



Effects of AI Chatbots on EFL Students' Communication Skills*

Hea-Suk Kim (Seoul Women's University) Yoonjung Cha (Hanshin University)
Na Young Kim (Sehan University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: June 15, 2021
Revised: August 20, 2021
Accepted: August 25, 2021

Hea-Suk Kim (1st author)
Professor, Department of General
Education, Seoul Women's
University
shskim@swu.ac.kr

Yoonjung Cha (corresponding author)
Professor, Peace and Liberal Arts
College, Hanshin University
yjcha@hs.ac.kr

Na-Young Kim (corresponding author)
Professor, Department of General
Education, Sehan University
nykim@sehan.ac.kr

ABSTRACT

Kim, Hea-Suk, Yoonjung Cha and Na Young Kim. 2021. Effects of AI chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics* 21, 712-734.

The purpose of this study was to examine how AI chatbots affect students' speaking skills and how they motivate and shape students' speaking experiences in the EFL classroom. Forty-nine university students who enrolled in a general English course took part in the study. They were divided into two different proficiency levels. The pre- and post-test design compared students' improvement in English speaking within and between the two levels. The test results indicated that both groups showed significant improvement in the two speaking tasks including read a text aloud and respond to questions. In terms of students' proficiency between the two levels, there was no significant difference in pronunciation. However, a notable difference in intonation and stress in the read a text aloud task was detected. Also, in terms of the second task: respond to questions, statistical differences were found in all of them. Regarding the fluency between the two levels, a statistical difference was likewise found. In addition, we explored the findings of a questionnaire illustrating students' perspectives toward using AI chatbots in their English classes. Finally, this study discusses how AI chatbots can help with language learning and how they can be used in EFL settings in the future.

KEYWORDS

EFL, communication skill, fluency, pronunciation, intonation & stress, WPS

* This paper was supported by Sehan University Research Fund in 2021.

1. Introduction

As a global lingua franca, the English language is used in almost every domain of communication around the world (Khamkhien 2011). At all stages of the education system in some countries including South Korea, English is the most widely taught foreign language (Khamkhien 2011), and speaking English, in particular, is seen as the most important goal of learning English (Ur 1996). To be sure, speaking itself is important in everyone's life. It is an essential means of communication and a channel through which people interact and share their ideas, thoughts, and feelings. As international communication in modern society increases, fluent speaking proficiency in English is increasingly expected from every language learner (Lintunen, Peltonen and Webb 2015).

Mastering foreign language speaking requires great effort and the skillful integration of certain language aspects like appropriate vocabulary, correct grammar and sentence structure, and clear pronunciation (Raba 2017). Fluency is also understood to be an important benchmark of foreign language speaking, so it is no surprise that it stands high among the curricular goals and is used in learner assessment (Lintunen et al. 2015). In fluency studies, measures of speed (i.e., how fast a speaker talks) are commonly used, (Skehan 2003) usually measured in terms of words or syllables per second or minute. Levelt (1989) noted that fluent speech is produced at a rate of two to three words per second. To master a language, what is necessary is to speak it fast (Hughes 2003).

However, a mismatch exists between industry requirements and university graduates' competencies regarding effective foreign language speaking skills (Aclan and Aziz 2015). Most students have difficulty in oral communication in a foreign language (Kim 2017). This prevents foreign language students from expressing themselves in an appropriate and professional way. In particular among EFL students, speaking provokes the most anxiety and nervousness (among the four language skills) (Kessler 2010). This can be attributed to rote learning (Aclan and Aziz 2015), minimal encouragement to practice speaking (Ho 2003), and the lack of opportunities to practice communication skills both inside and outside the classroom (Chiu, Liou and Yeh 2007).

According to Kim (2016), foreign language students should have maximum opportunities for communication via involvement with the conversational usages similar to everyday conversation. In this regard, the lack of speaking practice with a real audience in an authentic learning environment is one of the major limitations in EFL fields. To remedy these problems in the EFL curriculum, teachers have done their best to create opportunities to practice English speaking for their students. Particularly, language professionals have recommended the use of educational technology to increase students' output opportunities and to improve their speaking proficiency. Indeed, with significant advances in multimedia technology, educational innovations have taken hold in EFL contexts (Kim, Cha and Kim 2019).

The development of a wide range of communication tools has allowed students to have natural and genuine conversations with authentic audiences from all over the world. According to Kim (2017), this is particularly beneficial for promoting communication skills. Among them, artificial intelligence (AI) chatbots have drawn considerable attention in EFL fields. With the purpose of carrying on a conversation with human beings, the chatbot has provided foreign language students with a means for language practice (Haristiani 2019). Considering the limited opportunities to practice the target language, it is indeed a valuable resource for EFL students (Fryer and Carpenter 2006). With the advancement of AI technology, educators and educational practitioners have continuously investigated whether this advanced technology is indeed effective for foreign language teaching and learning (Kim, Kim and Cha 2021).

Nevertheless, empirical studies on AI chatbots are still scarce in EFL fields (Kim, Cha and Kim 2020), and many unanswered questions exist regarding the implementation of the chatbot (Yanguas 2010). Furthermore, previous research has had mixed results with respect to their effectiveness according to proficiency levels. For

example, according to Fryer and Carpenter (2006), chatbots are most useful to accomplished or higher-level foreign language students, but they are not good for beginner-level students who still have to master the basics. This is because the chatbots were originally designed to interact with native speakers, and thus, they are generally incapable of, or are poor at noticing, pronunciation, spelling, or grammar mistakes. However, Kim (2016) reported that all students at different proficiency levels – novice, intermediate, and advanced – can benefit from learning with an AI chatbot.

Taking all of this into consideration, there is a need to confirm the effects of AI chatbots on EFL learning according to language proficiency. In particular, given that English speaking skills are so important (Derakhshan, Tahery and Mirarab 2015), and that AI chatbots effectively contribute to the improvement of communicative competence (Kim 2017), this study explores their impacts on oral communication skills among EFL students at different levels of proficiency. Therefore, the current study aims to fill the gap by asking the following research questions:

- a. Is there a significant difference in students' communication skills between low and intermediate proficiency levels after experiencing AI communication?
- b. What are students' perceptions toward AI communication to improve English speaking?

2. Literature Review

2.1 Oral Communication Skills in EFL Settings

The focus for language teaching has moved from a one-way, single person, and staged lesson perspective to a more authentic, communicative, and interactive perspective (Kim 2017). Communicative language teaching (CLT) focuses on interaction, suggesting that comprehension of input as well as learner output are both necessary for foreign language acquisition. From this theoretical perspective, language is best learned and taught via communication (Cheon 2003). Warschauer (2001) added that successful language learning comes through communicating real meaning. In CLT, fluency has also been emphasized to meaningfully engage learners in language use. Today, the main apprehension of foreign language learners is whether they can use their target language independently and fluently in communicative situations (Thamarana 2015). According to Hughes (2003), CLT leads to the production of fluent learners.

In order to increase students' communication skills, CLT has been adopted in EFL classrooms (Khamkhien 2011). With the growing popularity of CLT, communicative competence has been emphasized and taken as a goal in EFL settings. According to Willis and Willis (2001), a foreign language classroom is where students use their target language for a communicative purpose. Nonetheless, CLT often fails to create enough opportunities for genuine interaction between students in the classroom (Khamkhien 2010). Interactive learning in EFL classrooms such as working in pairs or a small group cannot happen so easily because various factors such as a lack of class time or a large number of students in one class hamper this (Kim 2017). According to Chern (2010), although English is mandatory in EFL college curricula, classes meet for two hours per week on average and are usually large. Furthermore, the curriculum primarily focuses on vocabulary, grammar, or textbook reading and provides few opportunities to practice speaking (Huang 2015). This leads to limited communication opportunities for EFL students to put their knowledge they have learned into practice in an authentic environment (Ho 2003). Due to

such limited interactions, EFL students' communicative skills consequently fail to develop, and students continue to experience discomfort when communicating in English (Kim 2018).

However, in a classroom environment, teachers can effectively help students to enhance their communicative competence by encouraging them to be involved in a variety of strategies, activities, and tasks for communication. For the sake of improving EFL students' communication skills, there have been some empirical studies on this. Raba (2017), for example, investigated the effects of the think-pair-share (TPS) strategy on EFL students' oral communication skills. According to him, as one of the interactive skills, speaking skills can be enhanced by using interactive strategies like TPS. Since its main element is peer interaction, students are encouraged to interact with each other and are given opportunities to think, listen to, share, and reflect on their ideas as well as their peers'. Consequently, this leads to active engagement. After interviewing teachers and observing students' interaction in EFL classrooms, the researcher acknowledged that TPS plays a positive role in enhancing students' oral communicative skills. He also found its positive influence on increasing students' motivation to learn and in creating a cooperative learning environment. In conclusion, increasing the number of TPS activities is recommended in EFL classrooms.

In order to promote oral communication skills in EFL contexts, Zhang (2009) highlighted the appropriate selection of tasks. He suggested integrating oral communication into other lessons to improve foreign language students' communicative competence. For example, speaking can be added to reading and writing lessons. In doing this, students can receive essential practice in oral communication. According to him, this integration has many advantages since it adds variety to the lessons, encompasses students' different strengths, and generates interactive possibilities with a focus on both receptive and productive skills. Interesting topics accompanied with reading and writing lessons also lend themselves well to speaking tasks. Consequently, students' language skills can be facilitated with the topics to discuss, testing their language hypothesis. In particular, integrating speaking and reading skills can deepen the students' understanding of the reading material and allow them to apply the information they have read to authentic speaking practice. This results in an improvement in communicative competence (Zhang 2009).

Language proficiency does not depend on linguistic knowledge. Compared to the functional ability of communication, language knowledge is secondary. The only way to acquire complete communicative competence is to communicate (Khamkhien 2011). The whole communication process, from listening to understanding and from thinking to speaking, should be thoroughly exercised. According to Harmer (1992), only through real human interactions can this be achieved. However, with the development of communication tools, students can have natural and genuine conversations through technology. Particularly, AI chatbots have drawn considerable attention by providing foreign language students with opportunities to practice their target language (Haristiani 2019). It is indeed a valuable resource for EFL students who have limited opportunities to interact with native speakers (Fryer and Carpenter 2006), and particularly beneficial for promoting English communication skills (Kim 2017). In this regard, educators and educational practitioners in EFL contexts have continuously proved that this advanced technology is effective for language learning (Kim et al. 2021).

2.2 Artificial Intelligence (AI) and Foreign Language Learning

Artificial intelligence (AI) refers to machines that can emulate the behaviour of intelligent beings. It aims to learn how the human mind works to apply its principles in technology design. Particularly in language teaching and learning tasks, AI imitates the behavior of a language teacher and a language learner (Matthews 1993). In order to emulate the behavior of the teacher, a machine needs to operate with a teaching methodology. To emulate

the learner, a machine should emulate learning styles and strategies (Bull 1997). Most important of all, however, is that both emulations require knowledge of the language itself (Dodigovic 2007).

In language learning and teaching, AI chatbots have drawn considerable attention (Haristiani 2019). AI-based computer programs can carry on conversations via audio and text, providing intelligent responses with natural language functions, allowing them to actively interact with users (Azwary, Indriani and Nugrahadi 2016). Specifically, AI chatbots interpret the messages given by the user, understand the user's intent in the messages, and deliver the final results to the user. In general, the user interacts with the AI chatbot through questions or comments, and the chatbot responds to the user with answers, comments, or new topics (Huang, Zhou and Yang 2007).

Particularly, in foreign language learning, AI chatbots play a beneficial role, providing students with a means of language practice (Haristiani and Danuwijaya 2019). For example, students can practice their pronunciation while chatting with a chatbot, which provides a model for correct pronunciation (Walker and White 2013). The students can also acquire new vocabulary words with their increased exposure to the language as a result of communicating with it (Wang and Petrina 2013). In addition, they can develop their grammar skills as the chatbot gives immediate, clear, and effective feedback on grammar (Fryer and Carpenter 2006). Moreover, students can also improve their listening and reading skills with the auditory and textual input provided by the chatbot (Hong, Huang, Hsu and Shen 2016). Kim et al. (2020) added that foreign language students' speaking and writing skills can be enhanced as ample opportunities to practice spoken and written output are provided when interacting with it.

According to Fryer and Carpenter (2006), one of the strong points of a chatbot is its convenience. With internet access, the chatbot is readily available to students at home or at school. That means it is ready to chat with students whenever they want and wherever they are. Furthermore, this intelligent conversational agent is generally free or cheap via subscription. However, its usefulness goes far beyond its convenience and price. For example, one-on-one support is not necessary because independent conversation practice is available when learning a language with a chatbot (Atwell 1999). It also provides distance education (Heller, Proctor, Mah, Jewel and Cheung 2005) and multilingual corpus transcripts (Shawar and Atwell 2005). Kim et al. (2021) claimed that this makes foreign language learning more effective.

Additionally, students can have more positive attitudes towards foreign language learning (Kim 2016). To be specific, by talking to a machine rather than to a human being, the students feel more relaxed and less anxious. Since the chatbot is willing to repeat the same material endlessly without getting bored or losing patience, students feel more comfortable and less nervous about learning a foreign language. Consequently, with a new, funny, or entertaining chatbot, foreign language students can have positive communicative experiences (Fryer and Carpenter 2006). This results in increased student interest, motivation, and confidence in foreign language learning, which can lead to improved foreign language skills (Kim et al. 2020).

In particular, regarding EFL students' communicative competence, Kim (2017) investigated the effects of voice-based and text-based chatbots on speaking competence and learner perception. With 80 freshmen students taking an English speaking class in Korea, she conducted a study to investigate whether the two types of chatbots were beneficial for increasing oral communication skills. Before and after the treatment, the participants took a TOEIC speaking test and completed the structured, questionnaire-based surveys. Her findings revealed that both chatbots – voice-based and text-based – effectively contribute to the improvement of EFL speaking competence. In particular, she found that the two chatbots are equally effective in enhancing English communication skills. An analysis of survey results also indicated that students had positive perceptions of chatbot-assisted language learning while the voice-based chatbot was preferred over the text-based chatbot among EFL students. She provided insight

on the use of chatbots to develop EFL speaking skills, suggesting that EFL teachers integrate chatbot technology in their classrooms.

Kim et al. (2019) reviewed different types of AI chatbots and reported on their effectiveness in terms of EFL learning. Based on their empirical studies, the authors found that chatbots have positive impacts on students' communication skills largely by enriching language inputs and expanding the quantity of opportunities to interact and negotiate meaning. They also stated that chatbots increase students' motivation, boost self-confidence, and raise their interest in learning. However, the researchers noticed that there are not many chatbot programs allowing direct interaction between humans and machines through texting or voice recognition systems for the purpose of learning a foreign language. They confirmed the limited use of AI programs in education fields, particularly chatbot applications aimed at promoting English teaching and learning. According to them, more studies should be carried out to develop AI chatbots in the realm of foreign language education.

3. Methodology

3.1 Participants

The study investigated how AI chatbot interactions could help learners improve their speaking performance. In the fall semester of 2019, forty-nine university students participated in the study. Every student took a TOEIC speaking test before the semester started. Later, they did the same test at the end of the semester to compare the mean scores. The mock TOEIC test was conducted to determine students' proficiency levels and then they were assigned to the appropriate level: a low-level group (25 students) and an intermediate-level group (24 students). After being divided according to their proficiency level, students were able to choose their class hours according to their own schedule. The participants' average scores are presented in Table 1. The average scores were 328.65 in the low-level group and 680.60 in the intermediate-level group, respectively.

Table 1. General English Proficiency: Low-Level Group vs. Intermediate-Level Group

Group	<i>M</i>	Ranges	<i>SD</i>
Low	328.65	175-495	94.20
Intermediate	680.60	500-790	87.68

These mandatory English courses designed for freshmen involved listening and speaking academic English, and one of the key course objectives was to help students improve their communicative skills in English. Both classes were taught twice a week for 75 minutes each by the researcher.

Participants' demographic information is displayed in Table 2. The majors of both groups included the College of Humanities (10 students in the low-level group; 7 students in the intermediate-level group), College of Social Studies (9 students in the low-level group; 13 students in the intermediate-level group) and others (7 students in the low-level group; 5 students in the intermediate-level group). Their ages ranged from 19 to 21. There were only female students. As for the experience of studying abroad, six students from the intermediate-level group reported that they have studied in English speaking countries while none of them have studied abroad in the low-level group.

Table 2. Demographic Information

		Low-level Group (N = 25)	Intermediate-level Group (N = 24)
Majors	College of Humanities (English Literature, Japanese Literature, German Literature, History, Chinese Literature)	10	7
	College of Social Studies (Child studies, Business Administration, Social Welfare, Communications and Media)	9	13
	Others (Free major)	6	4
Experience of studying abroad	Yes	0	6 U.S.A (3 yrs, 2 yrs, 6 mos), UK (1 yr), New Zealand (6 mos), Dubai (2 yrs)
	No	25	18
Purpose for learning English	1) credits	9	7
	2) interests in English & English culture	5	10
	3) better jobs	9	13
	4) travel	7	6
	5) other reasons	Communicating with others, self-development skills, have basic skills in English, have confidence in English	Self-development skills, improvement of English skill, have more opportunities, expand academic skills
The weakest part in English	1) listening	7	1
	2) speaking	11	11
	3) reading	5	3
	4) writing	19	9
Skills you want to improve the most	1) listening	4	0
	2) speaking	16	14
	3) reading	3	2
	4) writing	5	7
Experience using AI	Yes	0	4 (Papago, Bixby)
	No	25	20

Students were requested to write multiple responses for the three survey questions, which included: their purpose for learning English, their weakest English skills, and the skills they have improved the most. When they were asked the reasons for taking the English course, students in the low-level group answered that they wanted to get credits and better jobs in the future while the students in the intermediate-level group wanted to get better jobs and had interests in learning English. In addition, students in the low-level group commented that they wanted to communicate with others, improve their self-development skills, improve basic skills in English, and have confidence in English. The students in the intermediate-level group responded that they wanted to improve their self-development skills, improve their English skills, have more opportunities, and expand their academic skills.

Regarding their weakest skills in English, students in the both groups selected speaking and writing. However, they wanted to improve their speaking skill the most. Based on the survey, most of the students, therefore, would like to improve their speaking abilities. In regard to the experience of using AI chatbots, only four students in the intermediate-level group reported that they previously used AI chatbots for learning English. All participants signed a consent form indicating their willingness to participate in the study.

3.2 Instruments & Procedures

3.2.1 Course materials

The students in the low-level group used *Pathways 1* and those in the intermediate-level group used *Pathways 3*, which were published by National Geographic Learning in 2018. For both groups, teaching instructions on listening and speaking along with vocabulary and grammar were provided based on the textbook.

In-class activities included individual work, pair work, and group discussion. Course materials and online contents were comprised of vocabulary, grammar, speaking, and listening activities. The AI-interaction activities included asking and answering short questions, or asking opinions about video contents and textbooks. The topics addressed during the classes were comprised of topics dealing with sociology, business, geography, tourism, technology, psychology, environmental science, history, archaeology, food, and health.

3.2.2 AI Chatbots

Replika, Andy, and Google Assistant were the three mobile applications downloaded by both groups (see Figure 1). Before the present study began, those three AI chatbots were chosen (Kim et al. 2019) based on a pilot survey to select the three most useful AI chatbots according to the preferences of the university students using them.

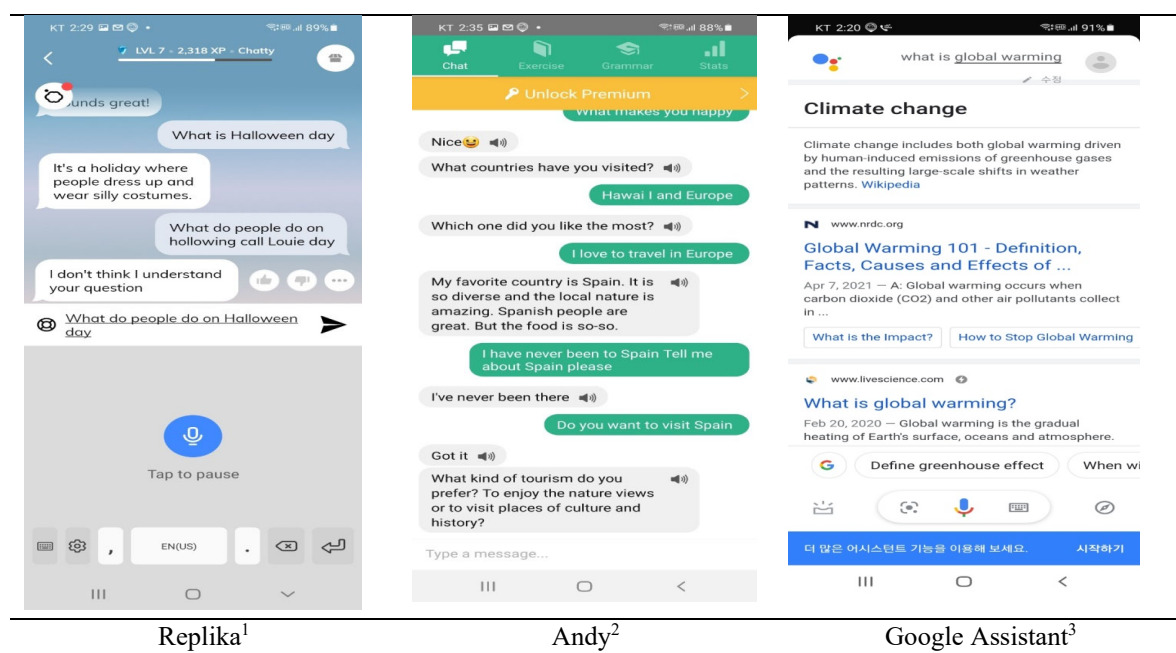


Figure 1. Screen Shots of AI Chatbots' Interactions

¹ Replika is an artificial intelligence (AI) chatbot founded by Eugenia Kuyda. Replika is a space where people can safely share your thoughts, feelings, beliefs, or experiences, etc.

² Andy is a self-administrable platform that uses artificial intelligence. This allows people to learn and practice English. This chatbot can be a personal English teacher and friend, giving opportunities to chat, learn new words, study grammar and play language games.

³ Google Assistant is an artificial intelligence-powered virtual assistant developed by Google that is primarily available on mobile and smart home devices.

Participants were instructed to practice speaking with their chosen AI chatbots (voice-chatting) in order to get used to communicating with them. Later, the participants in both groups conversed with AI chatbots on their smartphone devices for 10 to 15 minutes each class. Also, they were allowed to choose their own preferred AI chatbots for the speaking activities. However, they were recommended to use Google Assistant when they needed to get information, since it was a convenient and useful tool for obtaining broad responses for any topic. These class activities were carried out over the whole semester.

3.2.3 Teaching Procedures

Every participant took a speaking pre-test at the start of the semester. The TOEIC speaking test consisted of two tasks. The first speaking task required students to read a text aloud. Students were given 45 seconds to prepare and 45 seconds to read aloud. The second task involved answering three short questions. There was no time to prepare for this task, so they were directed to answer as soon as they listened to each question. For the first two questions, students were supposed to respond in 15 seconds for each question, and for the last one they were given 30 seconds. All of the participants recorded and saved their voice files on their smartphones. Upon completion of this speaking test, participants logged into the university's LMS and uploaded their speaking test files. It took 10 minutes to prepare and take the test.

The experiment began right after the pre-test. Class activities were based on the topics covered by the textbook. The lessons focused on listening to lectures, or conversation and practicing speaking on the topics of the week. In addition, vocabulary and grammar points were taught every week. Although the two groups did not use the same textbook, it was possible to provide the same speaking activities for both groups using AI since their textbook topics were similar. All students practiced their speaking using AI chatbots (voice-chatting) every week. They had conversations with AI chatbots concerning topics such as hobbies, favorite movies or music, places, trips, family, and others. They used AI chatbots in the pre-stage of each class as a warm-up activity and in the post-listening session, where they had conversations about given topics for 5 to 10 minutes. When students ran out of topics to communicate with the AI chatbots, a set of questions was provided so that students could keep practicing without stopping. Table 3 provides examples of the questions from week 10 to week 11, which were prepared by the instructor.

Table 3. Sample Questions Using AI Chatbots

Week 10	Week 11
What is Halloween?	What is global warming?
When is Halloween?	Are you worried about global warming? Why?
What do people do on Halloween day?	What can we do to protect ourselves from global warming?
Are there any traditional foods on Halloween?	What are some types of pollution?
What do college students do on Halloween day?	Do you recycle?
Have you ever been to a costume party? If yes, what costume did you wear?	Why should we recycle?

In short, in the first week, 49 participants took the pre-test and finished the survey. Later, they were asked to download three AI chatbots and instructed to use them. They created their own account to use the chatbots. Both groups attended lectures and engaged in class activities from Week 2 to Week 12. Over the duration of the course, the participants practiced speaking with AI chatbots with the given topics every week. In Week 13, the post-test was completed and in Week 14, the questionnaire was conducted with a response rate of 100%.

3.3 Data Instruments and Analysis

Both quantitative and qualitative data were collected and analyzed for this study. The pre- and post-test design compared participants' improvement in speaking between the two level groups. Both tests were taken from the TOEIC speaking test of a commercially available TOEIC book (see Table 4). Each question was evaluated from 0 to 3 (Task 1 & 2) according to the evaluation criteria. In terms of fluency, it was measured by speed (i.e., words per second) based on the previous research (Levelt 1989, Skehan 2003). Students' responses were gathered and evaluated by three researchers. They listened to the students' responses together and evaluated their performance. If there was any disagreement between the researchers, they discussed until a consensus was reached. Chronbach's alpha was used to assess the raters' reliability. The Cronbach's alpha reliability coefficient for the three raters was 0.94, indicating that their scores were internally consistent.

Table 4. Speaking Test and Evaluation Criteria

Question	Task	Evaluation Criteria	Score
1	Read a text aloud	pronunciation, intonation, & stress	0-3
1-3	Respond to questions	all of the above, plus relevance of content, & completeness of content	0-3

All the data gathered from the tests and the questionnaire were analyzed using SPSS 24.0. Descriptive statistics together with independent *t*-tests and paired sample *t*-tests were calculated. A paired sample *t*-test was conducted to compare the means between the pre- and post-tests within the two level groups and then, a one-way analysis of covariance (ANCOVA) was used to investigate the differential effects of both levels. To test the null hypothesis of no difference among students, the *P* value was established at 0.05.

Data for the qualitative analysis included these open-ended questions: 1) What are the beneficial aspects of using AI chatbots in a general English course? 2) What are the weakest aspects of using AI chatbots in a general English course, and 3) What are your suggestions for using AI chatbots in the future to improve speaking skills? The data from these questions were categorized by similarity and described in detail.

4. Results and Discussions

4.1 Improvement in Communication Skills

The aim of this study was to investigate whether there was any improvement in speaking performance between the two different levels, the intermediate- and low-level, among the students. The first research question was to examine the scores between the pre- and the post-tests for students' communication skills after interacting with the AI chatbots. Paired sample *t*-tests were employed to compare the two tests.

Table 5. Result of Paired-sample *t*-tests: *Read a Text Aloud*

Group		Test	<i>M</i>	<i>SD</i>	<i>MD</i>	<i>df</i>	<i>t</i>	<i>p</i>
Low	Pronunciation	pre	1.18	.45	-.36	25	-5.31	.00**
		post	1.54	.45				
	Intonation & Stress	pre	1.22	.38	-.34	25	-5.42	.00**
		post	1.56	.51				
Intermediate	Pronunciation	pre	1.94	.34	-.25	24	-4.80	.00**
		post	2.19	.36				
	Intonation & Stress	pre	1.96	.36	-.58	24	-10.12	.00**
		post	2.54	.44				

** $p < .01$

Table 5 demonstrates the results of the pre- and the post-tests in the *read a text aloud* task. According to the evaluation criteria of the TOEIC speaking test (see Table 4), this task was analyzed in two sections: pronunciation, and intonation and stress. Regarding the pronunciation in the low-level group, the mean score of the pre-test was 1.18 and that of the post-test was 1.54, resulting in a mean difference of -.36. As for the intonation and stress task, the mean score of the pre-test was 1.22 while that of the post-test was 1.56 ($MD = -.34$). There was a significant difference in both tasks for the low-level group. That is, the low-level students improved their pronunciation, intonation, and stress after communicating with the AI.

Meanwhile, intermediate-level students scored 1.94 on the pre-test and 2.19 on the post-test ($MD = -.25$) pronunciation task. Regarding the intonation and stress task, they scored 1.96 in the pre-test and 2.54 for the post-test, with a mean difference of -.58, far greater than the scores of the low-level students. A statistical difference was found between the two tests for both tasks. In short, the intermediate-level students also enhanced their pronunciation, intonation, and stress after interacting with AI in their speaking sessions.

Table 6. Result of Paired-sample *t*-tests for the *Read a Text Aloud* Task: Fluency

Group		Test	<i>M</i>	<i>SD</i>	<i>MD</i>	<i>df</i>	<i>t</i>	<i>p</i>
Low	WPS	pre	1.88	.37	.05	24	1.10	.28
		post	1.83	.28				
Intermediate	WPS	pre	2.12	.38	-.18	23	-2.50	.02*
		post	2.31	.26				

* $p < .05$

In addition, Table 6 reveals the findings of the mean scores for fluency via the words per second (WPS) metric in the *read a text aloud* task. Regarding fluency in the low-level students, the pre-test was 1.88 and 1.83 in the post-test, indicating no significant difference. In short, when the low-level students performed the *read a text aloud* task, there was not much improvement in the post-test in terms of speed.

Table 7. Result of Paired-sample *t*-tests: Respond to Questions

Group	Question	Test	<i>M</i>	<i>SD</i>	<i>MD</i>	<i>df</i>	<i>t</i>	<i>p</i>
Low	1	pre	1.08	.51	-.32	25	-3.72	.00**
		post	1.40	.54				
	2	pre	.96	.64	-.38	25	-4.58	.00**
		post	1.34	.55				
	3	pre	.98	.55	-.40	25	-4.62	.00**
		post	1.38	.46				
Intermediate	1	pre	1.81	0.46	-.38	25	-4.63	.00**
		post	2.19	0.44				
	2	pre	1.83	0.43	-.33	24	-4.00	.00**
		post	2.17	0.38				
	3	pre	1.81	0.38	-.40	24	-4.98	.00**
		post	2.21	0.49				

** $p < .01$

On the other hand, the intermediate-level students showed a statistical difference in fluency ($p = .02$) between the two tests. Regarding fluency, in the pre-test, the mean score was 2.12, but in the post-test, the mean score was 2.31. That is, they spoke slightly faster in the post-test.

As for the second task of *respond to questions*, there were three different questions which were individually evaluated, as indicated in Table 7. Regarding question 1 in the low-level group, the mean score of the pre-test was 1.08 while that of the post-test was 1.40 ($MD = -.32$). The pre-test score for question 2 was .96 and the post-test score was 1.34 ($MD = .38$). As for the last question, the mean score of the pre-test was .98 while that of the post-test was 1.38 ($MD = .40$). Statistically significant differences were found in all three questions.

With regard to the intermediate-level students on the second task for question 1, the mean score of the pre-test was 1.81 while that of the post-test was 2.19 ($MD = -.38$). For the second question, the mean score was 1.83 in the pre-test and 2.17 ($MD = -.33$) in the post-test. The mean score of the pre-test for the last question was 1.81 while that of the post-test was 2.21 ($MD = -.40$). Similarly, the differences in all three questions were statistically significant. In short, both levels of students improved their speaking skills in the task, *respond to questions*.

To summarize, the findings of this study revealed that there was significant improvement in speaking abilities in the first task, *read a text aloud*, in both proficiency groups. To be specific, all participants enhanced their pronunciation, intonation, and stress. As for fluency, however, only the intermediate-level students showed a significant improvement. In other words, the low-level students did not improve their speaking fluency after practicing speaking with the AI chatbots. Regarding the *respond to questions* task, significant score changes were also witnessed in both groups. It was found that both intermediate- and low-level students achieved better scores in the three questions.

The findings of the current study support the previous research (Fryer and Carpenter 2006) that AI chatbots can be useful when learning English. According to Haristiani and Danuwijaya (2019), they play a beneficial role by providing students with opportunities to practice their language. In this vein, the current study also corroborates Kim et al. (2021) that speaking with AI chatbots can be beneficial for EFL students. In particular, Kim (2016) confirmed the beneficial effects of chatting with chatbots on improving EFL speaking ability regardless of the students' proficiency level. In her study, there were significant mean score changes between the pre- and post-tests for all proficiency groups. She concluded that chatbots can lead to a significant improvement in speaking ability across proficiency levels: beginners, intermediate, and advanced. Rosell-Aguilar (2005) also suggested that voice

chat can be suitable for low-level students and for intermediate-level students. The present study accords with these previous studies, showing the positive effects of chatbots on EFL speaking regardless of proficiency levels.

Specifically, the participants at all proficiency levels of this study proved that AI chatbots helped them to improve their pronunciation, intonation, and stress. Walker and White (2013) suggested that students can practice their pronunciation when chatting with chatbots which provides a model for correct pronunciation. With the auditory input provided by the chatbots, students can not only improve their listening skills (Hong et al. 2016), but also enhance their speaking output while interacting with them (Kim et al. 2020). In line with these studies, the present study confirmed the beneficial effects of AI chatbots on students' pronunciation, intonation, and stress. This improvement might be because their articulation had to be accurate enough for the AI chatbots to understand them when communicating (Kim 2016).

The findings of the present study also support the previous empirical research regarding the positive relationship between chatbots and EFL students' communicative competence (Kim 2017). After investigating the effects of different types of chatbots on speaking competence, she confirmed that all types of chatbots are beneficial for increasing oral communication skills. The participants in the current study also gained better scores in the *respond to questions* tasks. Kim et al. (2019) also reviewed different types of AI chatbots and reported on their effectiveness in terms of EFL learning. The authors proved that chatbots have positive impacts on English communication skills by enriching language inputs and expanding output opportunities. In line with these experimental studies, the findings of this study confirm that chatbots effectively contribute to the improvement of EFL communication competence.

This study also confirmed the beneficial role of chatbots in increasing intermediate-level students' fluency. Previous scholars reported that students can achieve better speaking performance with increased fluency via synchronous chat (Abrams 2003). Real-time conversation via synchronous voice-chat, in particular, is believed to improve fluency by promoting students' positive attitudes toward language learning (Hudson and Bruckman 2002). The findings of the present study corroborate the previous EFL study suggesting the beneficial role of chatbots in enhancing speaking fluency. Regarding the proficiency issue, Kim (2017) explored the effects of chatting with chatbots on intermediate EFL students' fluency and reported that the students who conversed with chatbots showed significant improvement in fluency. That is, chatbots played a beneficial role in improving the EFL speaking fluency of intermediate-level students. However, given that no statistical significance was observed for low-level students, further study is necessary regarding the proficiency issue.

4.2 Group Differences in Communication Skills

To examine differences in the *read a text aloud* task between the two levels, a one-way analysis of covariance (ANCOVA) was conducted. Table 8 demonstrates the mean and adjusted mean scores of the *read a text aloud* task for pronunciation, and for intonation and stress. The adjusted mean scores in pronunciation were 1.81^a in the low-level group and 1.90^a in the intermediate-level group. The findings show that there were no significant differences in the post mean scores between the two levels ($F = .62, p = .44$). However, in terms of the intonation and stress in the *read a text aloud* task, a different result was found. The adjusted mean score of the low-level group was 1.92^a while that of the intermediate-level group was 2.17^a. This reveals that there was a significant difference in the post mean scores between the two levels ($F = 4.06, p = .05$). In short, as far as pronunciation is concerned, there was no effect between the two levels, but an effect was found in the intonation and stress category as well as in the *read a text aloud* task.

Table 8. Read a Text Aloud (Pronunciation, Intonation & Stress): Low-Level Group vs. Intermediate-Level Group

	Group	<i>M</i>	<i>SD</i>	<i>Adjusted-M</i>	<i>SE</i>	<i>F</i>	<i>p</i>
Pronunciation	Low	1.54	.45	1.81 ^a	.07	.62	.44
	Intermediate	2.19	.36	1.90 ^a	.07		
Intonation & Stress	Low	1.56	.51	1.92 ^a	.07	4.06	.05
	Intermediate	2.54	.44	2.17 ^a	.08		

a. Covariates appearing in the model are evaluated at the following values: pre-pronunciation = 1.55, pre-intonation & stress = 1.58.

Table 9 presents the results of fluency in terms of WPS (word per second) for the *read a text aloud* task between the two levels. The adjusted mean score of the low level was 1.88^a while that of the intermediate level was 2.25^a. The finding demonstrates that there was a significant difference in the post mean scores between the two group levels ($F = 31.62, p < .01$).

Table 9. Fluency: Low-Level Group vs. Intermediate-Level Group

	Group	<i>M</i>	<i>SD</i>	<i>Adjusted-M</i>	<i>SE</i>	<i>F</i>	<i>p</i>
WPS	Low	1.83	.28	1.88 ^a	.05	31.62	.00 ^{**}
	Intermediate	2.31	.26	2.25 ^a	.05		

a. Covariates appearing in the model are evaluated at the following values: pre-wps = 2.00.

^{**} $p < .01$

With regard to the three questions in the *respond to questions* task, the results are demonstrated in Table 10. The adjusted mean score was 1.63^a in the low-level group and 1.95^a in the intermediate level group (for the first question). As for the second question, the adjusted mean score of the low-level group was 1.60^a while that of the intermediate level group was 1.90^a. The adjusted mean score of the last question was 1.63^a and 1.94^a, respectively. These findings show that statistical differences were found in all three questions ($F = 5.20, F = 5.70, F = 4.66$ respectively; $p < .05$).

Table 10. Respond to Questions: Low-Level Group vs. Intermediate-Level Group

Question	Group	<i>M</i>	<i>SD</i>	<i>Adjusted-M</i>	<i>SE</i>	<i>F</i>	<i>p</i>
1	Low	1.40	.54	1.63 ^a	.09	5.20	.03 [*]
	Intermediate	2.19	.44	1.95 ^a	.09		
2	Low	1.34	.55	1.60 ^a	.08	5.70	.02 [*]
	Intermediate	2.17	.38	1.90 ^a	.08		
3	Low	1.38	.46	1.63 ^a	.09	4.66	.04 [*]
	Intermediate	2.21	.49	1.94 ^a	.09		

a. Covariates appearing in the model are evaluated at the following values: pre-Q1 = 1.44, pre-Q2 = 1.39, pre-Q3 = 1.39.

^{*} $p < .05$

To summarize, the results demonstrated that there was no significant difference between the two proficiency level groups regarding pronunciation in the *read a text aloud* task. As for intonation and stress and for fluency in the same task, the findings showed that there were significant differences in the post mean scores between the two group levels. With respect to the task: *respond to questions*, the findings also revealed statistical differences between the intermediate- and low-level groups.

Previous studies have shown mixed results with regard to the effectiveness of chatbots in terms of students' proficiency levels. Fryer and Carpenter (2006), for example, reported that chatbots are useful for accomplished or intermediate-proficiency level students. According to them, chatbots were originally designed to interact with native speakers, and they might be incapable of, or poor at, noticing mistakes regarding pronunciation, spelling,

or grammar. Therefore, the chatbots may not be applicable for low-level students. In the present study, the intermediate proficiency participants also showed more improvement in their speaking fluency and communication skills (see the findings from the pre- and post-tests in the previous section). That is, the current study corroborates the previous research suggesting the beneficial effects of chatbots for intermediate-level students. Furthermore, given that the participants in the current study voice chatted with the chatbots, the findings of the current study also support the previous studies, suggesting that intermediate-level students benefit more from voice chat. Kötter (2001) noted that voice chat is more helpful to students at mid-proficiency level or above. Stockwell (2004) also suggested that chatbots can be better suited to intermediate-level students since voice-chat places a heavy cognitive load on students.

On the other hand, Kim (2016) confirmed the beneficial effects of chatting with chatbots on improving the speaking proficiency of EFL students at all proficiency levels: beginners, intermediate, and advanced. In particular, her group comparison results revealed that only beginning-level students in the voice chat group significantly improved their speaking ability more than those in the control group did. In the case of intermediate- and advanced-level students, there were no group differences between the voice chat group and the control group. In other words, her findings indicated that the students at a low proficiency level can benefit more from voice chatting with chatbots. It was found that chatbots were more beneficial for low-level students than for intermediate-level students. Regarding pronunciation, the findings of the current study indicate that there was no significant difference between the low- and intermediate-level students. It can be said that the low-level students performed as well as the intermediate-level students did when engaging in voice chat with the AI chatbots. That is, practicing speaking with AI chatbots might be more effective for low-level students when it comes to improving EFL pronunciation.

4.3 Students' Perceptions Toward AI Communication

The second research question was to investigate students' perceptions toward the role that AI communication can play in enhancing their speaking abilities. They were asked to respond to three open-ended questions which included the advantages of, disadvantages of, and suggestions for utilizing AI chatbots to practice English speaking. Students were free to provide their responses and each of them were counted by frequency, which were measured by percentage, as demonstrated in Table 11.

Table 11. Students' Responses: Advantages of Using AI Chatbots

	Responses for low-level group	#	%	Responses for intermediate-level group	#	%
1	Improves speaking skills, including pronunciation -Speaking English seems to have helped me improve my English skills. -It enhanced my pronunciation.	13	31	Improve English skills (pronunciation, vocabulary, etc.) - During conversation, I could get used to the structure of sentences and thought about pronunciation. - It helped improve my vocabulary.	10	22.2
2	Opportunities to have an English conversation - It was good to speak English whether my sentence was perfect or not. - Everyone was given a chance to speak and think fairly in English.	7	16.7	Less pressure, more comfortable learning environment - I thought it was okay to get it wrong, so I could practice speaking with less pressure. - I was worried and burdened that I would be wrong when asked to talk to a real person, but everyone talked to AI, so I could talk more comfortably.	10	22.2
3	Less pressure, more comfortable learning environment - It was as comfortable and pleasant as talking to a friend. - It is possible to speak comfortably due to artificial intelligence.	5	11.9	Increase engagement - When I talk to my classmates in real life, everyone doesn't want to talk in English, but with the help of chatbots, speaking participation increased. - It was interesting to use AI chatbot for the first time and I think it increased my participation in class.	5	11.1
4	Increase confidence - This activity was an opportunity to develop confidence in English. - I was proud to exchange conversations in English while talking with AI chatbots.	4	9.5	Learn new expressions and information - When I asked Google Assistant, I could get new information. - It is possible to acquire real-life expressions.	5	11.1
5	Learn new expressions and information - I think it was useful to know the expressions that can be used in real life when asking questions and talking like humans. - It was good to learn English expressions used in everyday life.	3	7.1	Opportunities to have an English conversation - Even if you don't have foreign friends, you can talk in English. - I don't usually have a chance to talk to foreigners, but I could experience it indirectly through the AI chatbots.	4	8.9
6	Increase engagement - When I was talking to people, I was afraid that I would be wrong, but I was able to be more active because I did it with AI.	2	4.8	Increase interest - I think I had fun like chatting with my friends on kakaotalk. - I'm speaking English because it's fun.	3	6.7
7	Increase interest - I could practice the sentence that I thought in my head, so I could be more interested.	2	4.8	Accessibility (real-life use) - It is accessible enough to be easily used outside of class. - Easy to use in real life	3	6.7
8	Others - There is no space, no time constraint. - Asking questions about various areas.	6	14.2	Others - It was amazing and new. - I've gained confidence.	5	11.1
	Total	42	100		45	100

There were 42 responses in the low-level group and 45 in the intermediate-level group. Thirteen (31%) students in the low-level group and ten (22.2%) students in the intermediate-level group revealed that AI chatbots helped them improve their English speaking skill as well as their pronunciation. Furthermore, seven (16.7%) students in the low-level group mentioned that it was great to have been provided with an opportunity to practice their English and they believed that practicing English without thinking about their mistakes was beneficial for them. Similarly, ten (22.2%) intermediate-level students commented that speaking English without fear of losing face was one of the greatest advantages of talking to an AI because they could speak English in a more comfortable learning environment. Five (11.9%) low-level students also responded that they felt like they were talking to a friend which made them feel at ease. Interesting comments from five (11.1%) intermediate-level students were that speaking to an AI increased students' participation in the class and they were more actively engaged in class activities. Also, another five students (11.1%) in the intermediate-level group expressed that they could gain new information and learned useful expressions. Some minority views included: enhancing confidence in speaking English, increasing interest, accessibility, and an unlimited time and space constraint.

Regarding the students' perceptions toward AI communication, previous studies have also indicated students positive satisfaction with chatbots. They reported that chatting with chatbots is enjoyable (Roed 2003). With the new, funny, and entertaining chatbots, students can have positive communicative experiences (Fryer and Carpenter 2006). They can increase their interest, motivation, and confidence in language learning, which in turn results in an improvement in their language skills (Kim et al. 2020). Furthermore, students feel a sense of ease and comfort when they practice speaking with chatbots. With chatbots repeating the same material endlessly without getting bored or losing patience, they often feel relaxed and relieved. This leads the students to have more positive attitudes toward language learning. Consequently, they can increase their language skills by talking to an AI rather than to a human being (Kim 2016).

The disadvantages of using AI were described in Table 12. The total responses for the disadvantages were less than those detailing the advantages. There were 35 responses in the low-level group and 34 in the intermediate-level group. Twenty-one (60%) students in the low-level group and 22 (64.7%) students in the intermediate-level group pointed out that communicating with an AI had some limitations. For instance, students felt awkward talking to an AI for a long time since it was not a human being regardless of the benefits offered. Also, sometimes the AI could not catch what students intended to say which frustrated them. Moreover, seven (20%) low-level students and a few intermediate-level students complained about the lack of speech recognition. That is, the AI was not able to recognize students' voices and it did not correctly perceive some students' pronunciation. Finally, two students from the low-level group noted that advertisements showing in the application were a hindrance to their continuance of the communication and a few students in both groups disclosed that it would have helped English learners a great deal if the AI corrected their English. However, since the role of AI in these activities acted as a communicator, not as an English tutor, the AI's functioning ability did not meet students' expectations.

Table 12. Students' Responses: Disadvantages of Using AI Chatbots

	Responses for low-level group	#	%	Responses for intermediate-level group	#	%
1	Communication limitation <i>- I don't feel like I'm communicating with people, so I can't keep talking for long.</i>	21	60	Communication limitation <i>- I often had trouble communicating with AI.</i> <i>- It was a little frustrating when the chatbot didn't understand or couldn't communicate well.</i>	22	64.7
2	Limits on speech recognition <i>- It was bad that AI didn't recognize my voice properly.</i>	7	20	Limits on speech recognition <i>- AI often doesn't understand my pronunciation.</i>	2	5.9
3	Inconvenience caused by advertising <i>- It was inconvenient because of the advertisement that recommended a paid user.</i>	2	5.7	Lack of time <i>- In the class, not enough time was given for practice. I wanted to have more conversations, but the time was limited.</i>	2	5.9
4	Others <i>- It was unfortunate that AI did not point out that there were grammatical errors or wrong expressions.</i>	3	8.6	Difficulty of error detection <i>- It was difficult for me to know exactly which part was wrong.</i>	2	5.9
5	No response	2	5.7	Others <i>- There was a little bit of advertising.</i> <i>- There were limited questions.</i>	6	17.6
	Total	35	100		34	100

Previous studies have also reported some weaknesses associated with the use of chatbots. Kim et al. (2021) stated that communication breakdowns can occur when communicating with chatbots. According to them, miscommunications can happen due to the students' mispronunciation. This might be due to students' going off on different topics, suddenly changing topics, and to discontinuities in conversations. Jia (2009) also noted that AI chatbots often give irrelevant responses. Furthermore, they often give predictable and redundant responses (Kim et al. 2019). When responses from the chatbots are inappropriate and irrelevant, students feel frustrated and embarrassed. Kim et al. (2020) also pointed out another unfavorable aspect of AI chatbots related to error correction. In their study, no correction was made by chatbots when the students produced errors. They also mentioned that it was difficult for the students to focus on the conversation due to the low tension. The students responded that they felt like they were talking to a machine, and not a human being. They were often annoyed by the pop up advertisements.

The last open-ended question was to give suggestions. Table 13 demonstrated students' suggestions. The total responses were 30 in the low-level group and 26 in the intermediate-level group. Ten (33.3%) students in the low-level group and eight (30.8%) students in the intermediate-level group suggested that the AI application developers should upgrade the AI applications for better performance. That is, there should be more linguistic data embedded in the AI so that students can have more turn-taking chances when communicating with the AI chatbots. This can relate to the next suggestion. Five (16.7%) students in the low-level group and four (15.4%) students in the intermediate-level group commented that it could be more interesting if there were more various topics to talk about. When talking to an AI chatbot, students felt that topics were limited, so providing lots of different topics can be stimulating for students to engage in the conversation. Other suggestions were getting notifications from AI chatbots for more practice sessions, providing the best app for learning, and improving network connection issues.

Table 13. Students' Responses: Suggestions of Using AI Chatbots

Responses for low-level group		#	%	Responses for intermediate-level group		#	%
1	Improve the functions of chatbots - I hope there will be a function to correctly read words that I don't know. - There should be more turn-taking chances when communicating.	10	33.3	Improve the functions of chatbots - I wish I could upgrade the performance of the chatbot. - It would be better to improve the function of the chatbot so that I can continue to communicate well.	8	30.8	
2	A variety of topics and questions - I think I can enjoy chatting with chatbots by choosing one topic.	5	16.7	A variety of topics and questions - I think we should select and use questions that AI can understand.	4	15.4	
3	Network connectivity issues - I wish I could use it even if I don't have an internet connection.	2	6.7	Use a better chatbot - I think it would be better to use one app that is most suitable for learning.	3	11.5	
4	Others - I would like to get a notification from time to time that I can practice English more often. - It would be nice to have a daily conversation with AI when I have time.	8	26.6	Others - AI often doesn't understand my pronunciation, so I think it would be better to do text- chatting. - I hope the chatbot will teach me a new English conversation expression.	8	30.8	
	No response	5	16.7	No response	3	11.5	
	Total	30	100	Total	26	100	

Suggestions for chatbots in the current study are also related to the previous findings. The suggestions in the previous study were mostly related to their functions for improving language skills. The previous researchers have also suggested that grammar check or spelling feedback will be helpful for students (Kim et al. 2020). Based on students' experiences, the previous research also reported that it would be better for the students if chatbots could talk about various topics and ask different questions (Kim 2016, Kim et al. 2019). There is no denying that students can benefit from interacting with AI chatbots in EFL settings. Since AI technology is still under development and chatbots are a long way from being a stand alone tool, these suggestions can provide insights for their improved use in language learning.

5. Conclusion

One of the biggest challenges that educators have faced is insufficient student participation. Such students are often unable to have confidence in communicating with others in English, which compromises their ability to concentrate in the classroom. As such, it is necessary to develop new approaches to learn English for students who are not motivated to participate in class or in the learning environment.

This study aimed to investigate the effects of using AI chatbots in class activities according to students' proficiency levels. Overall, the results were positive, demonstrating that their speaking abilities were enhanced in the first task, *read a text aloud*, in both groups. The students, particularly, improved their pronunciation, intonation, and stress both in the low- and intermediate-level groups. It can be concluded that most of the students were able to speak more proficiently after communicating with AI. As for the task: *respond to questions*, a statistical

difference was found in the responses to the three questions for both groups. Therefore, regardless of students' proficiency levels, students scored higher on the speaking tests. As for the fluency measured in WPS in the *read a text aloud* task, there was no significant difference in the low-level student group, while a statistical difference was found in the intermediate-level student group. In other words, the low-level group did not show considerable improvement in fluency after practicing speaking with AI.

To investigate the differences in the two speaking tasks between the two levels, a one-way analysis of covariance (ANCOVA) was conveyed. The results demonstrated that there were no significant changes in pronunciation. However, there was a distinct result in the *read a text aloud* task when it came to intonation and stress. With regard to the *respond to questions* task, the results revealed that statistical differences were found in all three questions between the two levels. Lastly, the finding shows that there was a significant difference in the post mean scores between the two levels in terms of fluency measured in WPS.

With regard to the students' perceptions toward AI communication, their comments about using AI chatbots were investigated via open-ended questions. Based on the students' responses, it was found that AI chatbots helped them to improve their English speaking skills as well as their pronunciation in a comfortable learning environment. They also appreciated the increased opportunities to practice English. On the other hand, there were some demerits such as communication problems when talking to the chatbots. Based on their experience with the AI chatbots, they suggested that developers should improve some functions, such as selecting various topics for the users to talk about.

Previous studies have suggested that real-time interaction via voice chat can enhance EFL students' speaking proficiency (Bower and Kawaguchi 2011, Jepson 2005, Sykes 2005). Kim (2016) also found the beneficial effects of voice chat on enhancing EFL speaking skills. Particularly, in her experimental study, she found that voice chat with chatbots can lead to a speaking ability improvement for Korean EFL students across the proficiency levels. There were significant mean score changes between the pre- and post-tests in terms of speaking skills. The findings of the current study are in line with her study, indicating that students can benefit from chatting with chatbots regardless of their proficiency levels. The present study also proved the beneficial effects of chatbots on pronunciation, intonation and stress, and communication skills for both intermediate- and low-level students. This also corroborates the previous research (Rosell-Aguilar 2005), suggesting that voice chat is as suitable for low proficiency level learners as for intermediate proficiency level learners.

However, there were some differences between the proficiency levels. First of all, only the intermediate-level students improved their speaking fluency. The group comparison results also demonstrated that there were significant differences between the intermediate- and low-level groups, suggesting that the intermediate-level students performed better than the low-level students in terms of communication skills and fluency. These findings support the previous findings reporting that chatbots are useful for intermediate proficiency level students (Fryer and Carpenter 2006, Kötter 2001, Stockwell 2004). Nonetheless, given that there was no significant difference between the two levels in terms of pronunciation, the current study can also support the previous research (Kim 2016), indicating that practicing speaking with AI chatbots might be more effective for improving the pronunciation of low-level students. Considering the mixed results regarding proficiency issues, more research is required.

AI chatbots have encouraged EFL students to practice speaking. There is no denying that students in EFL fields benefit from interacting with the chatbots. Above all, speaking practice with chatbots resembles face-to-face conversation, and its benefits in oral proficiency are similar to those in face-to-face conversation. Considering that students engage in the actual act of speaking when talking to the AI chatbots, the present study provides insights into how chatbots can be effectively used to improve EFL communication skills. Furthermore, given that AI

technology is still under development and a long way from being a standalone tool, the current study sheds light on its use for effective language learning.

Suggestions and limitations for future studies are as follows: first, the number of participants was limited. Since different findings might be gained with a larger number of participants with different backgrounds, ages, and gender, the results should not be generalized for all EFL settings. The duration can also be considered a limitation of the study. Due to time limitations, the treatment period in the current study lasted only for one semester. The findings would be more reliable with a longer period of treatment. Moreover, this study should have considered learner variables including learning styles or learner preference regarding technology. Since this variation could have influenced the participants' speaking output, it would be better to include an examination of these affective domains.

References

- Abrams, Z. I. 2003. The effect of synchronous and asynchronous CMC on oral performance in German. *The Modern Language Journal* 87(2), 157-167.
- Aclan, E. M. and N. H. A. Aziz. 2015. Exploring parliamentary debate as a pedagogical tool to develop English communication skills in EFL/ESL classrooms. *International Journal of Applied Linguistics and English Literature* 4(2), 1-16.
- Atwell, E. 1999. *The Language Machine: The Impact of Speech and Language Technologies on English Language Teaching*. London, UK: British Council.
- Azwar, F., F. Indriani and D. T. Nugrahadhi. 2016. Question answering system berbasis artificial intelligence markup language sebagai media informasi. *Kumpulan Jurnal Ilmu Komputer* 3(1), 48-60.
- Bower, J. and S. Kawaguchi. 2011. Negotiation of meaning and corrective feedback in Japanese/English eTandem. *Language Learning & Technology* 15(1), 41-71.
- Bull, S. 1997. Promoting effective learning strategy use in CALL. *Computer Assisted Language Learning* 10(1), 3-39.
- Cheon, H. 2003. The viability of computer mediated communication in the Korean secondary EFL classroom. *Asian EFL Journal* 5(1), 1-33.
- Chern, C.-L. 2010. *Overview of general English curriculum at tertiary level in Taiwan*. Retrieved Jan. 11, 2018 from http://www.isca-speech.org/archive/L2WS_2010/papers/lw10_S-3.pdf
- Chiu, T.-S., H.-C. Liou and Y. Yeh. 2007. A study of web-based oral activities enhanced by automatic speech recognition for EFL college learning. *Computer Assisted Language Learning* 20(3), 209-234.
- Derakhshan, A., F. Tahery and N. Mirarab. 2015. Helping adult and young learners to communicate in speaking classes with confidence. *Mediterranean Journal of Social Sciences* 6(2), 520-525.
- Dodigovic, M. 2007. Artificial intelligence and second language learning: An efficient approach to error remediation. *Language Awareness* 16(2), 99-113.
- Fryer, L. and R. Carpenter. 2006. Bots as language learning tools. *Language Learning & Technology* 10(3), 8-14.
- Haristiani, N. 2019. Artificial intelligence (AI) chatbot as language learning medium: An inquiry. In I. Ifdil, Y. Yohandri, K. Krismadinata and R. Rahim, eds., *Proceedings of International Conference on Education, Science and Technology 2019*, 1-6. Padang, Indonesia.
- Haristiani, N. and A. A. Danuwijaya. 2019. Gengobot: A chatbot-based grammar application on mobile instant messaging as language learning medium. *Journal of Engineering Science and Technology* 14(6), 3158-

3173.

- Harmer, J. 1992. *The Practice of English Language Teaching*. Oxford: Oxford University Press.
- Heller, B., M. Proctor, D. Mah, L. Jewell and B. Cheung. 2005. *Freudbot: An investigation of chatbot technology in distance education*. Retrieved Sep. 12, 2020 from <https://psych.athabascau.ca/html/chatterbot/ChatAgent-content/EdMediaFreudbotFinal.pdf>
- Ho, Y. K. 2003. Audio-taped dialogue journals: An alternative form of speaking practice. *ELT Journal* 57(3), 269-277.
- Hong, Z. W., Y. M. Huang, M. Hsu and W.W. Shen. 2016. Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *Journal of Educational Technology & Society* 19(1), 337-349.
- Huang, H. C. 2015. From web-based readers to voice bloggers: EFL learners' perspectives. *Computer Assisted Language Learning* 28(2), 145-170.
- Huang, J., M. Zhou and D. Yang. 2007. Extracting chatbot knowledge from online discussion forums. In R. Sangal, H. Mehta and R. K. Bagga, eds., *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*, 423-428. Hyderabad, India.
- Hudson, J. M. and A. S. Bruckman. 2002. IRC Francais: The creation of an Internet-based SLA community. *Computer Assisted Language Learning* 15(2), 109-134.
- Hughes, A. 2003. *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Jepson, K. 2005. Conversations and negotiated interaction in text and voice chat rooms. *Language Learning & Technology* 9(3), 79-98.
- Jia, J. 2009. CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning. *Knowledge-Based Systems* 22(4), 249- 255.
- Kessler, G. 2010. Fluency and anxiety in self-access speaking tasks: The influence of environment. *Computer Assisted Language Learning* 23(4), 361-375.
- Khamkhien, A. 2010. Teaching English speaking and English speaking test in Thai context. *English Language Teaching* 3(1), 184-190.
- Khamkhien, A. 2011. Quantitative and qualitative views of Thai EFL learners' learning oral communication skills. *Academic Research International* 1(1), 90-103.
- Kim, H. S., Y. Cha and N. Y. Kim. 2020. Impact of mobile interactions with AI on writing performance. *Modern English Education* 21(2), 1-13.
- Kim, H. S., N. Y. Kim and Y. Cha. 2021. Is it beneficial to use AI chatbots to improve learners' speaking performance?. *The Journal of Asia TEFL* 18(1), 161-178.
- Kim, N. Y. 2016. Effects of voice chat on EFL learners' speaking ability according to proficiency levels. *Multimedia Assisted Language Learning* 19(4), 63-88.
- Kim, N. 2017. Effects of different types of chatbots on EFL learners' speaking competence and learner perception. *Cross Cultural Studies* 48, 223-252.
- Kim, N.-Y. 2018. Voice blogging versus text blogging: Effects of blogs on Korean EFL learners' oral proficiency, *Studies in English Language & Literature* 44(1), 175-197.
- Kim, N.-Y., Y. Cha and H.-S. Kim. 2019. Future English learning: Chatbots and artificial intelligence. *Multimedia Assisted Language Learning* 22(3), 32-53.
- Kötter, M. 2001. Developing distance language learners' interactive competence: Can synchronous audio do the trick? *International Journal of Educational Telecommunications* 7, 327-353.
- Levelt, W. J. M. 1989. *Speaking. From Intention to Articulation*. Cambridge, MA: ACL-MIT Press.

- Lintunen, P., P. Peltonen and J. Webb. 2015. Investigating English pronunciation. In J. A. Mompean and J. Fouz-Gonzalez, eds., *Tone Units as Indicators of L2 Fluency Development: Evidence from Native and Learner English*, 196-218. London: Palgrave Macmillan.
- Matthews, C. 1993. Grammar frameworks in intelligent CALL. *CALICO Journal* 11(1), 5-27.
- Raba, A. A. A. 2017. The influence of think-pair-share (TPS) on improving students' oral communication skills in EFL classrooms. *Creative Education* 8(1), 12-23.
- Roed, J. 2003. Language learner behavior in a virtual environment. *Computer Assisted Language Learning*, 16(2-3), 155-172.
- Rosell-Aguilar, F. 2005. Task design for audiographic conferencing: Promoting beginner oral interaction in distance language learning. *Computer Assisted Language Learning* 18(5), 417-442.
- Shawar, B. A. and E. S. Atwell. 2005. Using corpora in machine-learning chatbot systems. *International Journal of Corpus Linguistics* 10(4), 489-516.
- Skehan, P. 2003. Task-based instruction. *Language Teaching* 36(1), 1-14.
- Stockwell, G. 2004. Communication breakdown in asynchronous computer-mediated communication. *Australian Language and Literacy Matters* 1(3), 7-31.
- Sykes, J. M. 2005. Synchronous CMC and pragmatic development: Effects of oral and written chat. *CALICO Journal* 22(3), 399-431.
- Thamarana, S. 2015. A critical overview of communicative language teaching. *International Journal of English Language, Literature and Humanities* 3(5), 90-100.
- Ur, P. 1996. *A Course in Language Teaching, Practice and Theory*. Cambridge: Cambridge University Press.
- Walker, A. and G. White. 2013. *Technology Enhanced Language Learning: Connecting Theory and Practice*. Oxford, UK: Oxford University Press.
- Wang, Y. F. and S. Petrina. 2013. Using learning analytics to understand the design of an intelligent language tutor – Chatbot Lucy. *Editorial Preface* 4(11), 124-131.
- Warschauer, M. 2001. On-line communication. In R. Carter and D. Nunan, eds., *The Cambridge Guide to Teaching English to Speakers of Other Languages*, 207-212. Cambridge: Cambridge University Press.
- Willis, D. and J. Willis. 2001. Task-based language learning. In R. Carter and D. Nunan, eds., *The Cambridge Guide to Teaching English to Speakers of Other Languages*, 173-199. Cambridge: Cambridge University Press.
- Yanguas, I. 2010. Oral computer-mediated interaction between L2 learners: It's about time! *Language Learning & Technology* 14(3), 72-93.
- Zhang, Y. 2009. Reading to speak: Integrating oral communication skills. *English Teaching Forum* 47(1), 32-34.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary