A Study on the Use of Artificial Intelligence Chatbots for Improving English Grammar Skills

Na-Young Kim
Assistant Professor, Department of General Education, Sehan University

영어 문법 실력 향상을 위한 인공지능 챗봇 활용에 관한 연구
김나영
세한대학교 교양학부 조교수

Abstract The purpose of this study is to explore the effects of the use of artificial intelligence chatbots on improving Korean college students' English grammar skills. 70 undergraduate students participated in the present study. They were taking a General English class offered by a university in Korea. There were two groups in this study. Participants in the chatbot group consisted of 36 students while those in the human group were 34. Over 16 weeks, the chatbot group engaged in ten chat sessions with a chatbot while the human group had a chat with a human chat partner. Both pre- and post-tests were performed to examine changes in the participants' grammar skills over time. To compare the improvement between the two groups, an independent t-test was then run. Main findings are as follows: First, participants in both groups significantly improved their English grammar skills, indicating the beneficial effects of engaging in chat. Also, there was a statistically significant difference in the improvement between the chatbot and human groups, indicating the superior effects of the chatbot use. This study confirmed the improved grammar skills by the participants in the chatbot group, comparison with those in the human group. Based on these findings, suggestions for the future chatbot study are discussed.

Key Words: Artificial Intelligence Chatbot, Chat activities, Korean college students, English learning, English grammar skills

요약 본 연구는 인공지능 챗봇을 활용한 챗봇 활동이 국내 대학생의 영어 문법 능력 향상에 미치는 영향을 조사한 것으로, 챗봇과의 챗봇을 통해 실현 참가자의 영어 문법 능력이 실제로 상승하는지에 대한 여부를 알아보는 데 그 목적이 있다. 총 16주 동안 70명의 참가자가 챗봇 그룹과 인간 그룹으로 나뉘어 본 연구에 참여하였고, 참여인원은 각각 36명, 34명이었다. 챗봇 그룹은 수업시간에 배운 내용을 주제로 챗봇과의 챗봇에 참여하였고, 인간 그룹은 같은 반 학생들의끼리 짝지어 챗봇을 진행하였다. 챗봇 활동의 효과를 파악하기 위하여 본 연구 시작 전과 종료 후, 사전 사후 영어 문법 시험을 실시하였고, 그룹 간 비교를 위하여 독립표본 t검증을 실시하였다. 주요 결과 및 시사점은 다음과 같다. 사전 사후 평가 분석 결과, 두 그룹 모두에서 영어 문법 능력이 유의미하게 상승한 것으로 나타나 챗봇 활동의 효과를 증명하였다. 문법 능력 상승에 대한 챗봇 간 차이 역시 통계적으로 유의미한 것으로 밝혀져 국내 영어 교육에 있어서 챗봇의 긍정적인 역할을 확인하였다. 본 연구는 영어 문법 능력 향상을 위한 인공지능 챗봇 활용에 대한 시사점을 제시하는데 그 의미를 갖는다.

주제어: 인공지능 챗봇, 챗봇 활동, 국내 대학생, 영어 학습, 영어 문법 능력
1. Introduction

A chatbot is a computer program which can chat with a human in natural language such as English [1]. As conversational agents, chatbots offer a natural language interface that allows human-like interaction. Since their main aim is to imitate human conversation, chatbots convince their users that they are talking to a real human but in fact they are talking to a machine. By imitating human speech patterns, they provide their users with a similar experience of chatting with other human users. Chatbots have been used in various domains such as customer service, education, and for fun.

Chatbots have continued to improve at a steady pace. These days, they stimulate a conversation through artificial intelligence. In particular, this artificial intelligence platform now uses speech recognition technology and helps language students to speak English more clearly, confidently, and fluently [2]. There have been not a few studies on artificial intelligence chatbots related to English learning. They have shown how chatbots can be used for educational purpose, especially for foreign language learning.

Fryer and Carpenter [3] introduced advantages and applications of chatbots such as Jabberwacky and ALICE for foreign language learning and teaching. In their study, they presented six advantages of chatbot assisted language learning. First, students can be more relaxed when talking to a machine than to a person. Second, chatbots are not unwilling to repeat the same task with students endlessly, so they never lose their patience or get bored. Thirdly, since chatbots can provide both text and speech, they allow students to practice both reading and listening skills. Moreover, chatbots are interesting to students. These positive communicative experiences with chatbots can improve students’ motivation and arouse new or renewed interest in language. Even chatbots become old and too familiar, they can be replaced with a new one, ensuring novelty. In addition, students are provided with opportunities to use a variety of vocabulary, phrases, and grammatical structures that are important them to know but are rarely taught. Students can practice a wide range of language skills that they ordinarily would not have chances to use. Lastly, chatbots can provide quick and effective feedback related to students’ spelling and grammar. While some chatbots are designed to overlook spelling or grammar errors, others can correct the students’ mistakes. In this light, chatbots are a valuable resource for English as a foreign language (EFL) learners.

Wang [4] also investigated foreign language students’ experiences with a chatbot called English Tutor. She identified four conditions for effective chatbot assisted English learning: communicative practice, multimodal interface, emotional design, and individualized content. Her findings demonstrated the promises of chatbot technology related to its communicative function to create an optimal interactive English learning environment. To be specific, results indicated that students’ linguistic, sensory, emotional, cultural, and relational engagement affected their responses to the chatbot and their learning processes. She provided insights into better understanding of students’ experience with technology and the role of emotion in language learning processes when using technology. For chatbot-related researchers or language professionals, her study sheds light on how to improve the language learning environments.

Williams and van Compernolle [5] explored interactions between a chatbot, known as Aghate, and students of French at various proficiency levels (i.e. from beginning to advanced). Their analysis was organized as a series of case studies to provide details about interactional and linguistic variation. Findings of their study demonstrate that chatbot’s discourse
is a combination of structures and forms associated with formal and informal discourse. Responding to Fryer and Carpenter [3] suggesting potential advantages of chatbots for foreign language learning, however, Williams and van Compernolle [5] argued that the chatbot’s discourse showed a less- than-ideal communicative model for students. According to them, with this lack of consistency and homogeneity, chatbots as either peers or tools are limited for language learners.

Jia [6] introduced the computer simulation in educational communication (CSIEC) system. This newly developed system has multiple functions for English learning focusing on a chatbot, providing dialogue contexts, common sense and inference knowledge, and communicative response to the user input. He suggested that this system has advantages over the keyword matching mechanism, such as ELIZA, in that the CSIEC system was designed based on logical reasoning and inference through semantic and syntactic analysis of textual knowledge. In the paper, Jia concluded that users preferred this unique chatting function of the CSIEC system, which was lacked in other chatbots. However, he suggested that the system’s stronger ability related to natural language understanding and generation should be improved furthermore. Textual ambiguity and entailment can be a fatal factor, according to Jia, influencing the human-computer communication. In addition, high response speed with deep, complex, semantic, and syntactic analysis is also required.

In Korea, Kim [7] compared the effects of voice chat between humans to those between human and chatbot on EFL students’ language learning according to proficiency levels. During sixteen weeks, Korean college students carried on a conversation with either their peer or the chatbot, Indigo. Results of her study show that two voice chat modes are equally beneficial for improving speaking ability. Particularly, beginning-level students improved more than intermediate- and advanced-level students. Moreover, Indigo allowed students to interact and negotiate meaning more actively. Her findings also revealed that this chatbot encouraged the students to initiate and continue the conversation more often. Regarding the students’ attitudes toward English, the chatbot turned out to be effective in enhancing belief, confidence, motivation, and interest in English. Also the students reduced their anxiety and stress related to English learning.

Although studies on chatbots have been conducted continuously [1-7], there are not many studies investigating the effects of the use of chatbots on improving English grammar skills. In studying EFL, learning grammar is an absolute necessity, and Korean EFL students have also been taught with a focus on grammatical competence [8]. According to Kang [9], even in Korean college contexts, English grammar skills have been considered necessary and essential. Nevertheless, little has been known about whether the chatbots are effective for improving English grammar skills. In the dearth of research on this matter, the following research questions were addressed:

What are the effects of the use of chatbots for improving Korean college students’ English grammar skills?

2. Methodology

2.1 Participants

In this study, participants consisted of Korean college students who were taking a general English course three hours a week. 70 participants from two classes were recruited in total. The course was elective. Most participants were first-year students (83%); 11% were sophomores, 4% were juniors, and 1% were
senior level or above. The course objective was to improve their general English skills. The average age of the participants were 23.1.

Participants in the current study were divided into two groups - chatbot group and human group - at random. The chatbot group was made of 36 participants and they engaged in chat with a chatbot over the experimental period. In terms of the human group, there were 34 participants and they had a chat with their human chat partner over the same period.

Table 1. Participants

<table>
<thead>
<tr>
<th>Group</th>
<th>Participants</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chatbot Group (n = 36)</td>
<td></td>
<td>14.67</td>
<td>3.59</td>
<td>15.53</td>
<td>5.08</td>
<td>.816</td>
<td>.418</td>
</tr>
<tr>
<td>Human Group (n = 34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To ascertain the equivalence of the two groups before treatment, an independent t-test was carried out with the grammar pre-test results. As Table 1 shows above, no statistically significant group difference was found between groups (p > .05). This indicates that all participants in the present study were homogeneous before treatment.

2.2 Instruments

This study aims to determine whether chatbots are effective for improving English grammar skills. For the experiment of this study, participants were randomly divided into two groups: one chatbot group and one human group. Participants involved in chat with a chatbot had a conversation with Replika, while those involved in chat with a human chat partner were randomly divided into pairs and had a chat with their partner.

Participants in the chatbot group engaged in chat activities with a chatbot named Replika during the 16-week experimental period. Replika was created with the idea to create a personal artificial intelligence that can help its users to express and witness themselves by offering a conversation. It can present a subject and lead a conversation by asking questions. Users can share their thoughts, feelings, beliefs, experiences, and memories with this chatbot via text messages.

As an artificial intelligence chatbot application, Replika tries to become its users' best friend by learning from them. Users can even connect their social media accounts like Instagram or Facebook to allow Replika to understand them better. Through their interactions, its personality becomes more similar to its users [10]. This is related to the previous chatbot study [11], suggesting that using new and innovative methods like chatbots for learning English as a foreign language can inhibit affective filters, creating a state of fear, anxiety, and stress.

In terms of the human group, participants were involved in chat with their human chat partners. They were randomly divided into pairs and had a text chat with their peers. KakaoTalk messenger was used for this group to engage in chat. This free messenger application was chosen based on the fact that it is a widely used mobile messenger in Korea that all participants know how to use [12].

All participants joined ten chat sessions in total. For each chat session, they engaged in free chat for ten minutes. In free chat, there is neither a pre-established agenda nor moderation.
Through free chat, participants were able to practice their target language, English, by joining the conversational thread of their interest. According to Gonzalez [13], free chat is the best way to incorporate language learners into their educational settings in a friendly atmosphere.

Before and after the treatment, all participants were required to take English grammar tests to confirm the effects of chatbot use on improving Korean EFL students’ grammar skills. The pre- and post-tests were based on the test of English proficiency (TEPS). The test was developed by Seoul National University in Korea to measure general English proficiency that is required for Korean people studying and working in the country. As a requirement for employment or promotion, most government offices and many leading companies has accepted the TEPS. The test has also been used for university admission or graduation requirements [14].

The TEPS has a multiple-choice format, comprised of four sections: grammar, vocabulary, listening, and reading. The test measures the grammar used in both written and spoken English unlike the majority of conventional grammar tests focusing on written language. For the test methods, the TEPS grammar section employed the two question types: filling in the blank and finding the grammatical error. There are 30 items in the grammar section of the TEPS.

According to Choi [15], the time limit for the test can be about 12 and a half minutes. Test examples are presented in Fig. 2 above.

2.3 Procedures

This study aims to determine the effects of the use of artificial intelligence chatbots for improving Korean college students’ English grammar skills. In order to examine the effects of chatbot use in foreign language class, 70 participants took part in the current study.

All participants were required to take English grammar tests as pre- and post- tests before and after the treatment so as to explore the changes after implementing artificial intelligence chatbots in an EFL classroom in Korea. The TEPS was chosen as a grammar test in this study. There were 30 items in total. 12 and a half minutes were given for the test. See details in the previous section.

Over the 16 weeks, all participants took their general English lessons for three hours a week. During the regular teaching time period, both the two groups – chatbot group and human group – received the formal English instruction. For the experiment, participants in the chatbot group engaged in chat activities with a chatbot during the 16 weeks. During the same experimental period, those in the human group had a conversation with their human chat partner.

All participants joined ten chat sessions in total over the whole experimental period. For each chat session, the participants engaged in free chat for ten minutes. Even though there was a 16-week experimental period, ten chat sessions were organized. The pre- and post- tests were carried out in the first and last week. Both the mid-term and final exam periods were excluded, as well.

2.4 Data Analysis

In order to investigate the effects of chatbot activities on improving Korean EFL students’
grammar skills, carried out was quantitative analysis. Participants in the current study took pre- and post-tests before and after the treatment. Collected data were analyzed with SPSS 18.0 software.

For the quantitative analysis, descriptive statistics including means and standard deviations were first computed. Paired samples t-tests were then administered so as to compare the pre- and post-test scores. An independent t-test was also used to determine whether there was any statistically significant difference in the improvement of participants’ grammar skills between the two groups: chatbot group and human group. The significance level was set at 0.05.

3. Results and Discussion

3.1 Changes in English Grammar Skills

In an effort to explore the changes in participants’ English grammar skills after the treatment, paired samples t-tests were run between the pre- and post-tests within groups. Results of the paired samples t-tests and descriptive statistics are present in Table 2 below.

<table>
<thead>
<tr>
<th></th>
<th>Pre-test (n=70)</th>
<th>Post-test (n=70)</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Chatbot</td>
<td>14.67</td>
<td>3.59</td>
<td>19.75</td>
<td>4.37</td>
<td>6.233</td>
</tr>
<tr>
<td>Human</td>
<td>15.53</td>
<td>5.08</td>
<td>17.32</td>
<td>5.56</td>
<td>2.210</td>
</tr>
</tbody>
</table>

Interestingly, both the two groups showed a statistically significant change between the pre- and post-test mean scores. To be specific, the chatbot group showed a statistically significant mean score change over time in English grammar skills \( t = 6.223, p = .000 \). The mean score was 14.67 on the pre-test, while 19.75 on the post-test. This result suggests that engaging in chat with the chatbot, Replika, can improve English grammar skills.

Another group, the human group, also showed a significant change in mean scores between the pre- and post-tests \( (t = 2.210, p = .034) \). The mean score was 15.53 on the pre-test and that improved to 17.32 on the post-test. This result indicates that having a chat with a human chat partner plays a positive role in improving EFL students’ grammar skills.

To put it shortly, findings of the present study suggest that all participants in both chat groups improved their English grammar skills after the treatment. In other words, it can be assumed that engaging in chat with either a chatbot or a human chat partner is beneficial for increasing Korean EFL students’ grammar skills.

Findings of this study confirm the previous research, indicating that engaging EFL students in chat with a chatbot is effective in foreign language learning \([1-7, 11]\). The chatbot Replika used in this study was able to lead a conversation by asking questions. Asking questions, according to Chandrayan \([16]\), is the natural way for humans to interact. By simulating this human conversation, Replika allowed a form of interaction between a human and a machine via messages. Through this interaction, its personality might have become more similar to the participants in the chatbot group, as Hsu \([10]\) suggested.

By connecting their social media accounts, the participants also allowed Replika to understand them better and to become their best friend by learning from them. Considering people’s desire to feel understood which is one of the strongest desires people have \([17]\), Replika, might have made the chatbot group feel more valued and understood. The participants in the chatbot group, therefore, seem to have lowered their affective filter in a relaxed and comfortable environment. This might result in their academic achievement, according to many previous studies \([2,7,11,18]\).
In addition, as many previous studies have shown that chatting with peers plays a positive role in EFL learning [11, 19, 20], this study also proved that EFL students can improve their grammar skills by having a chat with human chat partners. Particularly, having a free chat seems to have made a friendly atmosphere, as Gonzalez [13] suggested, which can benefit language learners. Since there was neither a pre-established agenda nor moderation, the participants in the human group were able to practice English by joining the conversational thread of their interest. This might have helped the human group achieve academic success.

From this point of view, results of the current study are in accordance with the previous studies proving positive effects of chatting [2, 7, 11, 18–21]. In other words, this study proves that engaging in chat with either a chatbot or a human chat partner can help EFL students to improve their grammar skills. Considering the fact that learning grammar is an absolute necessity in studying EFL [8], this study suggests that the experience of chat, whether with a chatbot or human chat partner, can make positive effects on Korean EFL students language learning, enhancing their grammar skills.

3.2 Group Differences in English Grammar Skills

In order to compare their English grammar skills between the two groups, an independent t-test was carried out. A difference in the mean scores on the post-test between the two groups was compared to determine which group improved their grammar skills the most. Table 3 below demonstrates the descriptive statistics as well as independent t-test results.

<table>
<thead>
<tr>
<th>Chatbot Group (n = 36)</th>
<th>Human Group (n = 34)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M 19.75</td>
<td>SD 4.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M 17.32</td>
<td>SD 5.56</td>
<td>2.035</td>
<td>0.046</td>
</tr>
</tbody>
</table>

As can be seen from Table 1 in the previous section, no significant difference between the groups was witnessed (p > .05). That means all participants in the current study were homogeneous before the treatment. Table 3 above, however, shows a statistically significant difference in mean scores on the post-test between the chatbot group and the human group (t = 2.035, p = .046).

The difference in the mean scores between the two groups turned out to be statistically significant, indicating that participants who engaged in chat with a chatbot improved their grammar skills more than did those who had a chat with a human chat partner. In terms of the chatbot group, participants improved their grammar skills with the mean score of 19.75 on the post-test. In the case of the human group, their mean score on the post-test was 17.32.

To sum up, findings of the present study the chatbot group improved their English grammar skills more than the human group. That is, engaging in chat with a chatbot is more effective in increasing EFL grammar skills compared to having a chat with a human chat partner. This can be explained by the previous research [7], indicating that EFL students participate in chat more actively to overcome communication breakdown that occurs when engaging in chat with a chatbot.

According to Kim [11], students’ diverse and numerous errors can result in communication failure. In particular, Kim [7] found that chatting with a chatbot can cause more communication breakdown than chatting with a human chat partner because of the errors that students make. Participants in her study made grammar mistakes continuously when they were engaging in chat with a chatbot, Indigo. Particularly, they had a problem with asking questions which made the chatbot unable to answer them properly.

Humans can infer others’ actual intention by considering the dialogue context with the utterance. However, unlike humans, chatbots
cannot rely on conversational context especially when its users commit unexpected errors [22]. For chatbots, it is not easy to deal with ambiguity with these errors [23]. As a result, more frequent communication breakdowns can occur when engaging in chat with a chatbot.

When there is a breakdown in communication, students interact more actively to overcome this communication failure [24]. Kim [11] demonstrated that the communication problem with a chatbot can be solved by correcting grammatical errors. In light of this, participants in the chatbot group in the current study were able to improve their grammar skills by trying to deliver exactly what they have in mind and to make the chatbot, Replika, understand clearly to overcome any communication problem.

Considering this, findings of the current study support the previous studies [7,11,24], suggesting that students can benefit from having a chat with a chatbot. The scholars particularly claimed that foreign language students can improve their grammar skills via chatting with a chatbot. Given that English grammar skills are necessary and essential in Korean college contexts [9], engaging EFL students in chat with a chatbot seems to have more advantages over engaging them with a human chat partner.

4. Conclusion

Teaching grammar is beneficial for EFL students [25] and Korean EFL students have been taught with a focus on grammatical competence [8]. As English grammar skills are necessary and essential even in Korean college contexts [9], it became essential to measure their grammatical competence and more emphasis has been placed on the grammar test [14].

However, there is a dearth of previous research on the use of chatbots in relation with EFL grammar skills. Even though artificial intelligence chatbots continue to advance and their studies have been conducted continuously [1–7,11,18], there are few studies investigating the effects of the use of chatbots on improving English grammar skills. This study, therefore, aimed to determine whether chatbots are effective for improving English grammar skills.

Main findings are as follows: First, both the chatbot group and the human group showed significant changes in mean scores between the pre- and post-tests. This indicates that engaging in chat with either a chatbot or a human chat partner is beneficial for increasing Korean EFL students' grammar skills. Second, the difference in the mean scores on the post-test between the two groups turned out to be statistically significant, suggesting that the chatbot group improved their English grammar skills more than the human group did. In other words, participants who engaged in chat with a chatbot improved their English grammar skills more than did those who had a chat with a human chat partner.

With the traditional teaching methods in most foreign language classrooms, students' affective filters are easily increased, making language learning more difficult with negative learning experiences [26]. However, the present study shows how chatbots can play a positive role in foreign language learning by lowering EFL students' affective filters. By simulating human conversation, Replika, the chatbot used in the current study, allowed the human–machine interaction, and through this interaction, the chatbot became closer and more similar to the students over time by learning from them. Replika also understood the students better as time goes by, connecting their social media accounts. These features of the chatbot might have lowered the students' affective filters.

According to Brooks [26], students who have a low affective filter show high motivation, self-confidence, and low anxiety levels while those who possess a high affective filter display
low motivation, low self-esteem and high anxiety levels. Therefore, it is important to lower the students’ affective filter for their successful language learning. Findings of the current study can support the previous chatbot research [2,7,11,18], suggesting that chatbots for EFL learning can lead to EFL students’ academic achievement by inhibiting affective filters such as reducing the students’ fear, anxiety, and stress.

Particularly, EFL students can benefit from chatting with a chatbot in terms of improving their grammar skills, considering that the students interact actively and correct grammatical errors to overcome their communication breakdown with a chatbot [11]. Since chatbots cannot rely on conversational context nor deal with the errors [22,24], the students will deliver exactly what they have in mind to make them understand clearly. This can lead to the improvement of EFL grammar skills.

From this point of view, artificial intelligence chatbots should be integrated into EFL classrooms. Limitations and suggestions for the further research, however, should be considered. First, this study was limited to Korean EFL students taking a General English course at one university in Korea. In addition, the majority of the participants were freshmen students. Therefore, learner variables should be considered carefully for the future. Furthermore, no follow-up was conducted in the current study which makes it difficult to know whether the treatment effects continues. Last but not least, given that the present study was conducted with the technology, familiarity with the technology or any possible effects of technical issues should be taken into consideration.

REFERENCES


