Education and Information Technologies
https://doi.org/10.1007/s10639-023-11834-1

Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT

Jaeho Jeon[1] · **Seongyong Lee**[2]

Received: 21 February 2023 / Accepted: 17 April 2023

Abstract
Artificial Intelligence (AI) is developing in a manner that blurs the boundaries between specific areas of application and expands its capability to be used in a wide range of applications. The public release of ChatGPT, a generative AI chatbot powered by a large language model (LLM), represents a significant step forward in this direction. Accordingly, professionals predict that this technology will affect education, including the role of teachers. However, despite some assumptions regarding its influence on education, how teachers may actually use the technology and the nature of its relationship with teachers remain under-investigated. Thus, in this study, the relationship between ChatGPT and teachers was explored with a particular focus on identifying the complementary roles of each in education. Eleven language teachers were asked to use ChatGPT for their instruction during a period of two weeks. They then participated in individual interviews regarding their experiences and provided interaction logs produced during their use of the technology. Through qualitative analysis of the data, four ChatGPT roles (interlocutor, content provider, teaching assistant, and evaluator) and three teacher roles (orchestrating different resources with quality pedagogical decisions, making students active investigators, and raising AI ethical awareness) were identified. Based on the findings, an in-depth discussion of teacher-AI collaboration is presented, highlighting the importance of teachers' pedagogical expertise when using AI tools. Implications regarding the future use of LLM-powered chatbots in education are also provided.

Keywords ChatGPT · Large language model · Chatbot · AIEd · Human–computer interaction · Artificial intelligence · Large language model-powered chatbot

✉ Seongyong Lee
seongyonglee77@gmail.com

Extended author information available on the last page of the article

Published online: 02 May 2023

1 Introduction

The use of chatbots to enhance students' language learning experience has gained much attention in recent years (Huang et al., 2023; Hwang et al., 2020; Jeon, 2021; Jeon, 2022a; Lee & Jeon, 2022). Language researchers have employed different types of chatbots, including those produced commercially and those developed in research labs, and have identified novel educational opportunities provided by chatbots (Dizon, 2020; Fryer et al., 2020; Huang et al., 2022). Although research has consistently shown that chatbots can facilitate language learning, significant limitations inherent in current chatbot systems have also been revealed (Bibauw et al., 2019; Jeon, 2022b). These include chatbots' limited ability to engage in open-ended conversations with learners and to sustain an extended and goal-oriented conversation on a specific topic (Kuhail et al., 2022).

The recent public release of ChatGPT has marked an important advancement in the field of artificial intelligence (AI), specifically in natural language processing (NLP) (Brown et al., 2020). ChatGPT, a generative AI chatbot powered by a large language model (LLM) that has been trained on vast amounts of internet text data, is projected to overcome many limitations of previous chatbot technology and ultimately affect the way people learn (Heidt, 2023; Kasneci et al, 2023). In contrast to the majority of existing chatbots, which adhere to predefined dialogue paths or feature simplistic question-and-answer dialogue structures (for a review on existing educational chatbots, see Kuhail et al., 2022), this exponentially advanced chatbot can generate answers according to the context of a given prompt, and thus, can engage in a conversation that is more akin to human–human interaction than those of previous chatbots (Susnjak, 2022).

The popularity of ChatGPT, as evidenced by the rapidly growing number of registered users (Altman, 2022), suggests that this type of technology is poised to become increasingly integrated into society. Accordingly, educators anticipate a significant shift in various aspects of education, including the role of teachers (Stokel-Walker, 2023). As it proves able to provide novel resources that enrich learning experiences beyond conventional methods and ones provided by current chatbot technology (Jeon, 2022b), it may take over some roles traditionally carried out by teachers. For example, as suggested in initial theoretical explorations on the technology (Kasneci et al, 2023; Zhai, 2022), teachers may utilize the features of this chatbot to stimulate students' engagement with textbook content. Also, teachers may assign the task of initially grading student essays to the technology and focus their efforts on providing more detailed feedback. Conversely, some scholars, such as Kasneci et al. (2023), expressed a concern that some teachers may become excessively reliant on ChatGPT and fail to ensure opportunities to foster students' creativity, critical thinking, and problem-solving skills. However, despite such predictions on the impact of ChatGPT, its potential as a learning tool, how teachers utilize it, and the nature of the relationship between teachers and this technology have yet to be empirically explored.

In response to this research need, we conducted a qualitative investigation on the educational potential of ChatGPT in the context of language education.

Specifically, the perspectives and experiences of language teachers who deployed ChatGPT in their instruction were examined. The findings of this inquiry may contribute to our understanding of the pedagogical implementation of LLM-powered chatbots, the future development of the technology, and the professional development needs of teachers in the utilization of emergent forms of AI tools in education.

2 Literature background

2.1 Chatbots for language learning: From ELIZA to ChatGPT

Since the development of ELIZA, the initial version of NLP-powered chatbots (Weizenbaum, 1966), chatbots have undergone several technological advances, thereby increasing their potential for pedagogical use as a tool for supporting students' language learning. Jeon (2022b) identified three types of chatbots that have been employed in the context of language learning: (1) general-purpose chatbots that can engage in daily, simple conversation with users via question-and-answer dialogue structures (e.g., Dizon, 2020; Fryer et al., 2019), (2) specific-purpose chatbots with more complex structures designed by commercial developers for language learning (e.g., Wang et al., 2023), and (3) customized chatbots created by researchers or teachers for particular language learning contexts using visual chatbot development platforms (e.g., Hew et al., 2023; Lee & Jeon, 2022).

Recent research on these types of chatbots has consistently demonstrated the efficacy of chatbots in language learning, as evidenced in Bibauw et al.'s (2022) meta-analysis of previous chatbot studies. Meanwhile, several limitations in chatbot research have also been identified. First, little is known about the possible roles of chatbots in language education beyond that of a conversation practice partner that follows a limited number of pre-determined dialogue paths (Huang et al., 2022; Ji et al., 2023). Despite the potential for chatbots to assume a variety of more complex roles (Kasneci et al, 2023), such as facilitating the organization of students' thoughts, providing lesson plan information to teachers, and assisting with the preparation of instructional materials, these possibilities have not been developed or remain either at a theoretical or a prototype level (Bibauw et al., 2019). Second, chatbots previously examined were unable to participate in open-ended conversation comprising several turns with users, which restricted meaningful exchanges between chatbots and users to only a few turns (e.g., Jeon, 2021; Jeon, 2022b). As indicated by Bibauw et al. (2019) and Kuhail et al. (2022), the limitations of current chatbots can be attributed, at least in part, to the rule-based design structure for conversation and the limited sources of data to be fed into chatbot systems that were available to developers at the time of chatbot creation. These limitations resulted in the chatbots' inability to respond appropriately to a wide variety of user utterances.

The recent release of ChatGPT, a highly advanced version of a generative AI chatbot powered by an LLM, has successfully addressed some of the limitations of existing chatbots. An LLM, a type of machine learning model that is trained with a vast dataset of human language, significantly upgraded the overall performance of

NLP applications (Brown et al., 2020). Thus, it enables chatbots to generate a wider range of human-like responses than previously possible. Specifically, being built on LLM technology provides ChatGPT with the following three capabilities (OpenAI, 2023): (1) the ability to remember previous statements made by the user during the conversation; (2) the ability to comprehend follow-up corrections by the user; (3) the ability to decline inappropriate requests. These features enable ChatGPT to simulate a more human-like conversation and respond to a wide range of utterances if deemed appropriate. Moreover, the chatbot can maintain potentially limitless conversational turns in a goal-oriented conversation on a specific topic. These capabilities, the result of intensive development in NLP, suggest that LLM-powered chatbots may be adapted to a wider range of educational roles than previous forms of chatbots (Kasneci et al, 2023). In sum, the pedagogical use of LLM-powered chatbots has now become a viable option for implementation in education, including language learning.

2.2 Complementary relationships between humans and AI in education

Along with the recent proliferation of AI into classroom contexts, scholars have paid much attention to human-AI collaboration, positing that more effective learning can be jointly created by human facilitators and AI than by humans or AI working alone (Holstein et al., 2020; Kim et al., 2022; Xu & Ouyang, 2022). In this line of inquiry, teachers' agentive roles are emphasized and described as facilitating the positive impact of AI on education, rather than AI being considered as a potential substitute for teachers (Bower, 2019; Jeon et al., 2022). For example, using frameworks for adaptivity in education, Holstein et al. (2020) suggested four categories as areas in which human teachers may enhance the adaptability of AI technology, including goal augmentation, perceptual augmentation, action augmentation, and decision augmentation. In this theoretical framework for teacher-AI collaboration, they underscored the extensive roles of teachers in maximizing the effectiveness of AI. Of more relevance to the current research, Ji et al. (2023) reviewed 24 research studies on chatbots for language learning, focusing on the collaboration between teachers and chatbots. They found that only a limited amount of empirical evidence existed regarding teacher-chatbot collaboration. Based on the idea of classroom orchestration (Dillenbourg & Jermann, 2010), they suggested that researchers examine the potential that teacher-chatbot collaboration may offer, calling for more empirical research on how the collaboration can be conducted, how it facilitates learning, and how AI can reduce teachers' workloads.

In addition to these theoretical explorations, some, albeit limited, empirical research demonstrated how human teachers can actually complement AI (e.g., Holstein et al, 2019). For example, in Holstein and Aleven (2022), teachers utilized smart glasses that provided real-time data on student learning during class. They observed, "teachers then made a rich inference about the latent, underlying cause of the behavior and responded with support and flexibility that the AI tool could not provide" (p. 43). This observation shows that it was not the AI tool itself that facilitated learning, but rather the teachers' pedagogical expertise that was strengthened by the AI-generated data. On a similar note,

Zhang et al. (2022) offered another illustration on how collaboration between humans and AI might promote student learning. In a book-reading scenario involving an AI-powered storytelling app, parents, acting as classroom teachers, were provided with a series of individualized story questions via the app. Referring to those questions, parents could flexibly choose the most pertinent question at any given moment while ignoring irrelevant questions; this agentive action of parents promoted a rich conversation between parents and children.

These cases highlight the importance of human teachers' pedagogical decisions when using AI in education with empirical evidence, enriching our understanding of teacher-AI collaboration in education. However, despite their contributions, there remains an important gap in the literature, particularly concerning the features of AI tools used. Previous research has lacked relevant pedagogical insights regarding the rapidly changing landscape of AI technology, which blurs the boundaries between specialized areas of application. (Brown et al., 2020; Kasneci et al., 2023). That is, previous studies have primarily focused on tools that targeted specific aims during a particular instructional stage, as shown above in the cases of smart glasses and a storytelling app (Holstein & Aleven, 2022; Zhang et al., 2022). This emphasis on AI tools tailored to target specific educational aims has also been found in several review studies conducted in the field of AI in education (AIEd) (e.g., Celik et al., 2022; Hwang et al., 2020). For example, in a comprehensive review of 146 studies, Zawacki-Richter et al. (2019) found 150 AI application cases, which they categorized under four main application areas (i.e., profiling and prediction, intelligent tutoring systems, assessment and evaluation, and adaptive systems and personalization). This suggests that in most of the studies, investigations of AI tools were confined to specific roles within these four application areas. The tendency noted above is also salient in the literature that examined the role of chatbots in education (e.g., Ji et al., 2023). For example, Kuhail et al.'s (2022) review of studies on educational chatbots revealed that 37 chatbot application cases found in 36 studies were described as performing one or more of four distinct roles (i.e., teaching agent, peer agent, teachable agent, and motivational agent), again suggesting the specialized purposes of AI tools. That is, except for one study, chatbots in the reviewed studies were designed to assume only one of these four roles.

Consequently, scholarly examination of the rapidly changing field of AI technology may have been less exhaustive than necessary. As illustrated above, when teachers utilize AI tools that are specifically designed to achieve particular educational goals, their responsibilities may be limited to supplementing the tools' capabilities (Holstein & Aleven, 2022; Zhang et al., 2022); this may not necessarily reflect the nature of the emerging relationship between teachers and ChatGPT, which has the potential to be used in a variety of application areas and contexts (Kasneci et al., 2023; Zhai, 2022). As demonstrated by ChatGPT, AI is expected to undergo a paradigm shift in which it will be utilized with increasing versatility. This evolution may render the teacher-AI relationship in education ever more dynamic and difficult to predict than before (Kasneci et al., 2023; Zhai, 2022). However, how evolving AI, such as ChatGPT, can be integrated into education and what roles teachers will play as a result remains underexplored.

Therefore, this exploratory study aimed to investigate language teachers' uses and perceptions of ChatGPT to examine the pedagogical potential of this recently released LLM-powered chatbot. Specifically, this study attempted to identify the roles of ChatGPT in language education and how it may influence teachers' roles. The following research questions guided the study:

1. What roles can an LLM-powered chatbot play in language education?
2. What roles do teachers perceive as becoming more important with the use of LLM-powered chatbots?

3 Methodology

We adopted an exploratory qualitative approach to gain insight into the pedagogical value of an LLM-powered chatbot and how teachers perceived their roles would change with the use of the technology (Creswell, 2008). To achieve this goal, we collected data from two sources: individual semi-structured interviews as a primary source and teachers' chatbot-use logs as an objective supplement to the interview data.

3.1 Participants and chatbot

Eleven English language teachers from ten elementary schools in South Korea were recruited through ads posted on a social online forum for elementary school teachers (Table 1). We selected participants representing a variety of ages, genders, and teaching experiences from among those who contacted us and expressed their willingness to apply ChatGPT in their teaching. Prior to the study, none of the participating teachers had used ChatGPT for any purpose, and only one had heard about it through a news platform.

Table 1 Participant profiles

Teacher	Age	Gender	Years of teaching experience
1	35	Female	8
2	31	Female	5
3	32	Male	5
4	35	Female	7
5	30	Female	4
6	27	Female	2
7	34	Female	8
8	34	Female	9
9	38	Female	11
10	36	Male	9
11	32	Female	5

The latest version of an LLM-powered chatbot for text generation, ChatGPT, also called GPT-3.5 in the family of LLMs, was released by OpenAI for public use on November 30, 2022. It is built on OpenAI's generative pre-trained transformer-3.5 and is being continually fine-tuned with both supervised and reinforcement learning techniques (OpenAI, 2023).

3.2 Procedures

Figure 1 illustrates the procedures followed in this research. The recruited teachers adhered to the procedures individually, with their utilization of ChatGPT taking place throughout the months of January to February 2023. The teachers first participated in a 60-min seminar provided by one of the researchers, who first informed the participants of the purpose and procedures of the research, including the handling of their interview responses and interaction log data in future reports and publications. They were aware of their rights to withdraw at any stage and provided their informed consent. For about 20 min, the participants were introduced to ChatGPT and learned about the basic functions of the chatbot, including how to create an account and access the chatbot and dialogue histories. They were also shown how the chatbot responded to different types of questions. The rest of the session, which lasted for around 40 min, involved a Q&A discussion in which the presiding researcher and a teacher fielded questions from the participants, such as whether certain educational prompts would be recognized or how they could craft questions more accurately to obtain specific information. To respond to their questions, the researcher and teacher collaborated to seek answers by inputting specific prompts or questions and sharing their opinions about the outcomes. For the two weeks following the seminar, the participant teachers used ChatGPT however they chose in their own teaching contexts. It was emphasized that there were no constraints on the ways of using ChatGPT as long as it was appropriately used for facilitating teaching and learning. They were also asked to collect up to 10 noteworthy or debatable methods of use in the form of interaction logs with the chatbot. After the two weeks, the teachers participated in individual semi-structured interviews with one of the researchers in which they shared their experiences with and perceptions of the use of ChatGPT. The interviews lasted for an average of 60 min and were audio-recorded and transcribed, with parts translated into English.

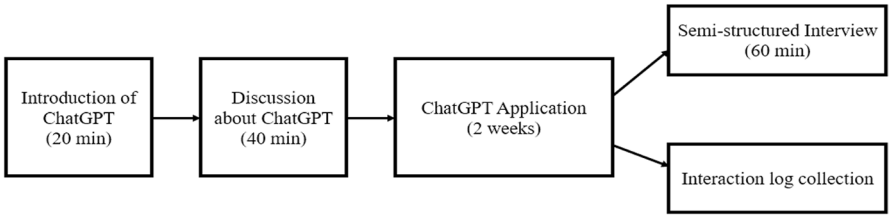

Fig. 1 Research procedures

3.3 Data collection and analysis

Data comprised the transcripts of the semi-structured in-depth interviews and the teachers' interaction logs with the chatbot. To guide the interview process, semi-structured protocols were developed based on the interview guide used by Timpe-Laughlin et al. (2022) in a study of teachers' experiences with and perspectives on the use of a chatbot system. The interviews were structured around the following questions:

- Can you tell me how you used the chatbot?
- Did you find using chatbots helpful when teaching English? If so, what functions of the chatbot were the most useful?
- What parts of your pedagogical work do you think can become easier/more important with the use of the chatbot?
- Do you think teachers are likely to be replaced by this type of technology? Why or why not?

Given the exploratory aim of the study, initial questions were designed to elicit teachers' general perceptions of and experiences with ChatGPT. Subsequent follow-up questions focused on the roles assumed by the teachers while using ChatGPT.

The transcribed interview data were subsequently subjected to qualitative analysis in order to identify recurring patterns of meaning that were pertinent to the research questions. In accordance with the methodology proposed by Braun and Clarke (2006), the analysis proceeded through several steps. First, we independently read through the transcribed data in order to gain a general understanding of the data. Next, utilizing the interview data from two of the participant teachers, we each developed an initial coding scheme by reviewing the transcripts for emergent themes, with a particular focus on the two research questions (i.e., specific applications of ChatGPT and teacher roles) and identifying subcategories within each theme. We compared our individual analyses and resolved discrepancies through an iterative process of repeated rounds of discussion. We then independently applied the reconciled coding to the remaining interview data. Upon calculating the results of the coding process, we obtained a Cohen's Kappa value of 0.89, which demonstrates a high level of inter-coder reliability as suggested by Lombard et al. (2002). We then resolved all remaining discrepancies through further discussion. To identify and examine patterns and thematic trends in the data, we calculated frequency counts of responses for coding. Representative answers were extracted to capture response patterns in the words of the teachers.

The teacher-chatbot interaction logs were then analyzed to identify the roles performed by ChatGPT and to supplement the interview data. All of the interaction transcripts provided by the teachers were thoroughly reviewed. In the interest of brevity, and as chatbot responses can be directly accessed by readers through the chatbot using the prompts the teachers provided (visit https://chat.openai.com/chat for chatbot use), here we provided only teacher prompts extracted from the interaction logs. The selected prompts were presented alongside interview excerpts to verify the chatbot's roles.

4 Findings

4.1 Roles of ChatGPT in language education

Based on the teachers' interview data and interaction transcripts, four roles of ChatGPT were identified: interlocutor, content provider, teaching assistant, and evaluator. Following are descriptions of how each role was performed by the chatbot.

4.1.1 Interlocutor

Table 2 provides sample prompts utilized by participant teachers that elicited the chatbots' performance of specific functions in the role of interlocutor. Ten teachers utilized the chatbot as an interlocutor, specifically as a role-player, and/or as an interactive game partner. The teachers indicated that they employed the role-play function of the chatbot to provide students with language experiences. To initiate role-plays with chatbots, the teachers provided various prompts such as "Act as a person…" All 10 teachers indicated that they demonstrated how a conversation with the chatbot would proceed and allowed students to contribute to the conversation, while two teachers also had groups of students or individual students practice using the chatbot independently.

Table 2 ChatGPT as interlocutor

Functions	Sample prompts
Role-playing	• Act as a waiter at a pizza restaurant, then I will act as a customer who wants to order • Can you role-play with me? ↳The scene is that I call my friend, Minho, and you answer the phone. You are Minho in this role-play. Can you go first?
Interactive language game	• Let's play a color-guessing game. Give me a hint then I will guess • Let's play a word-guessing game ↳Choose one word from among grade, club, sport, baseball, basketball, and volleyball, and give me a hint. I will guess • I want to play a 20-question game ↳Choose a basic word for 10-year-old EFL students and give me a hint • Let's play a would-you-rather game ↳The topic is food

The arrow "↳" indicates a prompt following a chatbot's response

Additionally, eight teachers reported having the chatbot perform the role of interlocutor as an interactive language game partner. The games included specific word-guessing games, including a color game, an animal guessing game, a 20-question game, and a would-you-rather game. Teacher 3 noted the benefit of the chatbot participating in such games, stating, "When I used language games before, it used to be just me versus students, but now each student can do it with the chatbot, and they can have many more opportunities to use English."

4.1.2 Content provider

As shown in Table 3, the chatbots also performed three different functions in the role of content provider: production and recommendation of materials, customization of materials, and provision of cultural knowledge. All 11 teachers used the chatbot to produce materials, including dialogue scripts, short stories, and sample words or sentences. Regarding this function, Teacher 3 mentioned, "In terms of resource availability, we now have many more options that we can choose from. We do not need to stick to the textbook or ready-made online materials." Additionally, some teachers leveraged the chatbot's expertise to recommend existing materials that were

Table 3 ChatGPT as content provider

Functions	Sample prompts
Production and recommendation of materials	• I want to do a two-member role play. This role-play is for practicing the question-and-answer structure, "What are you doing?" and "I'm …" Can you make some scripts to do the role-play using only simple English? • I can pronounce "good." Can you search for other words having the "oo" sound in the word "good"? • Can you recommend short English cartoon books for 5th-grade EFL students?" • Let me know the names of Disney songs about self-esteem • Can you teach me how to pronounce "interesting" in a Korean accent?"
Customization of materials	• I'm an elementary student in Korea and I don't use English so well. Can you use only simple words if you have the vocabulary list of Key Stage 1? ↳I just want you to use simple English like the above vocabulary • Can you write a story using the question "Are you okay?" and the answer "No, I'm not."? ↳I want you to use only simple sentences. And please write the story in 15 sentences
Provision of cultural knowledge	• Create simple dialogues about "What grade are you in?" ↳Can you add some cultural misunderstanding in the dialogue above? • Do you know the book "The Very Hungry Caterpillar"? ↳Can you show me some sample sentences from the book? ↳Can you exchange the words about the food above for Italian food? • Do people in the US take off their shoes in their homes? ↳What about Singapore?

The arrow ↳ indicates a prompt following a chatbot's response

tailored to the specific needs of their students, such as English storybooks on specific topics that were otherwise difficult to find.

All the teachers ($n = 11$) also mentioned that the chatbot could modify materials it produced or existing materials. This function was particularly useful for teachers who desired to provide individualized materials to students. They all expressed that the chatbot's ability to generate and revise different levels and types of materials was a primary reason for them to want to continue to use it. For example, Teacher 5 said, "I do not need to spend a lot of time on developing individual materials as in the past. I can simply ask the chatbot to make more difficult or easier versions of a material."

Last, some teachers ($n = 6$) stated that they could acquire cultural knowledge from the chatbot or ask the chatbot to make materials culturally relevant. The teachers stated that they used to have difficulty understanding the cultures of different countries. However, they found that the chatbot could not only provide cultural knowledge but also produce or modify lesson materials in manners that reflected cultural aspects, thereby making their language classes more engaging and culturally relevant at the same time than before.

4.1.3 Teaching assistant

Table 4 displays specific functions in the chatbots' role as teaching assistant and sample prompts provided by participant teachers. The teachers used the chatbot to help students resolve learning difficulties by allowing them to interact with the chatbot during the class or using it themselves. The chatbot performed the role of teaching assistant in three different ways. All 11 teachers interviewed reported utilizing the chatbot as a grammar checker or online dictionary. Among them, two teachers further stated that they also allowed students to use the chatbot during class to encourage them to find errors in their vocabulary or sentences or search for word meanings whenever necessary. In addition, a few teachers ($n = 2$) indicated that they used the chatbot as a background knowledge activator to give students a preview of what they were going to learn about a topic.

Table 4 ChatGPT as teaching assistant

Functions	Sample prompts
Grammar checker	• I wrote the answer "I want go to Jeju-do" to the question "What do you want to do?" Is there any error in my answer? • Can you check for any grammar mistakes?
Dictionary	• I want to know about specific contextual examples with the words "happy," "excited," and "exhilarated." What are the differences among those words? • What are some phrases that I can use instead of "How are you"? • What is the difference between "I am heading" and "I am headed"?
Background knowledge activator	• Can you tell me the origin of "banana"? • Can you show the words related to the category of "clothes"?

4.1.4 Evaluator

Table 5 shows specific functions the chatbots performed in the role of evaluator and sample prompts provided by teachers. Two teachers used their chatbots to provide initial grading of students' writing. Regarding this role, Teacher 10 stated that she elaborated on some feedback in the chatbot's initial assessment and ignored some information that she deemed irrelevant. Meanwhile, many of the teachers ($n = 9$) stated that they used the chatbot to produce testing materials. For example, the teachers stated that they asked the chatbot to produce multiple-choice questions, O/X questions, and gap-filling dialogues on specific topics.

Only two teachers noted that the dialogue records automatically saved on the chatbot website provided useful information about student performance. Teacher 7, who distributed tablet PCs to allow students to use the chatbot individually, mentioned:

> After a lesson, I brought back tablet PCs that I gave to students and examined the dialogue history with each student to determine the degree to which the student performed well or not. In this manner, I could obtain more accurate assessment information about my students.

4.2 Teachers' roles when using ChatGPT

In the interviews, all teachers acknowledged the educational potential of ChatGPT and expressed their willingness to use it in the future. They also agreed that ChatGPT would not replace human teachers, but rather support them in maximizing their professional expertise. That is, although they found ChatGPT to be a powerful tool that could automate many aspects of instruction, it was still just perceived as a tool that could enhance teachers' pedagogical effectiveness but that could not substitute for teacher-student interaction. Rather, the teachers believed that their teaching profession might need to be redefined with the assistance of this versatile technology in the following three roles of teachers, which emerged as important in the interviews.

Table 5 ChatGPT as evaluator

Functions	Sample prompts
Assessment	• This is a sixth-grade elementary EFL student's daily diary. Can you suggest some feedback for her?
Evaluation material production	• Let me know the full lyrics of the song, "Let It Go." ↳Make 10 gaps with a few words above • (After inputting a story from a textbook) ↳Make 10 O/X questions about it
Dialogue history	*Teachers assess student performance based on dialogue histories automatically saved in ChatGPT*

The arrow ↳indicates a prompt following a chatbot's response

4.2.1 Orchestrating different resources while making quality pedagogical decisions

All participants agreed that using ChatGPT would require teachers to become highly skillful at managing the plethora of teaching resources afforded by the technology so as to design creative, organized, and engaging lessons. With ChatGPT, teachers would now have significantly more control over creating and revising lesson materials than in the past, when they had access to fewer teaching resources and had to manually develop and revise materials.

However, the teachers also noted that the increased availability of helpful resources would not automatically raise the quality of instruction. They mentioned that it depended on human teachers' pedagogical knowledge and judgment in selecting materials appropriate to their students' needs and how effectively they used them within their instructional contexts. As Teacher 8 stated, "Although we can much more easily generate and modify lesson materials using ChatGPT, we must carefully consider how we will use the materials in the classroom," predicting that "there will be a significant disparity in the quality of lessons between teachers who possess the ability to effectively integrate the chatbot into their pedagogy and those who lack such proficiency." Similarly, emphasizing the importance of the connection between teachers and students, Teacher 9 stated:

> ChatGPT provides useful materials, but the ultimate acceptance of these materials by students can only be determined through a process of thoughtful reflection, taking into account factors such as the rapport between the teacher and students, individual students' needs and characteristics, and the collective dynamic of the classroom.

In conclusion, the teachers made it clear that ChatGPT has the potential to serve as a valuable instructional tool that can enrich teaching practice, but they emphasized that it would need to be employed in a manner that complements and enhances teachers' pedagogical expertise based on the nuanced understanding of their students and contexts.

4.2.2 Making students active investigators, not passive knowledge recipients

With regard to teachers' roles in using ChatGPT to promote students' learning, nine teachers emphasized the importance of formulating good questions and modeling effective questioning techniques to obtain high-quality results from the chatbot. Teacher 1 stated, "The ability to ask a question or in some instances a series of questions determines whether we can obtain desired information or not." Also, regarding this issue, Teacher 8 shared a story about developing good questioning strategies:

> I was creating a pronunciation guide for basic vocabulary for my beginning-level students. I wanted the chatbot to provide me with written pronunciations of English words in Korean, for example, the English pronunciation of ball can be written in Korean as 볼 [bol]. I initially attempted prompts such as "Can you tell me the pronunciation in Korean with this word, *interesting*?" or

> "Can you teach me how to pronounce *interesting* in a Korean accent?" I was unsuccessful and thought the chatbot was not able to do it. I gave up. Later, a coworker suggested trying the following prompt: *Can you write the English pronunciation of interesting in Korean?* It worked and I got the information I wanted."

In a similar vein, the teachers mentioned the importance of teaching students how to formulate questions effectively, which they believed influenced the extent to which students would benefit from their interactions with the chatbot. Teacher 4, for example, stated,

> When I encouraged my students to interact with the chatbot, some of them actively asked questions about what they wanted to know, experimenting with various question formats. However, I also discovered that some students did not even know how to ask questions. The most important thing that we teachers should do is to help them make appropriate questions and craft them to obtain the best information.

Four teachers said that they prepared worksheets consisting of questions for students to ask the chatbot, but Teacher 11 stated, "it was not possible to anticipate every question that students might ask." Instead, they all underscored the importance of the role of teachers that prepare students to construct and craft their own questions depending on their needs. That is, with the rich language resources of ChatGPT, the teachers agreed that it was now even more important not to treat students as passive knowledge recipients; rather, teachers shared that they would need to create a learning environment where students can be active investigators who learn by skillfully developing a variety of questions about the subject matter on their own.

4.2.3 Raising ethical awareness regarding the use of AI

All teachers interviewed emphasized the importance of ethical considerations in the use of ChatGPT and guiding students to use the technology responsibly. In this regard, two aspects of ethical awareness emerged in the interviews: making the use of the technology transparent and setting rules for ethically interacting with it.

First, all teachers emphasized the need for establishing a shared understanding of the degree to which students might draw on ChatGPT, while creating an atmosphere where students can openly talk about what information they obtained from the chatbot. Granting that it would not be feasible to monitor all students' uses of the technology, some mentioned that they should instruct students to not merely rely on ChatGPT to obtain what they need to complete tasks; instead, they argued that teachers should acknowledge students' efforts to obtain information from the chatbot while encouraging them to use it to further develop their own understanding. Specifically, Teacher 2 elaborated on this:

> Teachers have to make a classroom atmosphere where using the technology itself is not ethically wrong but where the behavior of hiding how they got information from it and pretending that the knowledge gained from it is

entirely their own is wrong. Students do not need to feel that they should hide their use of the chatbot. Rather, we have to acknowledge students' efforts to use the chatbot but only in a way that they can further their own learning. To do this, I believe we have to first make it acceptable for students to openly talk about what they obtained from the chatbot and how they used it to facilitate their learning experience.

Next, most of the teachers ($n = 10$) also expressed concerns about inappropriate or unethical student behavior in interactions with the chatbot and shared anecdotes regarding such behavior. Specifically, these teachers observed that some students tried to use inappropriate words with the chatbot or tried to obtain inappropriate information from it. Specifically, to address this issue, Teacher 7 mentioned, "A clear guideline outlining appropriate and ethical usage of the chatbot must be established before we introduce it to students." Overall, the consensus among the teachers was that ChatGPT should be implemented in conjunction with thoughtfully composed ethical guidelines covering a range of considerations from the utilization of information obtained through the technology to the manner in which students should interact with it.

5 Discussion

This research explored the educational potential of ChatGPT and language teachers' perceptions on how their roles would change as a result of its use. By qualitatively analyzing the teachers' chatbot-use logs and conducting in-depth interviews, we were able to identify the dynamic nature of the relationship between teachers and AI, with both parties found to complement each other. The findings highlight that as technology becomes increasingly multifunctional, teachers may need to assume more critical and professional roles to integrate the technology in a way that best benefits students. Therefore, this study provides important areas for discussion regarding both chatbots and teacher-AI collaboration in education.

Previous studies on educational chatbots have consistently identified the numerous limitations of chatbot technology, such as its limited capability to engage in open-ended conversations with users and to continue an extended, goal-oriented conversation on a particular topic (Bibauw et al., 2019; Jeon, 2022b). In response to these limits, the current research introduced an LLM-powered chatbot, as it can simulate a more human-like conversation and respond to a broader range of user inquiries. Furthermore, the augmented capabilities enabled the chatbot to perform various pedagogical roles compared to chatbots previously introduced in the literature (Ji et al., 2023). For example, Kuhail et al. (2022) discovered that the majority of educational chatbots introduced in research mainly fulfilled one of four roles (i.e., teaching agent, peer agent, teachable agent, and motivational agent). Similarly, Ji et al. (2023) showed that the primary role of chatbots in language education was the role of interlocutor. In contrast to these studies, teachers in the current study utilized ChatGPT in a broader capacity across several instructional phases. In other words, ChatGPT served not only as a content provider, teaching assistant, and evaluator, but

also assumed the role of a conversation practice partner as an interlocutor, showing its versatility and potential to play multiple roles across various instructional contexts (Celik et al., 2022). Overall, this study presented empirical evidence regarding the use of ChatGPT in education, enriching our understanding of how LLM technology may be incorporated into the educational context.

Regarding the relationship between the teachers and the chatbot, this study's findings are consistent with earlier research that supported teacher-AI collaboration (Holstein & Aleven, 2022; Holstein et al., 2020; Kim et al., 2022). Research on teacher-AI collaboration has shown the importance of teachers' agentive roles when using AI. It argues that more effective learning is jointly created by AI and human facilitators, rather than teachers or AI working alone (Bower, 2019; Holstein et al., 2020). For example, Holstein and Aleven (2022) showed that teachers could make better decisions based on students' past learning data provided by smart glasses. Zhang et al. (2022) illustrated that parents created a better storytelling environment by selecting the most pertinent questions at any given moment among different questions generated by AI. Similarly, we demonstrated that teachers' pedagogical expertise could be maximized with the help of the capabilities of the chatbot. It is of note that this study presented contradictory evidence to the concern that teachers may become overly dependent on LLM technology (Kasneci et al, 2023). Conversely, teachers in this study stated that without adequate student training and clear guidelines, students would not be able to use the chatbot in an effective and ethically appropriate manner, suggesting that the use of the chatbot may not diminish the importance of teachers' roles.

Additionally, by using ChatGPT, which can be utilized across many application domains, this study expands on the findings from earlier studies that examined teachers' relationship with AI tools built with a particular educational purpose (e.g., Holstein & Aleven, 2022; Zawacki-Richter et al., 2019; Zhang et al., 2022), presenting empirical evidence that the increasing versatility of AI technology requires teachers to make an even wider range of and more professional pedagogical decisions. This was demonstrated by three teacher roles that emerged during teacher interviews (i.e., orchestrating different resources with quality pedagogical decisions, making students active investigators, and raising ethical awareness regarding the use of AI). It is of significant note that these three roles underscore the importance of teacher-student interaction in understanding the teacher-AI relationship (Choi et al., 2023a; Luckin et al., 2022; Kim et al., 2022; Xu & Ouyang, 2022), as Teacher 9 stated that teachers' pedagogical decisions regarding ChatGPT should be made based on the rapport between the teacher and students, individual students' needs and characteristics, and the collective dynamic of the classroom.

6 Limitations

This research has some limitations to be considered in future studies. First, the participants were selected on the basis of their interest in and willingness to use ChatGPT. This indicates that they might have been more technology-friendly than the average among teachers. Also, the sample in this study comprised only a small

number of English teachers working at elementary schools located in one country; therefore, the findings may not be broadly generalizable. Also, we examined participant teachers as one group without consideration of their differences. Future investigations should involve teachers representing a diverse range of subjects, school levels, countries, and attitudes toward technology for a more comprehensive understanding of teachers' uses of LLM-powered chatbots to support further theorizing. Next, this study did not address the experiences and perceptions of students. An important contrast to this exploration will be an examination of the use of ChatGPT from the student perspective, such as a qualitative analysis of students' perceptions or a quantitative examination of the chatbot's effect on students' performance or motivation. This would provide an informative and interesting complement to this study. Finally, given the ubiquity and prevalence of new technologies, it would be relevant to investigate the impact of teachers' previous experiences with AI and their effect on teachers' acceptance or use of LLM-powered chatbots as educational tools (Choi et al., 2023b).

7 Implications and conclusion

Important implications for the use of AI in education can be drawn from this study, in particular, the value of using an LLM-powered chatbot as a multifunctional tool. First, this study is among the initial attempts to empirically examine the educational potential of the LLM technology, ChatGPT. To be specific, expanding on theoretical explorations with empirical evidence, this research found that the chatbot serves different instructional roles, including those of interlocutor, content provider, teaching assistant, and evaluator. Second, this study provided concrete examples of teacher prompts used to elicit useful responses. Teachers, teacher educators, and researchers can use this finding to initiate further exploration into the potential of LLM-powered chatbots and to develop different application methods based on their own educational contexts. Third, the study offered insight into the dynamicity of the human-AI relationship that changes in accordance with the rapid advance of AI technology. In the teacher interviews, it was indicated that the roles of teachers are predicted to increasingly become both multifaceted and specialized as AI tools continue to evolve more. Thus, this research provides an initial reference for further exploration into how this complementary relationship will unfold in the future, as technology develops even more and across different local educational contexts. Last, confirming the idea that teaching with AI requires specific teacher competencies (e.g., Celik, 2023; Choi et al., 2023a), this study shows a need for teacher training courses specifically designed for teaching with an LLM-powered chatbot. After the use of ChatGPT for a period of two weeks, the teachers in this study concurred that obtaining quality output from the chatbot depends on the teacher's ability to develop quality questions. In this regard, the need for preservice education and professional development programs to help teachers align the resources made available by technology with their pedagogical purposes will be another important issue (Jeon et al., 2022).

To conclude, this research took a crucial step forward to understand the pedagogical potential of LLM technology. Further investigation of this emerging technology, including different types of LLM technology, is necessary to determine how it can be incorporated into education as a tool for teachers, as well as students. This may require an in-depth understanding of the AI-human relationship that comes from a comprehensive knowledge of teachers' experiences with and perceptions of the ever-more sophisticated technologies that will inevitably become integrated into their profession.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflicts of interest None

References

Altman, S. (2022, December 5). *Twitter*. https://twitter.com/sama/status/1599668808285028353?s=20&t=j5ymf1tUeTpeQuJKlWAKaQ

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

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

Bower, M. (2019). Technology-mediated learning theory. *British Journal of Educational Technology, 50*(3), 1035–1048. https://doi.org/10.1111/bjet.12771

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, Jeffrey, Winter, C. …, & Amodei, D. (2020). *Language Models are Few-Shot Learners.* arXiv. https://arxiv.org/abs/2005.14165

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, 107468. https://doi.org/10.1016/j.chb.2022.107468

Celik, I., Dindar, M., Muukkonen, H., & Jarvela, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66*, 616–630. https://doi.org/10.1007/s11528-022-00715-y

Choi, S., Jang, Y., & Kim, H. (2023a). Influence of Pedagogical Beliefs and Perceived Trust on Teachers' Acceptance of Educational Artificial Intelligence Tools. *International Journal of Human–Computer Interaction, 39*(4), 910–922. https://doi.org/10.1080/10447318.2022.2049145

Choi, S., Jang, Y., & Kim, H. (2023b). Exploring factors influencing students' intention to use intelligent personal assistants for learning. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2023.2194927

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative approaches to research* (3rd ed.). Merrill/Pearson Education.

Dillenbourg, P., & Jermann, P. (2010). Technology for classroom orchestration. In M. S. Khine & I. M. Saleh (Eds.), *New science of learning* (pp. 525–552). Springer Science + Business Media.

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

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

Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior, 93*(December 2018), 279–289. https://doi.org/10.1016/j.chb.2018.12.023

Heidt, A. (2023). 'Arms race with automation': Professors fret about AI-generated coursework. *Nature*. https://www.nature.com/articles/d41586-023-00204-z

Hew, K. F., Huang, W., Du, J., & Jia, C. (2023). Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions. *Journal of Computing in Higher Education, 35,* 40-68. https://doi.org/10.1007/s12528-022-09338-x.

Holstein, K., Aleven, V., & Rummel, N. (2020). A Conceptual Framework for Human–AI Hybrid Adaptivity in Education. In: I. Bittencourt, M., Cukurova, K., Muldner, R., Luckin, & E., Millán (Eds.), *Artificial Intelligence in Education. AIED 2020. Lecture Notes in Computer Science* (vol. 12163). Springer, Cham. https://doi.org/10.1007/978-3-030-52237-7_20

Holstein, K., & Aleven, V. (2022). Designing for human–AI complementarity in K-12 education. *AI Magazine, 43*(2), 239–248. https://doi.org/10.1002/aaai.12058

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–ai complementarity. *Journal of Learning Analytics, 6*(2), 27–52. https://doi.org/10.18608/jla.2019.62.3

Huang, X., Zou D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Society & Technology 26*(1), 112–131. https://www.j-ets.net/collection/forthcoming-articles/26_1

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*, 237–257. https://doi.org/10.1111/jcal.12610

Hwang, G. J., Xie, H., Wah, B. W., & Gasevic, D. (2020). Vision, challenges, roles, and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jeon, J. (2021). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2021.1987272

Jeon, J. (2022a). Exploring a self-directed interactive app for informal EFL learning: a self-determination theory perspective. *Education and Information Technologies, 27*(4), 5767–5787. https://doi.org/10.1007/s10639-021-10839-y

Jeon, J. (2022b). Exploring AI chatbot affordances in the EFL classroom: young learners' experiences and perspectives. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2021.2021241

Jeon, J., Lee, S., & Choe, H. (2022). Enhancing EFL pre-service teachers' affordance noticing and utilizing with the Synthesis of Qualitative Evidence strategies: An exploratory study of a customizable virtual environment platform. *Computers & Education, 190*, 104620. https://doi.org/10.1016/j.compedu.2022.104620

Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: focusing on the collaboration with human teachers. *Journal of Research on Technology in Education, 55*(1), 48–63. https://doi.org/10.1080/15391523.2022.2142873. Advance online publication.

Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Jurgen, P., Poquet, O., Sailer, M., Schmidt, A., Seidel, T. …, & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies, 27*, 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2022). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*. https://doi.org/10.1007/s10639-022-11177-3. Advanced online publication.

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

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x

Luckin, R., Cukurova, M., Kent, C., & du Boulay, B. (2022). Empowering educators to be AI-ready. *Computers and Education: Artificial Intelligence, 3*, 100076. https://doi.org/10.1016/j.caeai.2022.100076

OpenAI. (2023). *ChatGPT*. OpenAI. https://chat.openai.com/chat

Stokel-Walker, C. (2023). AI Bot ChatGPT writes smart essays--should professors worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7. https://www.nature.com/articles/d41586-022-04397-7

Susnjak, T. (2022). *ChatGPT: The end of online exam integrity?* arXiv. https://doi.org/10.48550/arXiv.2212.09292

Timpe-Laughlin, V., Sydorenko, T., & Daurio, P. (2022) Using spoken dialogue technology for L2 speaking practice: what do teachers think? *Computer Assisted Language Learning, 35*(5–6), 1194–1217. https://doi.org/10.1080/09588221.2020.1774904

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education, 194*, 104703. https://doi.org/10.1016/j.compedu.2022.104703

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45.

Xu, W., & Ouyang, F. A. (2022). A systematic review of AI role in the educational system based on a proposed conceptual framework. *Education and Information Technologies, 27*, 4195–4223. https://doi.org/10.1007/s10639-021-10774-y

Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhai, X. (2022). *ChatGPT User Experience: Implications for Education.* arXiv. https://doi.org/10.2139/ssrn.4312418

Zhang, Z., Xu, Y., Wang, Y., Yao, B., Ritchie, D., Wu, T., Mo, Y.,Wang, D., & Li, T. J. J. (2022). StoryBuddy: A human-AIcollaborativechatbot for parent–childinteractive storytellingwith flexible parental involvement. In *CHI '22: CHI conferenceon human factors in computing systems* (pp. 1–21).https://doi.org/10.1145/3491102.3517479

Authors and Affiliations

Jaeho Jeon[1] · Seongyong Lee[2]

Jaeho Jeon
jaehojeon21@gmail.com

[1] Department of Literacy, Culture, and Language Education, Indiana University, 107 S. Indiana Avenue, Bloomington, IN 47405-7000, USA

[2] Department of English Education, Hannam University, 70 Hannam-Ro, Daedeok-Gu, Daejeon 34430, Republic of Korea