an information consultant and provide scenario-based problem-solving tasks to language learners. A similar approach was already made by Abbasi and Kazi (2014), who showed that conversation with a chatbot aided the learners in solving the given problems better than simple search engines.

Recently, the Korean government has been vigorously promoting the development and application of chatbots in the classroom, particularly in English classes for replacing the role of native English-speaking teachers. The Korean government has distributed this AI program to all elementary schools nationwide, thus starting a new chatbot-mediated method of teaching and learning English in early 2021 (Choi, 2021). Considering these rapid changes being made in the field, chatbot-mediated language education will inevitably be a favorable approach in L2 learning in the future, and more studies are called for to examine the ways in which chatbots can facilitate learning more effectively.

## References

- Abbasi, S., & Kazi, H. (2014). Measuring effectiveness of learning chatbot systems on student’s learning outcome and memory retention. *Asian Journal of Applied Science and Engineering*, 3, 57–66.
- Bowen, H. J., Kark, S. M., & Kensinger, E. A. (2018). NEVER forget: Negative emotional valence enhances recapitulation. *Psychonomic Bulletin & Review*, 25(3), 870–891.  
<https://link.springer.com/content/pdf/10.3758/s13423-017-1313-9.pdf>
- Choi, W. (2021). A survey of elementary school students’ recognition on the dialogue-based chatbot - EBSe AI Pengtalk. *Journal of the Korea English Education Society*, 20(4), 211–231.  
<http://dx.doi.org/10.18649/jkees.2021.20.4.211>
- Coniam, D. (2014). The linguistic accuracy of chatbots: usability from an ESL perspective. *Text & Talk*, 34(5), 545–567. <https://doi.org/10.1515/text-2014-0018>



- Dehn, D. M., & van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies*, 52(1), 1–22. <https://doi.org/10.1006/ijhc.1999.0325>
- Frude, N. J. (1983). *The intimate machine*. Century.
- Frude, N., & Jandrić, P. (2015). The intimate machine – 30 years on. *E-Learning and Digital Media*, 12(3–4), 410–424. <https://doi.org/10.1177%2F2042753015571830>
- Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3), 8–14. <http://hdl.handle.net/10125/44068>
- Fryer, L. K., Coniam, D., Carpenter, R., & Lăpuşneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology*, 24(2), 8–22. <http://hdl.handle.net/10125/44719>
- Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, 93, 279–289. <https://doi.org/10.1016/j.chb.2018.12.023>
- Google. (2020). *Dialogflow* [Website]. Google Cloud. <https://dialogflow.com>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125–132. <https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>
- Jia, J. (2009). CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning. *Knowledge-Based Systems*, 22(4), 249–255. <https://doi.org/10.1016/j.knosys.2008.09.001>
- Jia, J., & Chen, W. (2009). The further development of CSIEC project driven by application and evaluation in English education. *British Journal of Educational Technology*, 40(5), 901–918. <https://doi.org/10.1111/j.1467-8535.2008.00881.x>
- Kim, J. S. (2017). *The effects of human–AI assistant interactions on children’s collaborative language acquisition* [Unpublished master’s thesis]. Gwangju National University of Education.
- Kim, H., Shin, D., Yang, H., & Lee, J. (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. <http://doi.org/10.22251/jlcci.2019.19.1.89>
- Kiran, A. H., & Verbeek, P.-P. (2010). Trusting our selves to technology. *Knowledge, Technology & Policy*, 23(3–4), 409–427. <https://doi.org/10.1007/s12130-010-9123-7>
- Kwon, O.-W., Lee, K. S., Kim, Y.-K., & Lee, Y. (2015). GenieTutor: A computer assisted second language learning system based on semantic and grammar correctness evaluations. In F. Helm, L. Bradley, & S. Thouësny (Eds.), *Proceedings of the EUROCALL 2015* (pp. 330–335). Research-publishing.net.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. <https://doi.org/10.2307/2529310>
- Lee, S.-M. (2020). The impact of using machine translation on EFL students’ writing. *Computer Assisted Language Learning*, 33(3), 157–175. <https://doi.org/10.1080/09588221.2018.1553186>
- Lee, J., Kim, J., Na, K., Lee, D., Yin, J., Jeong, E., Kwon, M., Kim, D., Youn, K., Suh, M., Min, K., Choi, E., Kim, S., Jeong, H., Jang, J., Jeong, I., Kim, J., Shin, Y., & Song, J. (2018). *Elementary school English 5*. Daekyo Co.
- Lu, C.-H., Chiou, G.-F., Day, M.-Y., Ong, C.-S., & Hsu, W.-L. (2006). Using instant messaging to provide an intelligent learning environment. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Intelligent tutoring systems: ITS 2006* (pp. 575–583). Springer.

- Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, 33(5), 403–423. <https://doi.org/10.1111/jcal.12197>
- Nass, C., & Yen, C. (2010). *The man who lied to his laptop: What we can learn about ourselves from our machines*. Penguin Group.
- Oh, S. (2020). Second language learners' use of writing resources in writing assessment. *Language Assessment Quarterly*, 17(1), 60–84. <https://doi.org/10.1080/15434303.2019.1674854>
- Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018). *Chatbot as an intelligent personal assistant for mobile language learning* [Conference paper]. 2018 2<sup>nd</sup> International Conference on Education and E-Learning, Bali, Indonesia.
- Savin-Baden, M., Tombs, G., & Bhakta, R. (2015). Beyond robotic wastelands of time: Abandoned pedagogical agents and new pedalled pedagogies. *E-learning and Digital Media*, 12(3–4), 295–314. <https://doi.org/10.1177/2042753015571835>
- Savin-Baden, M., Tombs, G., Burden, D., & Wood, C. (2013). 'It's almost like talking to a person': Student disclosure to pedagogical agents in sensitive settings. *International Journal of Mobile and Blended Learning*, 5(2), 78–93. <https://www.learntechlib.org/p/186175/>
- Shamas, A. (2006). *The use of computerized pronunciation practice in the reduction of foreign language classroom anxiety* [Unpublished doctoral thesis]. Florida State University.
- Shawar, B. A. (2017). Integrating CALL systems with chatbots as conversational partners. *Computación y Sistemas*, 21(4), 615–626. <http://dx.doi.org/10.13053/cys-21-4-2868>
- Shin, D. (2019). *Korean Journal of Teacher Education*, 35(1), 41–45. <http://doi.org/10.14333/KJTE.2019.35.1.41>
- Shin, D., Min, H., & Joo, H. (2012). *The manual of English speaking and writing assessment for preparing NEAT* (ORM 2012-67). Korea Institute for Curriculum and Evaluation.
- Shum, H.-Y., He, X.-D., & 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>
- Sung, M. (2019). Development of a flowchart-based English-speaking chatbot for Korean primary students' negotiation of meaning. *Primary English Education*, 25(4), 101–122. <http://doi.org/10.25231/pee.2019.25.4.101>
- Tewari, A., Brown, T., & Canny, J. (2013). A question-answering agent using speech driven non-linear machinima. In R. Aylett, B. Krenn, C. Pelachaud, & H. Shimodaria (Eds.), *Proceedings of Intelligent Virtual Agents: 13th International Conference, IVA 2013* (pp. 129–138). Springer.
- Tsai, S.-C. (2019). Using Google Translate in EFL drafts: A preliminary investigation. *Computer Assisted Language Learning*, 32(5–6), 510–526. <https://doi.org/10.1080/09588221.2018.1527361>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, 64(7), 456–464. <https://doi.org/10.1177%2F0706743719828977>
- Veletsianos, G., & Russell, G. (2014). Pedagogical agents. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4<sup>th</sup> ed., pp. 759–769). Springer.

Yang, H., Kim, H., Shin, D., & Lee, J. H. (2019). A study on adopting AI-based chatbot in elementary English-speaking classes. *Multimedia-Assisted Language Learning*, 22(4), 184–205.  
<https://doi.org/10.15702/mall.2019.22.4.184>

## Appendix. Scoring Rubric

3	2	1	0
The content of words or sentences is related to the topic, and the use of vocabulary is appropriate. In addition, the use of language is correct and proficient, thus the meaning is clearly communicated without grammatical errors or mechanical errors in English writing rules such as alphabets and punctuation.	The content of words or sentences is usually related to the topic, but the main point is missing. Vocabulary is sometimes used incorrectly, but vocabulary is generally acceptable. In addition, grammatical errors and/or mechanical errors in English writing rules, such as alphabets and punctuation, occasionally occur but do not significantly affect the meaning.	The content of words or sentences is not very relevant to the topic, and the use of vocabulary is not appropriate. In addition, grammatical errors and/or mechanical errors in English writing rules, such as alphabets and punctuation, are frequent, negatively affecting meaning.	Consists of ideas irrelevant to the topic and incomprehensible sentences, making it impossible to convey meaning.

## About the Authors

Suh Keong Kwon (first author) is an assistant professor in the Department of English Education at Chonju National University of Education in the Republic of Korea. He holds a Ph.D. in Education (Language Testing) from University of Bristol. His research interests include language assessment, technology-assisted language learning and eye-tracking methodology.

**E-mail:** [sokekwon@cue.ac.kr](mailto:sokekwon@cue.ac.kr)

Dongkwang Shin (corresponding author) is a professor in the Department of English Education at Gwangju National University of Education in the Republic of Korea. He has received his Ph.D. in Applied Linguistics from Victoria University of Wellington. His expertise and interests are in applied corpus linguistics and AI-assisted language learning. He has published his recent work in *System*, *ELT Journal*, *ReCALL*, and *Language Learning & Technology*, among others.

**E-mail:** [sdhera@gmail.com](mailto:sdhera@gmail.com)

Yongsang Lee (co-author) is an associate professor in the Department of Education at Inha University in the Republic of Korea. He holds a Ph.D. in Education (Measurement and Evaluation) from University of California at Berkeley. His research interests include text mining and machine learning, explanatory item response modeling, and applications of hierarchical linear growth models.

**E-mail:** [yong21c@inha.ac.kr](mailto:yong21c@inha.ac.kr)