# A Study on Chatbots for Developing Korean College Students' English Listening and Reading Skills

**Na-Young Kim**

**Assistant Professor, Department of General Education, Sehan University**

# 국내 대학생의 영어 듣기 및 읽기 능력 향상을 위한 챗봇 활용 연구

김나영

세한대학교 교양학부 조교수

**Abstract**  In an effort to investigate the effects of chatbots on English listening and reading skills, 46 college students participated in the current study. Participants consisted of first-year students who enrolled in an English class at a university in South Korea. They were randomly divided into two groups: one experimental group (n=24) and one control group (n=22). During 16 weeks, the experimental group engaged in chats with a chatbot, named Elbot, while the control group did not. There were pre- and post-tests to confirm the effects of the chatbot usage. Major findings are as follows: First, participants in both groups significantly improved  listening and reading skills. On the post-listening test, however, the experimental group showed more improvements. Their listening proficiency level improved from intermediate to advanced level after engaging in chat with the chatbot. Limitations and implications for theory and practice are discussed at the end.

**Key Words :** Chatbots, Korean college students, English learning, English listening skills, English reading skills

요 약  본 연구는 챗봇의 활용이 국내 대학생의 영어 듣기 및 읽기 학습에 미치는 영향을 조사한 것으로, 실험 참가자의 영어 듣기 및 읽기 능력이 챗봇과의 채팅을 통해 실제로 상승하는지에 대한 여부를 알아보는 데 그 목적이 있다. 본 연구를 위해 총 46명의 참가자를 실험그룹과 통제그룹으로 나누어 16주 동안 실험을 진행하였으며, 실험 시작 전과 실험 종료 후 사전 사후 평가를 실시하여 챗봇 활용의 효과를 파악하였다. 본 연구의 주요 결과 및 시사점은 다음과 같다. 사전 사후 평가 결과 실험그룹과 통제그룹 모두에서 영어 듣기 및 읽기 능력이 유의미하게 상승한 것으로 나타났다. 특히 영어 듣기 능력과 관련하여 실험그룹이 통제그룹보다 사후 평가에서 더 많은 상승폭을 보임으로써 듣기 능력 향상에 대한 챗봇 활용의 효과를 증명하였다. 본 연구는 4차 산업혁명 시대에 따라 영어 학습을 위한 챗봇 활용에 대한 시사점을 제시하는데 그 의의를 갖는다고 볼 수 있다.

주제어 : 챗봇, 한국 대학생, 영어 학습, 영어 듣기 능력, 영어 읽기 능력

## 1. Introduction

With today's advanced technology, global communication has become more and more important; thus, the importance of English has increased. According to Khan and Shrivastava [1], proficiency in English as a medium of communication is now highly in demand for personal and professional growth.

Particularly, in Korean English as a Foreign Language (EFL) environment, English proficiency is mandatory to enter a prestigious school, to get a high-paid job, or to be promoted at work [2]. In order to prove their English proficiency, most students in Korea are required to submit their English test scores. Official English proficiency tests such as TOEIC and TOEFL are high stakes tests in EFL contexts and have a crucial effect on the students' futures [3]. For these reasons, students of English in Korea take extra English lessons, attend English academy classes, and study English abroad to prepare for and take English proficiency tests.

Korean EFL students invest an enormous amount of time and money studying English. Official English teaching in Korea begins in elementary school. From elementary to high school, the official time they spend on English learning at school is approximately 730 hours [4]. Along with official English classes at schools, most students in Korea participate in private tutoring, take on-line or off-line English classes, and use self-help materials. Some of them even go to English-speaking countries such as America or Britain to study English for a certain period of time. The Korean government, according to Jeon [2], also supports English education by creating English villages and cities as intensive language learning camps to learn English skills, hiring native speakers of English as teaching staff.

Nevertheless, Korean EFL students are still not good at English [5]. Despite all these efforts mentioned above to improve English language skills, there is no conspicuous improvement in their English proficiency. In particular, when measured by well recognized proficiency tests such as TOEFL, Korean students' English proficiency remains almost at the bottom. According to Kang [6], they obtained 18 points on average out of 30 on the speaking test. This is lower than 19.3 points, the world average. Korean students ranked 136th out of 161 nations in English speaking. Their score of 19 on the listening test was also lower compared to the world average score of 19.5. Their writing score showed similar results, scoring an average of 20 points out of 30 compared to 20.5, the world average. Only on the reading test did Korean students score above the world average. They obtained an average of 20 points out of 30, while the world average was 19.4 points.

According to Graddol [8], there are more than 750 million EFL students in the world. However, many of them have relatively few opportunities to meet English native speakers in their countries and little chance to practice English. A lot of scholars have pointed out that EFL students have   due to the lack of chances [9]. Fryer and Carpenter [7] suggested that chatbots are a potentially valuable resource for students in this setting. Chatbots, designed to carry on a human-like conversation with real human beings, have value as learning tools, acting like native speakers of English [9]. Another strong point of chatbots is their convenience. They are readily available to students with online access, at school or at home. Furthermore, they are generally free of charge. Shawar and Atwell [10] concluded that chatbots are useful alternatives to traditional EFL classrooms.

There have been some chatbot-related studies in EFL fields, suggesting that they are beneficial for English learning [11]. Since chatbots provide both textual and auditory input, students can practice both listening and reading skills [7]. Curiously however, most chatbot studies have centered around English productive skills such as speaking and writing, while only a few studies investigated receptive skills such as listening and reading [12]. Kim [9], for example, examined the effects of the use of chatbots for developing speaking skills. Her recent chatbot study [23] also focused on writing skills. Given the dearth of empirical studies examining the effects of chatbots on EFL receptive skills, the present study aims to report the effectiveness of the use of chatbots in relation to Korean college students′ English listening and reading skills. The research question is as follows:

What are the effects of chatbots on Korean college students' English listening and reading skills?

## 2. Methodology

### 2.1 Participants

Participants consisted of 46 college students who enrolled in a 16-week English class at a university in Korea. They were mostly freshmen students taking the same class named English 1. Their age ranged from 20 to 22. The class focused on general English skills. For graduation, this English course was a prerequisite at the university. Over the experimental period, the students took their English lessons for two and a half hours a week from the instructor who had 8 years of experience in teaching.

Table 1. Participants

|  | Experimental (n=24) | | Control (n=22) | | $t$ | $p$ |
|---|---|---|---|---|---|---|
|  | M | SD | M | SD | | |
| Listening | 359.29 | 61.75 | 325.95 | 60.88 | 1.841 | .072 |
| Reading | 343.88 | 53.54 | 333.36 | 54.00 | .662 | .511 |
| Total | 703.13 | 96.78 | 659.09 | 93.84 | 1.564 | .125 |

The participants' proficiency level was considered as intermediate based on their TOEIC pretest mean scores [13]. They obtained an average of 343.35 out of 495 on the pre-listening test and 338.85 out of 495 on the pre-reading test. They were randomly classified into two groups: one experimental group (n = 24) and one control group (n = 22). As Table 1 shows, there were no significant group differences between the experimental and control groups (p > .05), indicating that two groups were homogeneous at the beginning of the research. All participants had no experience with chatbots before.

### 2.2 Instruments

Participants in the experimental group engaged in

ten chat sessions in total for the experimental period and talked about their everyday life. There are various types of chatbots such as *Indigo* and *Cleverbot*. However, the chatbot called *Elbot* was employed for the present study. *Elbot* was chosen because it was widely used worldwide and also the winner of the Loebner Prize 2008 for Artificial Intelligence [14]. Developed by Fred Roberts, *Elbot* can carry on a human-like conversation with its users using natural language interaction (NLI) about a variety of topics. According to Floridi, Taddeo, and Turilli [15], *Elbot* deserves the prize considering that it might be indistinguishable from real humans.

Participants downloaded the *Elbot* application onto their smartphone or computer, registered their account, and conversed with the chatbot. This tireless chatbot acts as a native speaker to help its users practice speaking English. When a user speaks or types, *Elbot* understands and gives the user a human-like response. Users can hear the response as well as see the text on the screen. As an award-winning chatbot, its native-like responses are not only appropriate, but also correct. If a user speaks with a foreign accent and *Elbot* fails to understand, the user can type in questions and responses. Even though the user makes spelling or grammatical mistakes, this intelligent chatbot can recognize and correct them. It is a great way to practice English [14].

As pre- and post-tests, all participants were requested to take the TOEIC Listening and Reading Tests[1]. Given that most students in Korea are required to submit their official English test score to prove their English proficiency [3], TOEIC tests were chosen to meet the participants' practical needs.

Reflecting real-world tasks, the TOEIC listening and reading test provides a common standard of measurement for language skills of EFL students. It is a paper-and-pencil and multiple-choice assessment.

---

1) Educational testing service (ETS) offers a free practice test on its TOEIC listening and reading web site. https://www.ets.org/toeic/listening_reading/test_preparation/sample_questions

There are two timed sections with 100 questions each: 45 minutes for listening and 75 minutes for reading. For the listening test, test takers listen to short conversations or a variety of questions recorded in English. Based on what they have heard, they respond to the questions. For the reading test, test takers read various materials and answer questions at their own pace. By the number of correct answers, test scores are determined. They are then converted to a scaled score. These scores indicate how well the test takers can do in real-life situations in which they are asked to listen and read in English.

## 2.3 Procedures

The purpose of the current study was to examine the effects of chatbots on EFL receptive skills: listening and reading skills. In order to investigate the effectiveness of the use of chatbots, 46 participants took part in this study. All participants were requested to take the TOEIC Listening and Reading Tests as pre- and post-tests to investigate the changes after engaging in chatbot assisted language learning.

During the experiment, all participants took their English lessons for two and a half hours a week. For the experiment, participants were divided into two groups: One experimental group and one control group. Over the 16-week experimental period, those in the experimental group carried out an everyday conversation with the chatbot, *Elbot*, through a chat interface using both textual or auditory methods. They downloaded the *Elbot* application onto their smartphone or computer, registered their account, and conversed with the chatbot. As a homework assignment, participants in this group were required to engage in 20 chat sessions in total. Chat topics varied from business to environmental issues. Each chat session lasted more than 10 minutes. Although both the experimental and the control group received the formal instruction during the regular English teaching time period, the control group received no treatment.

## 2.4 Data Analysis

In order to examine the effectiveness of chatbots, carried out was quantitative analysis. Data from the current study were analyzed using SPSS 18.0 software. About the quantitative analysis, descriptive statistics including means and standard deviations were calculated. To compare the pre- and post-test scores, paired samples t-tests were run. To compare the mean differences between the experimental group and the control group, an independent t-test was administered. P values were set at .05.

# 3. Results

## 3.1 Changes in Listening and Reading Skills

To examine the changes in participants' listening and reading skills, paired samples t-tests were run between the pre-test and the post-test. Descriptive statistics and paired samples t-test results are present in Table 2.

Table 2. Changes in Listening and Reading Skills

|  |  | Pre-test (n=46) | | Post-test (n=46) | | t | p |
|---|---|---|---|---|---|---|---|
|  |  | M | SD | M | SD | | |
| Experi-mental | Listening | 359.29 | 61.75 | 419.54 | 44.26 | 7.881 | .000* |
|  | Reading | 343.88 | 53.54 | 403.67 | 68.18 | 5.601 | 000* |
|  | Total | 703.13 | 96.78 | 823.00 | 90.77 | 8.194 | 000* |
| Control | Listening | 325.95 | 60.88 | 369.18 | 83.32 | 3.440 | 001* |
|  | Reading | 333.36 | 54.00 | 393.59 | 65.97 | 6.952 | 000* |
|  | Total | 659.09 | 93.84 | 762.59 | 115.47 | 6.654 | 000* |

Interestingly, both two groups showed significant changes in mean scores between the pre-test and the post-test. For the experimental group, there were significant mean changes in both listening skills (t = 7.881, p = .000) and reading skills (t = 5.601, p = .000). To be specific, regarding the listening skills, the mean score was 359.29 on the pre-test, while 419.54 on the post-test. As for the reading skills, the mean score was 343.88 on the pre-test, whereas 403.67 on the post-test,

suggesting that participants improved both listening and reading skills after engaging in chats with the chatbot.

For the control group, significant changes were also observed in listening skills (t = 3.440, p = .001) as well as reading skills (t = 6.952, p = .000). Specifically, the mean score regarding listening skills improved from 325.95 on the pre-test to 369.18 on the post-test. In terms of reading skills, the mean score also increased from 333.36 on the pre-test to 393.59 on the post-test, indicating that the control group also improved their listening and reading skills over time. This can be attributed to the general practice effect by retaking the same test [16]. The control group might be able to improve numerically on the post-test due to the equivalent and marginal test-retest practice effect.

Nonetheless, it is important to note that the experimental group did rank at a higher level on the post-listening test after the intervention, suggesting that they advanced to the next proficiency level. According to Liao [13], participants' listening proficiency can be divided into three levels based on TOEIC listening test scores: Scores around 200 (novice), 300 (intermediate), and 400 (advanced). The mean score of 359.29 on the pre-listening test reveals that the participants in the experimental group were intermediate-level students at the beginning of the study. However, the mean score of 419.54 on the post-listening test indicates that their listening proficiency level reached the advanced level, as a result of engaging in chats with a chatbot. On the contrary, control group participants' listening proficiency level did not change over time, with the mean score of 325.95 on the pre-test to 369.18 on the post-test, indicating that they were still intermediate-level students.

## 3.2 Group Differences in Listening and Reading Skills

In order to explore the effects of chatbots between the two groups, independent t-tests were performed on the post-listening and reading tests. Table 3 represents the descriptive statistics and t-test results. At the beginning of the research, as seen from Table 1 in the previous section, there were no significant differences between the experimental and control groups (p > .05). However, a statistical significance was found on the post-listening test (t = 2.591, p =. 013), as shown in Table 3.

Table 3. Group Differences in Listening and Reading Skills

| | Experimental (n=24) | | Control (n=22) | | t | p |
|---|---|---|---|---|---|---|
| | M | SD | M | SD | | |
| Listening | 419.54 | 44.26 | 369.18 | 83.32 | 2.591 | .013* |
| Reading | 403.67 | 68.18 | 393.59 | 65.97 | .508 | .614 |
| Total | 823.00 | 90.77 | 762.59 | 115.47 | 1.981 | .054 |

As seen from Table 3, the post-test results reveal that the experimental group significantly performed better than the control group, regarding the listening skills, with mean scores of 419.54 (experimental group) and 369.18 (control group). In terms of reading skills, the mean score on the post-reading test for the experimental group was 403.67, while 393.59 for the control group. However, no significant differences were observed.

According to Shah, Warwick, Vallverdú, and Wu [14], chatbots can be a good conversationalist with a successful artificial dialogue system. Designed to interact with human users, they can select appropriate and corresponding response [7]. Particularly, Lee et al., [17] and Movellan, Eckhardt, Virnes, and Rodriguez [18] reported that chatbots in English classroom play a role of native speakers carrying on conversations in natural language. Johnson [19] also claimed that they are beneficial when implemented in education.

Results of this study are in line with previous studies suggesting the beneficial effects of chatbots for developing language skills. Han [20] emphasized this current trend of using robot-like characters to assist language learning and teaching using the term RALL, known as robot-assisted language learning. Papert [21]

also claimed that this type of interaction improves language teaching effectiveness. In their empirical study, Hong, Huang, Hsu, and Shen [12] reported the obvious progress in Taiwan EFL students' two language skills, listening and reading skills, considering chatbots as beneficial tools in education improving teaching effectiveness. They suggested that more and more robotic applications are developing activities that can be beneficial for teaching and learning English as a foreign language.

It is notable that the experimental group significantly outperformed the control group in terms of listening skills. According to Kim [9], the embedded speech recognition technology provides students with opportunities to experience a voice chat with the chatbot and this creates value and enhances the participant experience via sending and receiving both textual and auditory messages. Zakos and Capper [22] also suggested that chatbot software programs are helpful for developing conversational language skills including listening skills. They claimed that language skills can be acquired by conversationally interacting with this chatbot in a natural and humanized way. However, in the current study, no group difference was observed in terms of reading skills. One possible reason is that the TOEIC test measures a different type of reading skills than the chatbot provides. While the TOEIC listening test is related to face-to-face communication, meetings, and telephone conversations, the TOEIC reading test is in association with skills necessary for email, reports, and other forms of business correspondence.

## 4. Conclusion

English communicative competence has been emphasized in Korea [2]. Since the TOEIC test was designed to evaluate a test taker's communicative competence [13], EFL students have taken the TOEIC test as an official test to prove their proficiency. Given

that Korean college students' main goal of studying English is to perform well on standardized tests like the TOEIC test [3], chatbots can be beneficial for their communicative language development improving learning effectiveness.

Considering that EFL students including Korean students of English have few opportunities to meet English native speakers in their countries and little chance to practice English [9], chatbots, acting like native speakers of English, seem to be valuable as learning tools, providing authentic input. In EFL settings where there is lack of English input, the advances in chatbot technologies can provide students with authentic English language input. By doing so, as Fryer and Carpenter [7] put, chatbots can offer opportunities to learn listening and reading.

Findings of the current study, from this point of view, provide the empirical evidence of educational effects of chatbots on receptive skills in Korean EFL settings giving authentic input. The present study evaluated the effects of chatbots on Korean college students' listening and reading skills. Major findings from this study are as follows: First, participants in both experimental and control groups significantly improved both listening and reading skills. In particular, it turned out that the listening proficiency level of the participants in the experimental group improved from intermediate level to advanced level after engaging in chats with a chatbot. Group comparison results also revealed that the experimental group significantly outperformed the control group on the listening test.

Results of this study are in line with a lot of previous studies have suggesting the beneficial effects of chatbots for developing language skills [14, 18, 21]. However, most of them have focused on productive skills such as speaking and writing. Although Fryer and Carpenter [7] suggested that chatbots provide both textual and auditory input, encouraging foreign language students can practice both listening and reading skills, there is lack of studies investigating these receptive skills [12]. As more practical studies on

chatbots have been called for [9], this paper provides evidence of educational effectiveness of chatbots for Korean EFL students. Particularly, this research demonstrated that chatbots provide a great opportunity to foster Korean college students' English language learning, especially their receptive English language skills such as listening and reading skills.

Some pedagogical implications can be drawn from the findings of the study. Given that EFL students have few opportunities to meet English native speakers and little chance to communicate in English, EFL teachers can employ chatbots as a potentially valuable resource for students in this setting. Designed to hold a human-like conversation, they can play a role as a native speaker of English, providing authentic language input. In addition, considering that chatbot assisted language learning not only improves English productive skills such as speaking and writing [12], but also increases receptive skills such as listening and reading, chatbots can be integrated into EFL classroom to develop the four major language skills.

Limitations and suggestions for further research are also given. First of all, learner variables such as age, nationality, and language proficiency should be carefully considered because participants in the current study were limited to intermediate-level Korean college students of English. Furthermore, the small number of participants in the experimental group (n = 24) was another limitation of this study. Because different results can be obtained with a larger number of participants in other settings, caution should be made when generalizing the result of the study. Moreover, although a pre-test post-test control group design was implemented in the current study to determine the effectiveness of chatbots, findings of the study would be stronger with an additional comparative group. The improvement in English skills for the experimental group might be because of the increased exposure to English language. In addition, since the Hawthorne effect was an unavoidable bias in the current study, it can be difficult to conclude that the chatbot was the only parameter to influence the participants' language performance and it should be taken into account when analyzing the results. Future study can include another tangible instructional tool similar to chatbots to investigate whether chatbots still have significant effects. Lastly, with technology literacy, any possible effects of technical issues and familiarity with the technology should be taken into consideration.

# REFERENCES

[1] J. Khan & S. Shrivastava. (2016). English language as a medium of communication in India. *New Man International Journal of Multidisciplinary Studies, 3(10),* 12-15.

[2] J. Jeon. (2010). Issues for English tests and assessments: A view from Korea. In Y. I. Moon, & B. Spolsky (Eds.) *Language assessment in Asia: local, regional or global* (pp. 55-82). Seoul: Asia TEFL.

[3] T. Y. Kim. (2009). The dynamics of L2 self and L2 learning motivation: A qualitative case study of Korean ESL students. *English Teaching, 64(3),* 49-70.

[4] J. Jeon & J. Paek. (2008). *Developing framework to evaluate the process and performance of government English language policies.* Seoul: Korean Ministry of Education, Science, and Technology.

[5] J. I. Han & N. Y. Kim. (2016). The effects of post-task CMC activities and task types on Korean EFL learners' oral performance. *STEM Journal, 17(2),* 109-135.

[6] S. W. Kang. (2009, April 1). *Koreans ranked bottom in English proficiency test.* The Korea Times.

[7] L. Fryer & R. Carpenter. (2006). Bots as language learning tools. *Language Learning & Technology, 10(3),* 8-14.

[8] D. Graddol. (1997). *The future of English?* London: British Council.

[9] N. Y. Kim. (2017). Effects of different voice-chat conditions on EFL learners' topic negotiation according to proficiency levels. *Modern English Education, 18(1),* 49-74.

[10] . A. Shawar & E. Atwell. (2007). Chatbots: Are they really useful? *LDV Forum, 22(1),* 29-49.

[11] Y. F. Wang & S. Petrina. (2013). Using learning analytics to understand the design of an intelligent language tutor? Chatbot Lucy. *International Journal of*

*Advanced Computer Science and Applications, 4(11),* 124-131.

[12] Z. W. Hong, Y. M. Huang, M. Hsu & W. W. Shen. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *Educational Technology & Society, 19(1),* 337-349.

[13] C. W. Liao. (2010). *TOEIC listening and reading test scale anchoring study.* Princeton, NJ: Educational Testing Service.

[14] H. Shah, K. Warwick, J. Vallverdu & D. Wu. (2016). Can machines talk? Comparison of Eliza with modern dialogue systems. *Computers in Human Behavior, 58,* 278-295.

[15] L. Floridi, M. Taddeo & M. Turilli. (2009). Turing's imitation game: Still an impossible challenge for all machines and some judges? An evaluation of the 2008 Loebner Contest. *Minds and Machines, 19(1),* 145-150.

[16] N. Y. Kim. (2016). *Effects of different voice-chat conditions on Korean EFL learners' speaking ability, oral interaction, and affective factors.* Unpublished doctoral dissertation, Ewha Womans University, Korea.

[17] S. Lee, H. Noh, J. Lee, K. Lee, G. Gary, S. Sagong & M. Kim. (2011). On the effectiveness of robot-assisted language learning. *ReCALL, 23(1),* 25-58.

[18] J. R. Movellan, M. Eckhardt, M. Virnes & A. Rodriguez. (2009). Sociable robot improves toddler vocabulary skills. In M. Scheutz, F. Michaud, P. J. Hinds, & B. Scassellati (Eds.), *Proceedings of the 4th ACM/IEEE Human Robot Interaction* (pp. 307-308). New York: ACM.

[19] J. Johnson. (2003). Children, robotics and education. *Artificial Life and Robotics, 7(1-2),* 16-21.

[20] J. Han. (2012). Emerging technologies? Robot assisted language learning. *Language Learning & Technology, 16(3),* 1-9.

[21] S. Papert. (1993). *Mindstorms: Children, computers, and powerful ideas.* New York: Basic Books.

[22] J. Zakos & L. Capper. (2008). CLIVE - An artificially intelligent chat robot for conversational language practice. In J. Darzentas, G. Vouros, S. Vosinakis, & A. Arnellos (Eds.) *Proceedings of 5th Hellenic Conference on Artificial Intelligence* (pp. 437-442). Berlin: Springer.

[23] N. Y. Kim. (2018). Effect of text chat on EFL writing fluency, accuracy, and complexity by interlocutors. *Foreign Languages Education, 25(1),* 27-54.

김 나 영(Kim, Na Young)                    [정회원]

▪2017년 2월 : 이화여자대학교 영어교육학과 (문학박사)
▪2018년 3월 ~ 현재 : 세한대학교 교양학부 조교수
▪관심분야 : 영어 교육, 멀티미디어를 활용한 영어 교육, 기업 교육
▪E-Mail : nykim@sehan.ac.kr