

## A Meta-analysis of the Effects of Using AI Chatbot in Korean EFL Education

Je-Young Lee\* · Yohan Hwang\*\*  
(Jeonju University)

Lee, Je-Young & Hwang, Yohan. “A Meta-analysis of the Effects of Using AI Chatbots in Korean EFL Education.” *Studies in English Language & Literature* 48.1 (2022): 213-243. The main purpose of this study is to perform a meta-analysis from a macro perspective, synthesizing the results of experimental research for the use of AI chatbots on educational purposes and practices in EFL environments. From the analysis of the results of 58 individual research cases in 16 experimental trials, EFL classes employing chatbot technology showed a mean effect size of .689, which is near to a large effect size. The following are the key findings in terms of moderators (e.g., school level, publication kind, treatment period, chatbot type, interaction method, and device type). The results show that 1) the lower the school level, the larger the effect size was calculated, 2) treatment that lasted fewer than 8 weeks proved to be more beneficial, and 3) the effect of using a purpose-built chatbot was greater than that of a general-purpose chatbot. On the other hand, the effect sizes calculated for dependent variables regarding linguistic competence and affective domains showed values greater than the medium effect size. The overall outcomes of this study shed light on why AI chatbots have become popular in an EFL setting as a learning and teaching tool to support a set of English capabilities and communication skills. This research has confirmed how AI chatbots may improve the quality, equity, and efficiency of education, as well as where they can optimize their potential and achieve their goals. Based on the findings, this study further suggests valuable insights into future AI chatbot-based research and education. (Jeonju University)

**Key Words:** AI chatbot, AI speaker, meta-analysis, research synthesis, effect size

---

\* First author

\*\* Corresponding author

## I. Introduction

Artificial Intelligence (AI hereafter) is a constellation of state-of-the-art technologies that enable machines, especially computer systems, to comprehend, speak, and act with human-like levels of intelligence. From a wide-ranging branch of computer science, it is now a key driver of growth and innovation across all industries. Educational domains are not different. Although there is inconsistency in prescribing an overdose of AI-based therapy and its efficacy in education, AI technology has changed the conventional learning process all over the world. In South Korea, after the government's AI education blueprint, so-called "National Strategy for Artificial Intelligence" in December 2019 (The Government of the Republic Korea, 2019), the integration of AI technology into the K-12 curriculum and postsecondary education has become one of the most powerful trends in education. After a couple of years later, AI-based education has not fully bloomed yet; however, it has enhanced the quality, equity, and efficiency of education, increasing the possibility of application in many areas.

English education is no exception. English as a Foreign Language (EFL hereafter) educators and researchers have valued AI as a learning and teaching tool to support a set of English capabilities (Hyun & Lim, 2019; H. Kim, Shin, Lee, Y. Kim, Yang, 2019; S. Kim & Lee, 2020; Shin, 2019). Among many service-centered and more teaching-oriented devices, AI speakers and chatbots that undertake tasks thought to require human speech recognition and synthesis are gaining immense popularity in an EFL environment. They serve as a supplementary learning tool for proper and effective inter- and intrapersonal communication. To be specific, an AI speaker helps EFL students comprehend and produce spoken language, which has contributed to reducing their speaking anxiety and improving pronunciation in terms of authenticity and correctness (Choi, 2020; D. Han, 2020; G. Park, 2020). The management and utilization of an AI chatbot system help students improve their (digital) literacy skills by understanding and expressing ideas through two-way communication. From

this perspective, extensive research has shown that the capabilities of AI speaker and chatbot technology improve learning processes and outcomes, making English language learning more engaging and teaching more interactive (Bae, 2020; Choi, 2020; I. Kim & B. Kim, 2020; Min, 2019).

Evidence suggests that AI technology gives wings to English education. As educators begin to explore the ‘Information Superhighway’ up in the air, there should be aerial navigation that guides effective integration and application of this new technology for their safe take-off and landing in an AI-based world.

Recently, research on AI chatbots and speakers, including theoretical inquiry, practical applications, and development, has begun in South Korea. Furthermore, there is much research progress to be made relative to the experimental examination of auditory and text-based chatbots. Such approaches, therefore, need a systematic review and meta-analysis from a more macroscopic point of view by synthesizing the results of experimental studies on the effects of AI chatbots for educational purposes and practices in EFL contexts. In order to address this issue, this study addresses the following two research questions:

- 1) What is the overall effect of English teaching using chatbot?
- 2) What is the effect of English teaching using chatbot by moderator and dependent variables?

## **II. Literature Review**

### **2.1 Foundation and Current Development of AI Chabots**

A chatbot is a computer application designed on the basis of a model of Natural Language Processing (NLP) from information gathering to request routing (Bansal & Khan, 2018). Through auditory or textual data processing, it stimulates

Human-Computer Interaction (HCI). In other words, by using text-based or text-to-speech instant messaging systems that conduct an online conversation, chatbots provide human-like conversation, artificially intelligent meaning-making with one or more human users (Khan, 2017). The history of AI chatbot models can be traced back to the 1960s. The first chatbot, called ELIZA, was developed by MIT professor Joseph Weizenbaum and his team, which simulated simple conversations using pattern matching and substitution techniques based on the information provided by a user (Perez-Marín & Pascual-Nieto, 2011). At this early stage, chatbots were far behind to catch up on human conversation or carry out complex tasks. However, the invention of A.L.I.C.E (Artificial Linguistic Internet Computer Entity) in 1995, recognized as a descendant of ELIZA, provided a steady upgrading process of a chatbot system that performs human-like intelligence functions and catalyzed the next decades of AI chatbot research. As one of the accomplishments, the Watson computer chatbot system, artificially born by an IBM team, competed on a famous quiz show, Jeopardy, against human champions and won the prize in 2011 (Best, 2013).

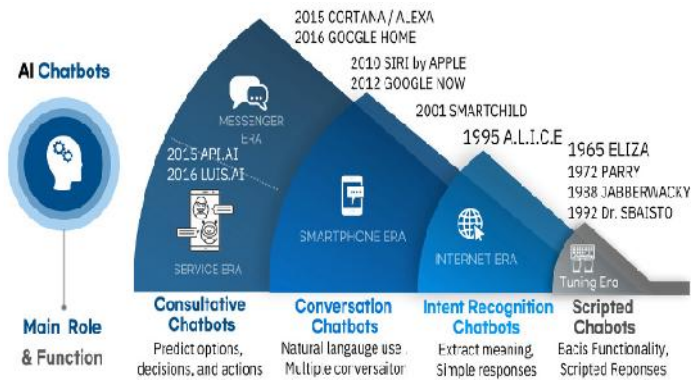


Figure 1. A Brief History of AI Chatbot Development

In the late 2000s and early 2010s, the smartphone had risen in popularity, which was the accelerator that pushed chatbot development and deployment forward in the

social networking world. Early in 2016, the first wave of AI technology in the form of chatbots was established (Dodelin, 2021). Developers, for example, have set up a chatbot for their firm or service on social networking sites like Facebook, assisting customers with some of the routine tasks using their messaging platform (Adamopoulou & Moussiades, 2020; Dodelin, 2021). Many people are thrilled about the advancement of AI technology since the potential for a new method of communicating has been exponentially enhanced. Thanks to the scalability of computer systems and the accessibility of smartphones, chatbots have been used in various public sectors, including e-commerce, entertainment, education, finance, health, news, and productivity (Wollny et al, 2020).

According to Adamopoulou and Moussiades (2020), chatbots can be classified based on service provided (interpersonal, intrapersonal, and inter-agent), knowledge domain (generic, open domain, closed domain), goals (informative, conversational, task-oriented), response generation methods (rule-based, retrieval-based, and generative), human-aid (human-mediated, autonomous) and communication channels (text, voice, and image). From a more educational perspective, Pérez, Daradoumis, and Puig (2020) defined chatbot activities through two applications: “service-oriented chatbots” and “teaching-oriented chatbots.” In light of this, a number of educators and researchers have studied a variety of applications of “teaching-oriented chatbots” in education, which has appeared to benefit learners’ accomplishment, motivation, and pleasure, as well as self-directed learning (Hobert & Wolff, 2019; N. Kim, 2016, 2017; Wollny et al, 2020).

## 2.2 Theory, Creation and Utilization of AI Chatbots in EFL Context

This section presents a literature review investigating the areas of EFL education where “teaching-oriented chatbots” have been adopted. According to Hwang and Lee (2021), AI chatbots in the EFL field can be pointed to three main research domains on a large scale: 1) Theory (theoretical explorations and imperativeness of

AI-chatbots); 2) Creation (design and development of AI chatbots for educational purpose and future potential), and 3) Utilization (experimental research using AI chatbots in education, examining pedagogical implications and possibilities).

### 2.2.1 Theoretical Concerns with AI Chatbots

The literature on an AI chatbot in EFL education has highlighted its future potential and anticipated problems based on the theoretical groundwork of why and how its usage and development have to be taken into consideration. I. Kim and B. Kim (2020) suggest the application of the current state-of-the-art chatbot technology to develop English activities is an inevitable trend in and outside of the classroom. Apart from basic copying of a language conversation or being entertained, chatbots can aid students both linguistically and cognitively in a complex and goal-oriented approach.

As a new AI system that can help perform a certain activity better, it intrigues students' interest and active participation while it increases the level of intrinsic motivation in the language learning process (D. Han, 2020; Hwang & Lee, 2021; Park, 2020). Yoon and Park (2020) highlight that not only is the improvement of the accuracy of speech recognition and synthesis crucial, but detailed search and design for specific learning contents are also prerequisites for successful instruction with AI chatbots in the English classroom. In the same vein, Fryer and Nakoa (2019) claimed why AI chatbots are not yet a substantial instrument for language learning engagement and practice in a Japanese EFL context. They found that even in digital learning environments, teachers were still solely responsible for tailoring course contents to specific student groups and suggested that "the qualitative experience of having 'learned more' with the chatbot was strongly connected to task interest, even when reporting communication difficulties" (p. 279). This also lays the theoretical ground for the importance of the human teacher's role who provides scaffolding (Gabrielli, Rizzi, Carbone, & Donici, 2020), recommending (Xiao, Zhou, & Fu, 2019), and informing (Kerly, Ellis, & Bull, 2008). This is the main reason

why a robot can hardly substitute a human teacher despite the innovative power of AI technologies in the current technological paradigm shift.

In a similar vein, one of the other major theoretical issues that have dominated the AI field concerns as to whether AI chatbots can fully imitate human conversation and go beyond. Again, there is still uncertainty among some researchers if a chatbot could learn various language uses depending on different sociocultural norms and contexts from a pragmatic perspective. However, others forecast that since most chatbot systems can be automatically and/or artificially developed with up-to-date information, and technology feeds itself, a chatbot can be easy and convenient access to socially situated language use and will be evolving itself in terms of a prescriptive approach to language. Overall, some of the theoretical concerns presented here highlight both the potential and limitations of AI chatbots for educational purposes.

### 2.2.2 Creation of Tasks-oriented AI Chatbots

There is a large volume of published studies describing the process and results of creating and developing AI chatbots for educational purposes. Even just a few years ago, the creation and design of AI chatbots were considered a specialty of the computer science or educational engineering fields. However, recently, as API (Application Programming Interface) technology has become easily available, it is possible to use an interface that can control the functions provided by a programming language without professional coding skills (Hwang & Lee, 2021). If a teacher can create a purposeful chatbot system for a certain task in a teaching-oriented fashion, it can help students perform better based on a specific direction and refrain them from getting distracted from too much information and inappropriate language uses that AI chatbots may provide. More importantly, a teacher can supplement the information that existing AI chatbot programs cannot provide to students due to technical limitations.

In light of this, there is an increasing number of studies verifying the educational effects of the creation of AI chatbot programs for educational practice. For example, Chu and Min (2019) examined the possibility of creating an AI chatbot to facilitate English communication. The participants created three interactive tasks adopted from elementary English textbooks such as exchanging small talk, doing information gap activities, and solving problems. The result shows that the prompts for English activities, created by a chatbot API, Google Dialogflow, facilitate “negotiation of meaning and elicitation to proceed interaction among learners” (p. 27). In another study conducted by Min (2019), from the analysis of the 19 pre-service teachers’ developing chatbot experiences with Google Dialogflow, he sheds new light on a collective effort to create and implement AI chatbots for EFL teacher preparation and practice. Similarly, Hwang and Lee (2021) examined the pre-service EFL teachers who created task-oriented chatbots for writing and speaking activities by the same API program, Dialogflow. The results suggest that a task-oriented chatbot built by a teacher is substantially beneficial not only for teachers to harness their potential as a ‘future teacher’ but also for students to achieve language and learning objectives in a more effective way.

Although most of the studies use a chatbot builder API that requires little coding knowledge and skill, several attempts have been made to create an algorithm and a more complex coding-based program for educational chatbots. For example, Chu and Min (2020) used the natural language toolkit (NLTK) and an AI chatbot to create an algorithm that automatically assesses English vocabulary from student conversations. Kim (2020) and her colleagues created their own chatbot called, Ellie, and examined its educational value in terms of a task-based language teaching (TBLT) approach. Specifically, Yang et al. (2019) introduced Ellie into English classrooms in Korean elementary schools and found that students were actively involved in the large number of conversational turns and using various conversational techniques, especially when communication failed with Ellie. The study also exposed the limitations of the current chatbot, such as its inability to make long and fast utterances and its lack of



voice recognition capability. Despite some of the shortcomings, the studies suggested here far provide evidence that building and creating AI chatbots for specific learning tasks can deliver benefits to both teachers and students alike.

### 2.2.3 Experimental Studies on AI Chatbots

A number of experimental studies on AI chatbots on EFL education have reported on their learning potential such as designing activities and courses as well as formative and summative evaluation possibilities. Through an in-depth analysis of existing chatbot programs, they also examine the educational effects of chatbots on a set of four skills of English (i.e., reading, writing, speaking, and listening) (Hyun & Lim, 2019; H. Kim et al., 2019; S. Kim & Lee, 2020; H. Lee, 2019; Shin, 2019; Sung, 2020). In terms of literacy development, when students can use AI chatbots in a well-designed and planned activity, it has a positive influence on improving reading and writing abilities. For example, teachers can encourage students to use an AI chatbot as their virtual assistant to figure out the meaning of vocabulary, search Wikipedia entries, and probe relevant issues in a reading passage (Hwang & Lee, 2021; H. Kim et al, 2019).

Furthermore, chatbots can potentially provide spontaneous responses, formative feedback, and descriptive suggestions for fixing spelling errors and building grammatically correct sentences in a writing prompt. For example, Shin (2019) attempted to explore the feasibility of AI chatbots in EFL students' writing process. In the experiment, 27 university students chatted with the existing chatbot software, Kuki (<https://www.kuki.ai/>) and Cleverbot (<https://www.cleverbot.com/>) while engaged in seven writing tasks (greeting, self-introduction, home, job, hobby, a historical event 10 years ago, and closing). The result shows that it was not difficult for students to complete writing tasks using the expressions of the two chatbots because both of them can give 90% or more of their words among the top 3,000 words during discussions. In the same vein, a study conducted by N. Kim (2018)

reveals that using a voice-chatbot, Alexa by Echo, can help students in writing activities such as exchanging small talk, commanding performance, asking for information as well as problem-solving tasks. N. Kim (2018) also found that students in voice-chatbot immersion significantly improved listening skills as well because they concentrated on the information-gathering process by the auditory mode.

Moreover, there is a growing body of literature that recognizes the importance of chatbots in speaking activities. To be specific, chatbots can provide students an opportunity to practice English speaking without a human partner, which helps them overcome a fear of making mistakes and decrease the level of affective filters (Choi, 2020; Han, 2020; Park, 2020; Park & Go, 2020). If affective filters are low, students' talk will be encouraged, and language acquisition will be facilitated. As mentioned earlier, the information programmed in a chatbot system can be found and provided via the auditory mode and verbal control. Thus, it can help students practice both listening and speaking skills (Hwang & Lee, 2021; H. Kim et al, 2019). During pronunciation and speaking practice, if students provide a wrong answer, a chatbot will provide a fallback intent (e.g., I don't understand, Can you repeat that again?, etc.), which in turn gives students another opportunity to try again for speaking exercise.

According to experimental research conducted by Fryer and Carpenter (2006), out of 211 learners using two automatically answering applications in the classroom, most of the students claimed they felt more confident in chatting with a chatbot although a human teacher or peer is much more intelligent than a programmed robot. Bao (2019) looked at a variety of indicators, including anxiety measures, attitudes toward everyday chatbot usage, oral interviews, and IELTS English-speaking skill tests. The analysis of the results shows that the acceptance of AI chatbots as a non-threatening interlocutor boosted students' confidence and improved specific IELTS skills. Goda and her colleagues (2014) used a chatbot, Eliza, in their experiment, and the findings were cautiously encouraging: a previous interaction with a chatbot increased the quantity of contributions made by students to

dialogues, as well as the number of conversations in which they participated. The researchers further highlight that pre-discussion with a chatbot could lower students' affective filters, boost critical thinking awareness, and develop questioning mindsets. The fact that most EFL students tend to feel more relaxed talking to a chatbot than to a teacher or peers can eventually create a student-machine speaking bond in a comfortable atmosphere (H. Kim et al, 2021), which aids in the dismantling of a psychological barrier that limits input and hence blocks language comprehension.

Across the globe, the number of students learning English with AI technology is dramatically increasing. Literature reviews on AI chatbots in EFL contexts have been undertaken in three main categories focusing on theoretical inquiry and imperativeness, pedagogical implication, and creation of AI chatbots for specific tasks. Taken all together, the evidence reviewed here seems to suggest a pertinent role AI chatbots can play in English language learning and teaching. Such approaches, however, have been less successful to defend teaching-oriented AI chatbots against negative results from research. In addition, although the proportion of experimental research has gradually and steadily increased, the vast majority of studies conducted for research synthesis have been qualitative. Thus, this paper conducts a quantitative meta-analysis of educational AI chatbot uses and their research trends in EFL contexts to establish statistical significance and generate a quantitative estimate of effect magnitude.

### **III. Methods**

#### **3.1 Subjects**

The subjects of this study were searched from various Korean online academic databases, including RISS, NDSL, Kyobo Scholar, DBpia, Kiss. The keywords used for searching were chatbot, chatterbot, AI speaker, artificial intelligences, and so on. The overall stages of collecting the relevant studies are presented in Figure 2 below.

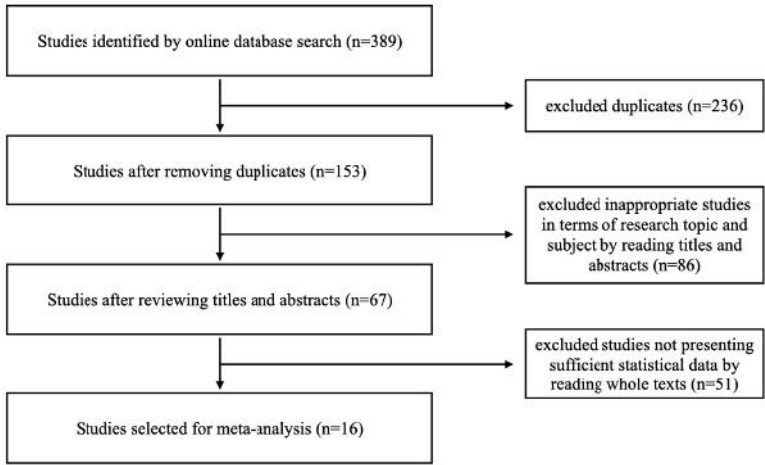


Figure 2. Flowchart of Research Selection

At first, 389 chatbot-related studies were identified and downloaded from the Korean academic databases mentioned above. Then 236 duplicates were removed manually. Next, the studies not related to using the chatbot as a teaching-learning tool and that do not fall into experimental research were excluded by reviewing the titles and abstracts of the research. Finally, the research which does not contain sufficient statistical data for the meta-analysis (e.g., absence of sample sizes, means, standard deviations, or *t/p* values) was ruled out from the analysis. In the process of research selection, two chatbot-related experimental studies (e.g., T. Kim, 2008; Yoon & Han, 2004) were excluded because they used the chatbots developed using the previous generation of AI technology. Seo (2020) was not selected as the subject of this meta-analysis since it applied chatbot-based language teaching to elderly people, not to primary, secondary, or tertiary students. After these screening processes, 58 individual statistical results from 16 chatbot-related experimental studies were selected as the subjects of the current meta-analysis.

### 3.2 Procedures and Instruments

Each study was classified with the following coding schemes (see Table 1) in order to identify the effects of chatbot-based L2 instruction according to the various moderators and dependent variables. The coding results and calculated effect sizes are presented in Appendix 1.

Table 1. Coding Scheme of the Current Meta-Analysis

Category	Variables
1. Publication Types	(1) Journal Article (2) MA/Doctoral Thesis
2. School Level	(1) Primary School (2) Secondary School (3) University
3. Treatment Period	(1) 8 Weeks or Less (2) 9 Weeks or more
4. Chatbot Type	(1) Global (2) Local
5. Proportion of Chatbot Activity	(1) Leading (2) Supplementary
6. Interaction Modes	(1) Spoken (2) Written (3) Spoken + Written
7. Devices for Interaction	(1) AI Speaker (2) Application (3) AI Speaker + Application
8. Dependent Variable I	(1) Linguistic Competence (2) Affective Domain
9. Dependent Variable II	(1) Listening (2) Reading (3) Reading (4) Writing (5) Grammar (6) Vocabulary
10. Dependent Variable III	(1) Interest (2) Confidence (3) Motivation (4) Anxiety (5) Belief (6) Others

After classifying each research case according to the criteria above, the coding results were first summarized in the MS Excel program. Then meta-analysis was implemented with Comprehensive Meta Analysis (CMA) 3.3. When two or more statistical results were presented for the one dependent variable (e.g., writing), multiple effect sizes were integrated into one effect size to minimize the impact of specific study cases on the overall mean effect size.

### 3.3 Analysis

The current meta-analysis targets to synthesize the results of various experimental studies implemented in diverse environments. Therefore, it is necessary to implement the homogeneity test (Q-test) in order to check whether the individual effect size comes from the homogenous population or not (Borenstein, Hedges, Higgins & Rothstein, 2009). If homogeneity was secured, a Fixed Effect Model was used. If not, a Random Effect Model was adopted for the meta-analysis. Finally, Meta-ANOVA was conducted to investigate the effect of moderators.

Each effect size was interpreted according to Cohen (1988)'s standard. He suggested that 0.2 be considered a 'small' effect size, 0.5 means a 'medium' one and 0.8 represents a 'large' one. This means that if the effect size is less than 0.2, the difference between the two groups' means is negligible, even if the calculated effect size is statistically significant.

## IV. Results

### 4.1 Overall Effects of L2 Instruction with Chatbots

In order to investigate the overall effects of English language teaching, the effect sizes of 58 individual studies were calculated, and then the mean effect size was computed. Each effect size that belongs to the same domain (linguistic competence or affective domain) was integrated into one effect size to prevent one study from having too large an impact on the overall effect size. For example, N. Kim (2016) reported three statistical results on speaking skills, and three effect sizes were integrated into one effect size. Through these processes, 29 effect sizes were produced from 58 individual effect sizes. Homogeneity among studies does not exist ( $Q=157.834$ ,  $df=28$ ,  $p=.000$ ). Thus, the Random Effect Model was adopted for the

analysis. The mean effect size of all the chatbot-related research was .689 ( $z=8.149$ ,  $p=.000$ ), which is close to the large effect size.

There has been no meta-analysis for synthesizing the effects of using a chatbot for English language teaching in the Korean EFL context until now. However, Seo, Sohng, and Lee (2020) integrated the effect of computer-mediated communication by meta-analysis and presented a similar effect size ( $g=.69$ ). According to the result of this study, the effect of using a chatbot is similar to the result of Seo, Sohng, and Lee (2020).

Table 2. Overall Result: Mean Effect Size

	<i>k</i>	<i>d</i>	se	95% CI	<i>z</i>	<i>p</i>
fixed	29	.572	.034	.505 ~ .639	16.787	.000
random	29	.689	.085	.523 ~ .854	8.149	.000

## 4.2 Effects of Moderator Variables

### 4.2.1 School Level

The effect of using a chatbot for teaching English language was examined by dividing the subjects of the study into primary school, secondary school, and university. Since the number of research conducted for junior high school students and high school students was relatively low, two school levels were integrated into one category, secondary school students. The results are presented in Table 3.

Table 3. Mean Effect Sizes by School Level

	<i>k</i>	<i>d</i>	se	95% CI	<i>z</i>	<i>p</i>
primary	6	1.030	.283	.475 ~ 1.584	3.637	.000
secondary	5	.708	.154	.406 ~ 1.010	5.647	.000
university	18	.617	.104	.414 ~ .820	5.963	.000

At first, looking at the number of cases, the number of studies on university students was the largest with 18, and the number of research on elementary and secondary school students was relatively small with 6 and 5, respectively. This shows that chatbot-related studies were largely conducted in colleges.

The effect size was higher than the large effect size for elementary school students ( $d=1.030$ ,  $z=3.637$ ,  $p=.000$ ), and the medium effect size or greater for secondary ( $d=.708$ ,  $z=5.647$ ,  $p=.000$ ) and college students ( $d=.617$ ,  $z=5.963$ ,  $p=.000$ ). It was found that the effect size tends to decrease as the school level increases. However, no statistically significant difference was found between the mean effect sizes by school level ( $Q=1.926$ ,  $df=2$ ,  $p=.382$ ).

4.2.2 Publication Type

After dividing the publication type into journal articles and dissertations, the mean effect size of each was calculated. The number of master’s theses and doctoral dissertations was two and one, respectively. Since their number was small, the two cases were unified into one variable, the dissertation. The number of studies was much larger for journal articles ( $k=24$ ) than for dissertations ( $n=5$ ).

Table 4. Mean Effect Sizes by Publication Types

	<i>k</i>	<i>d</i>	se	95% CI	<i>z</i>	<i>p</i>
journal	24	.666	.088	.494 ~ .839	7.579	.000
dissertation	5	.936	.318	.313 ~ 1.559	2.944	.003

The effect size of the dissertation ( $d=.936$ ,  $z=2.944$ ,  $p=.003$ ) was higher than that of the journal article ( $d=.666$ ,  $z=7.579$ ,  $p=.000$ ), and both were close to the medium effect size. The difference between the two effect sizes was not statistically significant ( $Q=.669$ ,  $df=1$ ,  $p=.413$ ).



### 4.2.3 Treatment Period

The analysis of the difference in the effects of using a chatbot by treatment period was conducted, based on the week of treatment. Basically, assuming that the semester runs for 16 weeks, the mean effect size was calculated by dividing eight weeks, which is half of the semester, into two categories: 8 weeks or less and 9 weeks or more. The number of study cases for more than 9 weeks was 22. It is greater than those for less than 8 weeks ( $k=7$ ).

Table 5. Mean Effect Sizes by Treatment Period

	<i>k</i>	<i>d</i>	<i>se</i>	95% CI	<i>z</i>	<i>p</i>
8 weeks or less	7	1.089	.252	.595 ~ 1.582	4.323	.000
9 weeks or more	22	.586	.081	.427 ~ .745	7.227	.000

As a result of examining the effectiveness of using a chatbot based on the week, a large effect size ( $d=1.089$ ,  $z=4.323$ ,  $p=.000$ ) was produced when the treatment was conducted for 8 weeks or less. On the other hand, a medium effect size ( $d=.586$ ,  $z=7.227$ ,  $p=.000$ ) was found for chatbot-based classes for more than 9 weeks. The difference in effect size in the two cases was quite large, about .5, but not statistically significant ( $Q=3.615$ ,  $df=1$ ,  $p=.057$ ).

The results above are inconsistent with the common belief that students' academic achievement is higher as the amount of learning time increases. Some previous research (e.g., J. Lee, 2021; J. Lee, Change & Jeon, 2017) showed that there was no statistically significant difference even when the duration of treatment was increased. On the other hand, the longer treatment showed lower performance in J. Lee's (2017) study. These results showed that too long treatment and intervention, no matter how effective teaching-learning activities or teaching techniques are, can reduce the freshness of the technique and cause students to be bored to learn (Harmer, 2015).

4.2.4 Types of Chatbot

In this part, the types of chatbots were divided into purpose-built chatbots and standard (or non-purposeful) chatbots, and the mean effect size for each type was estimated. According to Lee, Kim, Shin, and Yang (2019), purpose-built chatbots, such as Mitsuku, Replika, Cleverbot, and Elbot, are developed for specific purposes (e.g., task suggestion, English learning in specific situations, etc.), and standard chatbots, usually developed by Dialogflow, play a role as human-like conversation partners. Looking at the number of research cases, the use of purpose-built chatbots for teaching English language (k=4) was much less common than that of standard chatbots (k=25).

Table 6. Mean Effect Sizes by Chatbot Types

	<i>k</i>	<i>d</i>	<i>se</i>	95% CI	<i>z</i>	<i>p</i>
standard	25	.649	.083	.486 ~ .812	7.808	.000
purpose-built	4	1.287	.556	.197 ~ 2.377	2.315	.021

As seen in Table 6 above, the effect size of utilizing a purpose-built chatbot was 1.287 ( $z=2.315$ ,  $p=.021$ ), which was higher than that of using a standard chatbot ( $d=.649$ ,  $z=7.808$ ,  $p=.000$ ). The difference in two types of the chatbot, however, was not statistically different ( $Q=1.288$ ,  $df=1$ ,  $p=.256$ ).

Lee et al. (2019) argued that a purpose-built chatbot was more suitable for students' foreign language learning. Although it is not directly related to the field of chatbots, J. Lee (2012), which meta-analyzed the effect of English education using corpus, reported that the case where the corpus was modified and presented according to the individual differences of the learner (e.g., English proficiency, interest, etc.) had a statistically higher effect than the case that presented without modification. These results showed that language materials tailored to the characteristics of the learner can be of great help in improving students' English proficiency.

#### 4.2.5 Proportion of Chatbot Use in Class

The mean effect size was calculated after classifying the target studies into a case in which chatbots were used as the main tool and a case in which they were used as a supplementary means. If chatbots were used in more than half of the class, it was classified as the main tool. On the other hand, when chatbots were used in less than half of the classes or used for a post-class assignment tool, they were considered supplementary tools. The number of cases in both was relatively similar.

Table 7. Mean Effect Sizes by Proportion of Chatbot Use

	<i>k</i>	<i>d</i>	<i>se</i>	95% CI	<i>z</i>	<i>p</i>
main	12	.823	.143	.543 ~ 1.103	5.764	.000
supplementary	17	.609	.107	.399 ~ .820	5.673	.000

The effect size when chatbot was used as the main tool was .823 ( $z=5.764$ ,  $p=.000$ ), which is a slightly larger result than that of the supplementary tool ( $d=.609$ ,  $z=5.673$ ,  $p=.000$ ). However, this difference is not statistically significant ( $Q=1.434$ ,  $df=1$ ,  $p=.231$ ).

#### 4.2.6 Modes of Instruction

Modes of interaction refer to what medium a learner communicates through when interacting with a chatbot application or AI speaker: spoken language, written language, or both. When using an AI speaker, interaction is usually done through spoken language. On the other hand, in the case of the chatbot application, communication is done by using only the written language or using both the written language and the spoken language at the same time. The number of study cases was in the order of spoken language ( $k=13$ ), written language ( $k=10$ ), and both ( $k=6$ ).

Table 8. Mean Effect Sizes by Interaction Mode

	<i>k</i>	<i>d</i>	<i>se</i>	95% CI	<i>z</i>	<i>p</i>
spoken	13	.824	.123	.583 ~ 1.065	6.691	.000
written	10	.463	.118	.233 ~ .694	3.938	.000
spoken + written	6	.790	.216	.366 ~ 1.213	3.656	.000

The cases of interaction with the spoken language ( $d=.824$ ,  $z=6.691$ ,  $p=.000$ ) and with both spoken and written language ( $d=.790$ ,  $z=3.656$ ,  $p=.000$ ) showed results close to the large effect size. In the case of interaction through written language, on the other hand, a result corresponding to the medium effect size was produced ( $d=.463$ ,  $z=3.938$ ,  $p=.000$ ). Although the difference among modes of interaction was not statistically significant ( $Q=4.921$ ,  $df=2$ ,  $p=.085$ ), the effect size of interaction through spoken language tended to be higher. This result implies that students immersed in voice-communication with a chatbot can focus on the auditory mode of information collecting. N. Kim (2017) also reported that both types of chatbots can contribute to the development of Korean EFL students' speaking ability and the voice-based chatbot was more effective in prompting students' interest and motivation in second language learning. These results are similar to those of this meta-analysis, that is, both modes have a statistically significant level of positive effect, but the effect of the voice-based chatbot is a little larger.

4.2.7 Devices for Interaction

Communication with AI-based chatbots is usually done through applications (e.g., CleverBot or Mitsuku) or AI speakers (e.g., Google Assistant or Amazon Echo). In this study, the effect size of each device was computed by dividing the devices used for interacting with the chatbot into three cases: application, AI speaker, and both application and AI speaker. In the number of cases, application ( $k=17$ ) was the most often used, followed by AI speaker ( $k=9$ ), both ( $k=3$ ).

Table 9. Mean Effect Sizes by Interaction Devices

	<i>k</i>	<i>d</i>	se	95% CI	<i>z</i>	<i>p</i>
AI speaker	9	.686	.146	.399 ~ .972	4.690	.000
application	17	.675	.117	.446 ~ .904	5.779	.000
AI speaker + application	3	.779	.160	.464 ~ 1.093	4.855	.000

The effect size was large in the order of using both AI speaker and application, AI speaker, and applications. The three effect sizes were quite similar, about .7, and no statistical difference was found among them ( $Q=.293$ ,  $df=2$ ,  $p=.864$ ). It indicates that which device is used to implement chatbot-based English language instruction does not have a significant effect on the overall effect.

### 4.3 Effects of Dependent Variables

#### 4.3.1 Linguistic Competence vs. Affective Domain

After classifying the dependent variables of the chatbot-related experimental studies into linguistic proficiency and affective domains, each effect size was calculated. The number of studies that set linguistic competence as a dependent variable ( $k=19$ ) was slightly higher than the number of studies that examined the effect on the affective domains ( $k=10$ ).

Table 10. Linguistic Competence vs. Affective Domains

	<i>k</i>	<i>d</i>	se	95% CI	<i>z</i>	<i>p</i>
linguistic	19	.672	.111	.455 ~ .889	6.076	.000
affective	10	.725	.135	.462 ~ .989	5.390	.000

The effect sizes of the linguistic ability and affective domains were .672 ( $z=6.076$ ,  $p=.000$ ) and .725 ( $z=5.390$ ,  $p=.000$ ), respectively, which were slightly below the large effect sizes. The difference between the two cases was not

statistically significant ( $Q=.095$ ,  $df=1$ ,  $p=.758$ ). Youn and Hwang (2020) reported that flipped learning has the medium effect size ( $g=.486$ ) on English competence and the large effect size ( $g=.735$ ) on affective domains. Comparing two studies, the result in the affective domain is almost similar, but the effect size of this study was larger .2 than Youn and Hwang (2020)’s result in the case of linguistic competence.

4.3.2 Dependent Variables in Linguistic Competence

In this part, the mean effect size was calculated by dividing the dependent variables related to language abilities into more detailed variables such as listening, reading, speaking, writing, grammar, and vocabulary. The number of studies dealing with speaking and writing skills belonging to productive skills was relatively large, and only 1 or 2 studies related to other variables existed.

Table 11. Mean Effect Sizes by Linguistic Variables

	<i>k</i>	<i>d</i>	se	95% CI	<i>z</i>	<i>p</i>
listening	1	.765	.306	.165 ~ 1.364	2.501	.012
speaking	8	.653	.141	.377 ~ .929	4.634	.000
reading	1	.150	.296	-.429 ~ .729	.508	.611
writing	5	.802	.232	.348 ~ 1.256	3.462	.001
grammar	1	.488	.243	.012 ~ .963	2.010	.044
vocabulary	2	1.959	1.683	-1.340 ~ 5.258	1.164	.245

Four (listening, speaking, writing, grammar) of the six calculated effect sizes were statistically significant, while the other two (reading, vocabulary) were not. The four statistically significant values were in the order of writing ( $d=.802$ ,  $z=3.462$ ,  $p=.001$ ), listening ( $d=.765$ ,  $z=2.501$ ,  $p=.012$ ), speaking ( $d=.653$ ,  $z=4.634$ ,  $p=.000$ ), and grammar ( $d=.488$ ,  $z=2.010$ ,  $p=.044$ ). Except for grammar, which had a medium effect size, the remaining three values were close to a large effect size, and there was no statistically significant difference between the effect sizes for each variable ( $Q=4.348$ ,

df=5,  $p=.500$ ). However, some research (e.g., H. Kim et al, 2021; N. Kim, 2016; N. Kim 2018c) reported low effect sizes of 0.2 or less, and some reported negative effect sizes. Therefore, it is necessary to examine which factors positively influence the effectiveness of utilizing chatbots in more detail through follow-up studies.

#### 4.3.3 Dependent Variables in Affective Domain

The mean effect size was calculated by dividing dependent variables related to the affective domain into each sub-variable, including interest, confidence, motivation, and so on. Some variables (e.g., engagement, self-regulation, and attitude) had only one case. Therefore, these variables were combined into one variable, 'others.' The number of cases was diverse, from 2 to 7 as seen in Table 12.

Table 12. Mean Effect Sizes by Affective Variables

	<i>k</i>	<i>d</i>	se	95% CI	<i>z</i>	<i>p</i>
Interest	7	.662	.256	.161 ~ 1.164	2.587	.010
Confidence	3	.815	.210	.404 ~ 1.255	3.886	.000
Motivation	6	.947	.274	.410 ~ 1.484	3.455	.001
Anxiety	3	.861	.286	.299 ~ 1.422	3.004	.003
Belief	4	.590	.135	.325 ~ .855	4.368	.000
WTC	2	.748	.214	.330 ~ 1.167	3.504	.000
Others	4	.249	.266	-.272 ~ .770	.937	.349

The computed effect sizes were in the order of motivation ( $d=.947$ ,  $z=3.455$ ,  $p=.001$ ), anxiety ( $d=.861$ ,  $z=3.004$ ,  $p=.003$ ), confidence ( $d=.815$ ,  $z=3.886$ ,  $p=.000$ ), WTC (willingness to communicate;  $d=.748$ ,  $z=3.504$ ,  $p=.000$ ), interest ( $d=.662$ ,  $z=2.587$ ,  $p=.010$ ), belief ( $d=.590$ ,  $z=4.368$ ,  $p=.000$ ), and others ( $d=.249$ ,  $z=.937$ ,  $p=.349$ ). The effect sizes of all variables except others were statistically significant. The effect sizes of interest, WTC, and belief were above the medium effect size, and that of confidence, motivation, and anxiety were above the large effect size. There was no statistically

significant difference in these effect sizes ( $Q=4930$ ,  $df=6$ ,  $p=.553$ ). In the affective domain, as in the case of linguistic ability, statistically insignificant effect sizes were reported in some studies (e.g., Han, 2020b; H. Kim, Cha, & N. Kim, 2020).

## V. Conclusions

Recently, there has been renewed interest in a pertinent role AI chatbots can play in English education. Despite the fact that the proportion of experimental research has continuously increased, the vast majority of studies have been qualitative. Therefore, this study attempts to collect experimental studies using AI chatbots in English language instruction in the Korean EFL context and synthesized the results included in each study through a meta-analysis. By integrating the results of 58 individual research cases in 16 experimental studies, English language classes using chatbot technology had an effect size of .689, which is close to a large effect. This is very similar to the results ( $g=.69$ ) of Seo, Sohng, and Lee (2020), which meta-analyzed the effects of using Computer Mediated Communication (CMC) in teaching English language.

The main results in terms of moderators (e.g., school level, publication type, treatment period, type of chatbot, proportion of chatbot use, interaction mode, and device type) are as follows. First, the lower the school level, the larger the effect size was calculated. Second, treatment for less than 8 weeks was more effective. Third, the effect of using a purpose-built chatbot was greater than that of a general-purpose chatbot. However, all differences in effect size for each moderator were not at a statistically significant level. On the other hand, the effect sizes calculated for dependent variables regarding linguistic competence and affective domain showed values greater than the medium effect size.

The findings in this study provide a new understanding of why AI chatbots gain enormous popularity in an EFL environment as a learning and teaching tool to support a set of English capabilities and communication skills. In addition, the contribution



of this study has been to confirm that how AI chatbots can enhance the quality, equity, and efficiency of education and where they can maximize the potential and achieve their fulfillment. Notwithstanding the relatively limited sample for meta-analysis, this study offers valuable insights into future AI chatbot-based research and education.

## Works Cited

- Adamopoulou, E., and Moussiades, L. "Chatbots: History, technology, and applications." *Machine Learning with Applications* 2.1 (2020): 1-18. Print.
- Bae, K. "Development of a simple program with a chatbot for class discussion." *The Journal of Modern British & American Language & Literature* 38.1 (2020): 169-190. Print.
- Bansal, H., and Khan, R. "A review paper on human computer interaction." *International Journal of Advanced Research in Computer Science and Software Engineering* 8.4 (2018): 53-56. Print.
- Bao, M. "Can home use of speech-enabled artificial intelligence mitigate foreign language anxiety: Investigation of a concept." *Arab World English Journal* 5 (2019): 28-40. Print.
- Best, J. "IBM Watson: The inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next." *TechRepublic* (2013). Retrieved on September 05, 2021, from <https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. *Introduction to meta-analysis*. West Sussex: John Wiley & Sons, 2009. Print.
- Choi, W. "Process-oriented speaking assessment of primary English using AI chatbots: Possibilities and limitations." *Primary English Education* 26.1 (2020): 131-152. Print.
- Chu, S. Y., and Min, D. G. "A study of using task-based artificial intelligence(AI) chatbot for further interaction in English and the analysis of students' production." *Primary English Education* 25.2 (2019): 27-52. Print.
- Chu, S. Y. and Min, D. G. "A study on the development of an automated algorithm using natural language toolkit (NLTK) and artificial intelligence (AI) chatbot for primary English vocabulary assessment." *Primary English Education* 26.2 (2020): 55-80. Print.
- Dodelin, K. "Putting the human back into the equation: Personalization and AI-driven digital banking experiences." *The Machine* (2021): Retrieved on September 05, 2021, from <https://venturebeat.com/2021/09/20/putting-the-human-back-into-the-equation-personalization-and-ai-driven-digital-banking-experiences/>

- Fryer, L. K., and Carpenter, R. "Bots as language learning tools." *Language Learning & Technology* 10 (2006): 8-14. Print.
- Fryer, K. L., and Nakao, K. "Chatbot learning partners: Connecting learning experiences, interest and competence." *Computers in Human Behavior* 93 (2019): 279-289. Print.
- Gabrielli, S., Rizzi, S., Carbone, S., and Donisi, V. "A chatbot-based coaching intervention for adolescents to promote life skills: Pilot study." *JMIR Human Factors* 7.1 (2020): 1-7. Print.
- Goda, Y., Yamada, M., Matsukawa, H., Hata, K., and Yasunami, S. "Conversation with a Chatbot before an Online EFL Group Discussion and the Effects on Critical Thinking." *Journal of Information and Systems in Education* 13.1 (2014): 1-7. Print.
- Han, D. G. *AI Speaking with Google*. Seoul: Crown, 2020. Print.
- \*Han, D. E. "The effects of AI chatbots on Korean EFL middle school students' speaking competence and affective domains." *Asia-pacific Journal of Convergent Research Interchange* 6.7 (2020a): 71-80. Print.
- \*Han, D. E. *Effects of AI chatbot on Korean EFL learners' speaking ability and affective factors*. Unpublished doctoral dissertation, Chonnam National University, Gwangju, 2020b.
- Harmer, J. *The practice of English language teaching*. 5th ed. Essex, UK: Pearson Education, 2015. Print.
- Hobert, S., and Wolff, M. R. "Say hello to your new automated tutor: A structured literature review on pedagogical conversational agents." *Proceedings of the 14th International Conference*. Eds. V. Pipek and T. Ludwig. Wirtschaftsinformatik, Siegen, Germany, 301-314. Print.
- Hwang, Y. and Lee, H. "Artificial intelligence and its potential in English education: Deployment and creation of an AI chatbot by pre-service EFL teachers." *Multimedia-Assisted Language Learning* 24.1 (2021): 104-133. Print.
- Hyun, J., and Lim, H. "Analysis and implications of AI speakers as English learning tools." *The Mirae Journal of English Language and Literature* 24.1 (2019): 201-219. Print.
- Kerly, A., Ellis, R., and Bull, S. "CALMsystem: A Conversational Agent for Learner Modelling." *Knowledge-Based System* 21.3 (2008): 238-246. Print.
- Khan, R. "Standardized architecture for conversational agents a.k.a. chatbots." *International Journal of Computer Trends and Technology* 50 (2017): 114-121. Print.
- Kim, H. S., Shin, D., Lee, J. H., Kim, Y., and Yang, H. *Application and creation of an AI chatbot for English learning*. Gyeonggi-do: Kyoyookbook, 2019. Print.
- \*Kim, H. S., Cha, Y., and Kim, N. Y. "Impact of mobile interactions with AI on writing performance." *Modern English Education* 21.2 (2020): 1-13. Print.
- \*Kim, H. S., Kim, N. Y., and Cha, Y. "Is it beneficial to use AI chatbots to improve learners' speaking performance?" *The Journal of Asia TEFL* 18.1 (2021): 161-178. Print.

- Kim, I. S., and Kim, B. "The current state of interactive AI chatbots and the development of chatbots in English language education." *English21* 33.2 (2020): 1-20. Print.
- \*Kim, J. S. *The effects of human-AI assistant interactions on children's collaborative language acquisition*. Unpublished MA thesis, Gwangju National University of Education, Gwangju, 2017.
- \*Kim, N. Y. "Effects of voice chat on EFL learners' speaking ability according to proficiency levels." *Multimedia-Assisted Language Learning* 19.4 (2016): 63-88. Print.
- \*Kim, N. Y. "Effects of different types of chatbots on EFL learners' speaking competence and learner perception." *Cross-Cultural Studies* 48 (2017): 223-252. Print.
- \*Kim, N. Y. "A study on chatbots for developing Korean college students' English listening and reading skills." *Journal of Digital Convergence*, 16.8 (2018a): 19-26. Print.
- \*Kim, N. Y. "Chatbots and Korean EFL students' English vocabulary learning." *Journal of Digital Convergence* 16.2 (2018b): 1-7. Print.
- \*Kim, N. Y. "Effect of text chat on EFL writing fluency, accuracy, and complexity by interlocutors." *Foreign Languages Education* 25.1 (2018c): 27-54. Print.
- \*Kim, N. Y. "Different chat modes of a chatbot and EFL students' writing skills development." *Studies in Foreign Language Education* 32.1 (2018d): 263-290. Print.
- \*Kim, N. Y. "A study on the use of artificial intelligence chatbots for improving English grammar skills." *Journal of Digital Convergence* 17.8 (2019): 37-46. Print.
- Kim, S., and Lee, S. "An analysis of a primary school English speaking lesson using an AI-powered conversational English learning application: A focus on student engagement as well as learners' and teachers' perceptions of the class." *Primary English Education* 26.3 (2020): 177-202. Print.
- \*Kim, S. Y., and Kim, J. R. "Affective effects of English digital textbook lessons using AI chatbot." *Journal of Learner-Centered Curriculum and Instruction* 21.10 (2021): 37-49. Print.
- Kim, T. W. "The effects of computer interest levels and chatting method (with AI chatting robot: Chatterbot) on teaching and learning." *Journal of Engineering Education Research* 11.3 (2008): 19-33. Print.
- Lee, D. H. "A developmental plan for an English conversation learning system through the application of a talking robot based on artificial intelligence." *STEM Journal* 20.1 (2019): 189-211. Print.
- Lee, J. H., Kim, H., Shin, D. K., and Yang, H. "A study on the conversational interaction with chatbot for foreign language learning." *Multimedia-Assisted Language Learning* 22.1 (2019): 132-153. Print.
- Lee, J. Y. "Trends of Korean corpus-based CALL research: A meta-analysis." *Multimedia-Assisted Language Learning* 15.3 (2012): 83-111. Print.

- Lee, J. Y. "Effects of using subtitles of video contents on L2 learners' listening and vocabulary development: A meta-analysis." *Korean Journal of Applied Linguistics* 33.2 (2017): 137-158. Print.
- Lee, J. Y. "Effects of utilizing storytelling on English competence: A meta-analysis." *Journal of Language Sciences* 28.1 (2021): 137-161. Print.
- Lee, J. Y., Chang, K., and Jeon, Y. "Effects of English language instruction for underachievers in the secondary school: A meta-analysis." *Studies in English Education* 22.4 (2017): 251-278. Print.
- \*Lee, S. "The effects of gamification-based artificial intelligence chatbot activities on elementary English learners' speaking performance and affective domains." *Primary English Education* 25.3 (2019): 75-98. Print.
- Min, D. G. "A study on the development of artificial intelligence(AI) chatbot by pre-service student teachers." *Primary English Education* 25.4 (2019): 169-190. Print.
- Park, G. *55 tips for English classes by artificial intelligence*. Seoul: Chunjae Education, 2020. Print.
- Park, G., and Ko, K. *Okay Google, Please Take Care of My English*. Seoul: Saramin, 2020. Print.
- \*Park, M. Ho., and Kim, J. R. "An analysis of high school students' task interaction using AI chatbots." *Journal of Learner-Centered Curriculum and Instruction* 21.9 (2021): 1-13. Print.
- Pérez, J. Q., Daradoumis, T., and Puig, J. M. "Rediscovering the use of chatbots in education: A Systematic Literature Review." *Computer Applications in Engineering Education* 28.1 (2020): 1549-1565. Print.
- Perez-Marin, D., and Pascual-Nieto, I. *Conversational agents and natural language interaction: Techniques and effective practices*. Hershey, PA: Information Science Reference, 2011. Print.
- \*Seo, J. I. *The effects of AI chatbot on the acquisition of English vocabulary and affective areas of elementary school students*. Unpublished MA thesis, Cyber Hankuk University of Foreign Studies, Seoul, 2021.
- Seo, S. Y., Sohng, H. S., & Lee, H. J. "Efficacy of CMC-based English learning Environments: Meta-analysis." *Korean Journal of Teacher Education* 36.4 (2020): 373-400. Print.
- Seo, S. J. *Learning English vocabulary with a smart speaker: Opportunities and challenges for the elderly*. Unpublished MA thesis, Hallym University, Gangwon, 2020.
- Shin, D. K. "Exploring the feasibility of AI chatbots as a tool for improving learners' writing competence of English." *Korean Journal of Teacher Education* 35.1 (2020): 41-55. Print.
- Sung, M. C. "Pre-service primary English Teacher's AI chatbots." *Language Research* 56.1 (2020): 97-115. Print.

- The Government of the Republic Korea. (2019). *National Strategy for Artificial Intelligence*. Retrieved from September 05, 2021, from <https://english.msit.go.kr/SYNAP/skin/doc.html?fn=14acc067ebaf2780a558e24993a560f0&rs=/SYNAP/sn3hcv/result/>
- Yang, H., and Kim, H. "Development and application of AI chatbot for cabin crews." *Korean Journal of English Language and Linguistics* 21 (2021): 1085-1104. Print.
- Yoon, S. K., and Han, J. I. "A case study of the effects of the use of chatter bot and messenger on the productive skills of Korean EFL learners." *Multimedia-Assisted Language Learning* 7.2 (2004): 269-292. Print.
- Yoon, Y., and Park, M. "Artificial intelligence and primary English education: with special reference to chatbots." *Journal of Korea Elementary Education* 31 (2020): 77-90. Print.
- Youn, J. M., & Hwang, M. "A meta-analysis of the effects of flipped learning on the learners' English proficiency and affective traits." *English Language Teaching* 32.2 (2020): 127-148. Print.
- Wollny, S., Schneider, J., Mitri, D., Weidlich, J., Rittberg, M., & Drashsler, H. "Are we there yet? A systematic literature review on chatbots in Education." *Frontiers in Artificial Intelligence*, 4.654924 (2020): doi: 10.3389/frai.2021.654924
- Xiao, Z., Zhou, M. X., and Fu, W. T. (2019). "Who should be my teammates: Using a conversational agent to understand individuals and help teaming." Paper presented at the 24th International Conference on Intelligent User Interfaces, Marina del Ray, California, USA, 2019. (unpublished proceedings)

Je-Young Lee (Jeonju University / Professor)

Address: (55069) Dept. of English Education, Jeonju University, 303 Cheonjam-ro, Wansan-gu, Jeonju-si, Jeollabuk-do

Email: jylee@jj.ac.kr

Yohan Hwang (Jeonju University / Professor)

Address: (55069) Dept. of British and American Language and Culture, Jeonju University, 303 Cheonjam-ro, Wansan-gu, Jeonju-si, Jeollabuk-do

Email: yvh5101@jj.ac.kr

Received: December 31, 2021 / Revised: February 7, 2022 / Accepted: February 9, 2022



Kim, N. (2018d)	Article	University	10 Weeks	NP	SU	S	Application	Linguistic	Writing	2.497
						W	Application	Linguistic	Writing	.892
									Writing	1.038
Kim, N. (2019)	Article	University	16 Weeks	NP	SU	W	Application	Linguistic	Grammar	.488
Kim, S. & Kim, J. (2021)	Article	Primary	8 Weeks	NP	M	S+W	AI Speaker + Application	Affective	Confidence	.943
									WTC	.644
Lee (2019)	Article	Primary	12 Weeks	P	M	S+W	AI Speaker + Application	Linguistic	Speaking	.504
								Affective	Affective Domain	1.042
Park & Kim, J. (2021)	Article	Secondary	8 Weeks	NP	M	W	Application	Linguistic	Writing	1.123
Seo (2021)	MA	Primary	4 Weeks	P	M	S	AI Speaker	Linguistic	Vocabulary	3.683

Note. Pub: Publication, TP: Treatment Period, CT: Chatbot Type, PoC: Proportion of Chatbot,

P: Purpose-Built, NP: Non-Purpose, M: Main, SU: Supplementary, S: Spoken, W: Written