



An AI-based English education platform during the COVID-19 pandemic

Hansuk Um¹ · Hisam Kim² · Dain Choi³ · Hyungna Oh⁴

Accepted: 31 August 2023 / Published online: 6 October 2023
© The Author(s) 2023

Abstract

This study examines whether the use of AI-Pengtalk, an AI-based Conversational English programme, provided by a broadcasting company (EBS) that specializes in public education can significantly improve conversational English skills and bridge the English language proficiency gap associated with parental socioeconomic status. Over the course of four weeks from April 27 to May 22 in 2020, 108 fourth-grade classes in 54 elementary schools voluntarily participated in this experiment. Two classes in each school were designated as a treatment group and a control group. For the treatment group, a tablet installed with a pilot version of AI-Pengtalk was provided and students were encouraged to make use of the programme. Two sets of surveys and English tests were placed pre and post hoc. After 4 weeks, test scores, log files, and survey responses of participants were analysed. A series of DID analyses demonstrate that the use of AI-Pengtalk improves the treatment group's self-evaluation of their English abilities, confidence in using English, preference on English itself, and amount of time spent on studying English during the pilot experimental period compared to the control group. When other variables were controlled, the use of AI-Pengtalk also helped the treatment group achieve higher test scores. This study implicates that the use of smart English education like AI-Pengtalk may especially be able to better compensate for academic setbacks caused by low parental SES or, in the case of English learning, the reluctance to converse in English with other students.

Keywords Data science applications in education · Distance education and online learning · Human–computer interface · Mobile learning

1 Introduction

English is not a primary language in Korea. As a result, very few people find themselves using English for communication purposes in their daily lives, which results in English being taught (particularly in the early stages of education) purely in the context of an academic subject. Korea's primary education is provided on both a public and private basis, and each type provides a different quality of education. In the case of English, public schools often utilize Korean teachers lecturing to students in a non-native context. However, private schools can afford to hire native English teachers while also providing a wide array of extracurricular activities to further foster learning. Due to this discrepancy in budgets, private schools are more expensive than public schools. Consequently, children's English proficiency is influenced by parental socioeconomic status (henceforth, SES). This gap is larger for other subjects, such as mathematics [1, 2], but in the context of students' English skills, it widens as further English education takes place outside the regular curriculum

✉ Hyungna Oh
h.oh@khu.ac.kr

Hansuk Um
hsum@ebs.co.kr

Hisam Kim
hisamkim@gist.ac.kr

Dain Choi
dainchoi@kdi.re.kr

¹ Korea Educational Broadcasting System, Seoul, Republic of Korea
² Gwangju Institute of Science and Technology (GIST), Gwangju, Republic of Korea
³ Korea Development Institute, Sejong, Republic of Korea
⁴ Kyung Hee University, Yongin, Republic of Korea

of schools, such as in private institutions or homes. Many Korean children learn English at private institutions with full financial support from their parents, which may occur before a student even enters elementary school and tends to continue alongside the regular curriculum.

EBS, the South Korean public educational broadcasting service, has been making continuous efforts to compete with expensive private education by providing quality online education for free. It provides recorded lectures on many subjects, including but not limited to English. This feature of EBS came into the spotlight as the need for online education increased with the COVID-19 outbreak. During the onset of the pandemic, many schools were forced to switch from offline classes to e-classes to prevent the spread of COVID-19. However, most schools at the time lacked the capacity and infrastructure to offer quality online classes. To prevent the complete suspension of public education, the government requested that EBS provide online class materials for elementary, middle, and high school students. However, online classes with recorded one-way lectures that lacked interaction between teachers and students had difficulty motivating or capturing the academic interest of students. In addition to concerns about the overall degradation of the quality of education, it was expected that the gap in students' learning incurred by differences in parental SES would be exacerbated during the pandemic. Such gaps were likely to become larger for elementary English education, which is largely influenced by students' parents. A study conducted in Japan in 2002 showed that the gap widened as Saturday classes in schools were cancelled, resulting in students spending more time at home [3].

Before the outbreak of COVID-19, EBS started a project to develop an artificial intelligence-based English tutoring platform called *AI-Pengtalk* with technical support from the Electronics and Telecommunications Research Institute (ETRI) and the NHN consortium. *Pengsoo*, the mascot of EBS that is very popular with children, is an AI English teacher that teaches English conversation through an education platform called *AI-Pengtalk*. *Pengsoo* is equipped with ETRI's perception and learning algorithms, and *AI-Pengtalk* adopted NHN's game-design elements and game principles.¹ This project was funded by the Ministry of Education in South Korea. The original EBS plan was to provide *AI-Pengtalk* as a supplement for English classes in schools. Due to the outbreak of the COVID-19 pandemic, *AI-Pengtalk* was introduced earlier than planned to supplement online English classes. Accordingly, a study to test the effectiveness of *AI-Pengtalk* was conducted ahead of the original plan. Through nationwide recruitment, 108 fourth-grade classes in 54 elementary schools voluntarily participated in this study.

In each school, one class was randomly selected to use *AI-Pengtalk* in English classes, while another class was told to use EBS's conventional online materials. The experiment was conducted over the course of four weeks, from April 27 to May 22, 2020.

This paper aims to investigate whether *AI-Pengtalk* contributes to improving students' English proficiency and attitudes towards English. It also aims to determine whether *AI-Pengtalk* mitigates the disparity in English ability caused by differences in students' endowment, including parental SES. Data obtained from the experiment, such as pre- and post-experiment English test scores, were used for the empirical exercise.

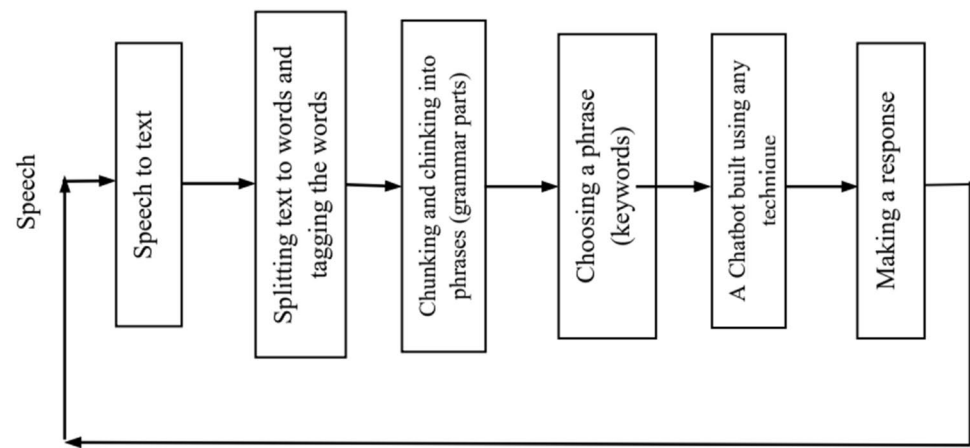
2 Related literature

Artificial intelligence (AI), enhanced with the recent development of big data and machine learning, has been applied to education services. AI that has the explicit purpose of tutoring has various names. For example, a chatbot, a portmanteau of “chat” and “robot”, is a computer programme with the ability to connect with a human using natural language speech [4]. Early chatbots were chiefly text-based programmes that understood and answered text entered by users. Recently, however, they have been developed into voice-based operations or multimedia files that can be used to communicate through a combination of STT (sound-to-text) and TTS (text-to-sound) technologies. Chatbots can provide immediate user feedback through dialogue during learning and customized content based on feedback, making interaction very easy. Because of these characteristics, they are mainly used in the field of education with a focus on language. Duolingo, the quintessential example, uses AI technology to conduct real-time conversations with people in specific situations. Eggbun conducts learning based on an interactive interface and is designed to enable feedback through preset menus and answers for word and sentence learning. Other types of chatbots currently being developed and used for language studies include Alexa (developed by Amazon), Google Assistant (developed by Google), and Cleverbot (developed by the British scientist Rollo Carpenter).

Recently, research has been conducted to determine the possibility of using AI-based software for learning English and its educational effectiveness [5–11]. Randall [12] provides an in-depth literature review on the effects of using robots in language learning. One of the most representative studies is Köse and Arslan [5], which was conducted among university students in Turkey, Italy, and Romania. In this study, researchers compared a control group in which English classes were conducted using only the conventional face-to-face method and an experimental group that used

¹ NHN is a Korean cloud company. NHN started its business as a game company, “*Hangame*”.

Fig. 1 The main steps of analysis and processing to perform human–computer conversation source: Abdul-Kader and Woods (2015)



source: Abdul-Kader and Woods (2015)

both the face-to-face method and AI-based online learning software, as shown in Fig. 1. They found that higher academic performance and test scores were achieved when AI-based online learning software was combined with the existing method [5].

The main advantages of AI-based learning software are that learners can learn autonomously at their own pace; learning materials are provided according to each learner's level; learning is possible anytime and anywhere and without time and space constraints, which is convenient for repetitive learning; and it can lead to increased motivation, an increased sense of ease and comfort in a learning environment, the stimulation of essential learning behaviours, and increased smoothness of information and communication processes [5, 6, 13–17]. However, AI chatbots continue to be limited in their capacity to train students' conversational English skills. It has been observed that AI chatbots often fail to recognize non-native speakers' English pronunciation and that conversations on free topics with AI chatbots remain limited due to the lack of availability of a knowledge database for dialogue [18–22].

Despite the rapid expansion of AI technology in second-language education, there are few studies that evaluate its effectiveness using the actual academic performance data of participants. Some studies use test scores to show the effect of robot-assisted language learning. However, the data used in these studies are apparently not generated by a randomized experimental design and/or include limited samples. For example, Ruan et al. [9] used 60 samples collected from voluntary participants at a university in China to statistically evaluate the learning benefits of *EnglishBot*. Wang et al. [23] used 327 samples collected from a primary school in China; therefore, their sample represented only a limited part of the entire population or peer group. To our knowledge, this study is the first to conduct a nationwide randomized experiment, collect students' academic performance data, and

empirically test whether using AI-based software improves students' conversational English learning.

3 Study design

3.1 *AI-Pengtalk*, a two-way English tutoring AI

AI-Pengtalk is a chatbot (chat robot) powered by AI that interacts with students, perceives their personality and language skills, and provides customized English conversation classes. As shown in Fig. 2, the popular mascot of EBS, *Pengsoo*, is introduced as a native English teacher and/or as a friend in *AI-Pengtalk*. According to Randall [12], humans' perception of the emotional expressions of AI plays a crucial role in human–robot interaction. *AI-Pengtalk* therefore employs *Pengsoo* and uses facial expressions and body movements to convey actions, intentions, and emotions.

AI-Pengtalk works in both online and offline educational environments to enable students to learn English regardless of their location. The content of *AI-Pengtalk* for fourth graders was developed based on the content of the five most commonly selected English textbooks in Korea. It is built on advanced machine learning technology, including artificial intelligence and voice recognition. Students can have real conversations with the 3D character “*Pengsoo*” and see the scores of their conversation practice immediately. *AI-Pengtalk* is linked with a scientific LCMS (Learning Content Management System), which uses learning big data to enable students to learn English with minimum effort (Fig. 3). When students interact with an AI-embodied tutor (software), they feel less nervous about making mistakes [9, 23, 24]. Furthermore, the app used in this study, *AI-Pengtalk*, encourages students to talk and practice repeatedly and regularly. This helps students build confidence in English over time [25].

Fig. 2 Pengsoo in “AI-Peng-talk”

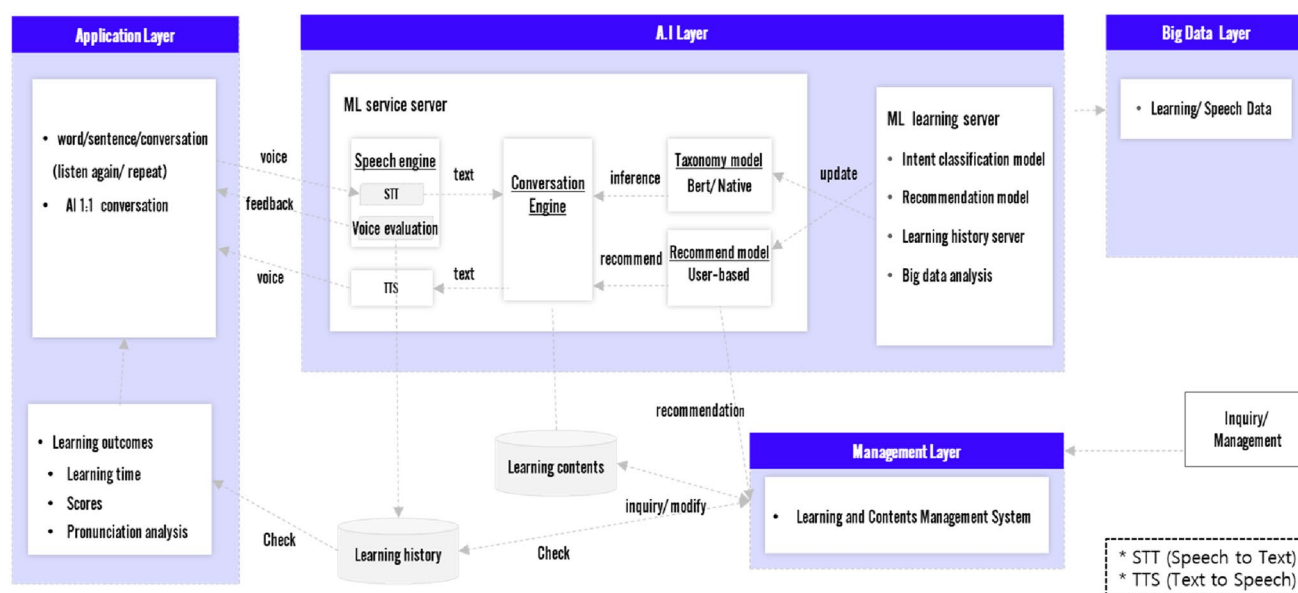


Fig. 3 AI-Pengtalk system

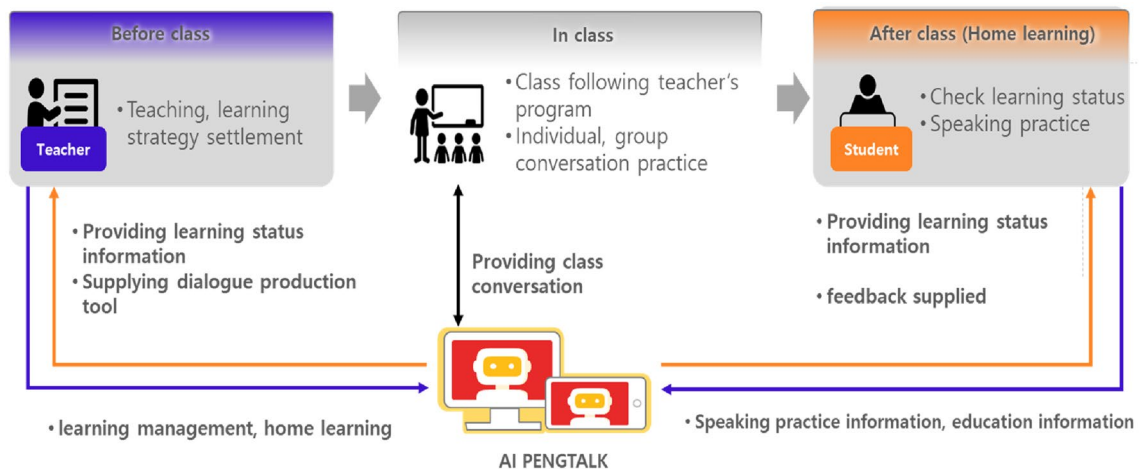
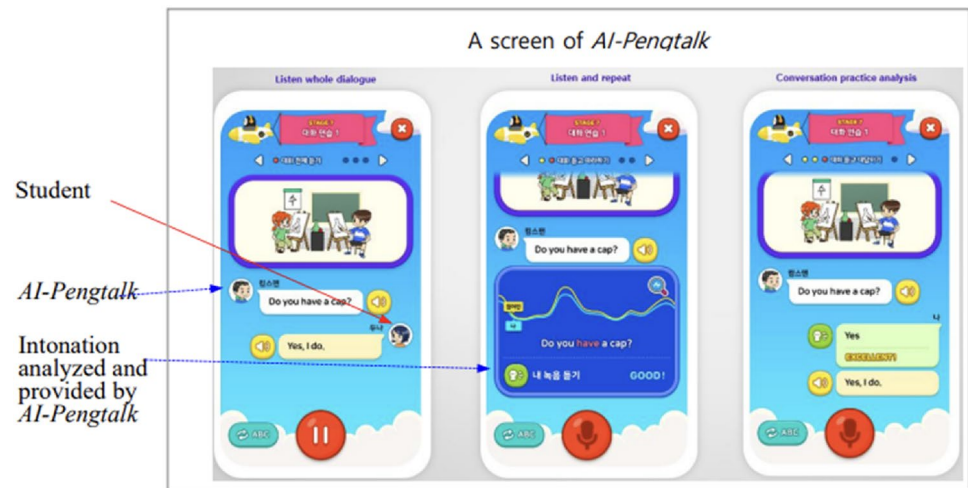
Intonation, one of the most difficult aspects for second-language English learners, can be improved by real-time visualization of a student's own intonation compared with a prerecorded native speaker (see Fig. 4).

The effectiveness of immediate feedback with visualization has been well addressed by previous studies [26, 27]. Teachers are able to closely monitor students' learning status through the teacher's portal on the website. Moreover, teachers can check the overall status of students' progress before class and plan upcoming classes accordingly. Teachers trace how students have studied with *AI-Pengtalk*, the duration of their study time, and their progress in more than

five categories, such as "Topic World", "Speaking", "Let's Talk", "Scan It", and "School Talk". During class, teachers can use *AI-Pengtalk* for both individual and group work. After class, students can continue their studies and reflect upon feedback from their teachers (Fig. 5).

3.2 Experimental design

We used a randomized control trial that randomly assigned *AI-Pengtalk* users. Figure 6 depicts the study design. Reflecting the regional distribution of all elementary schools nationwide, we set the target number of school participants for 18

Fig. 4 English conversation analysis in *AI-Pengtalk***Fig. 5** *AI-Pengtalk* learning system

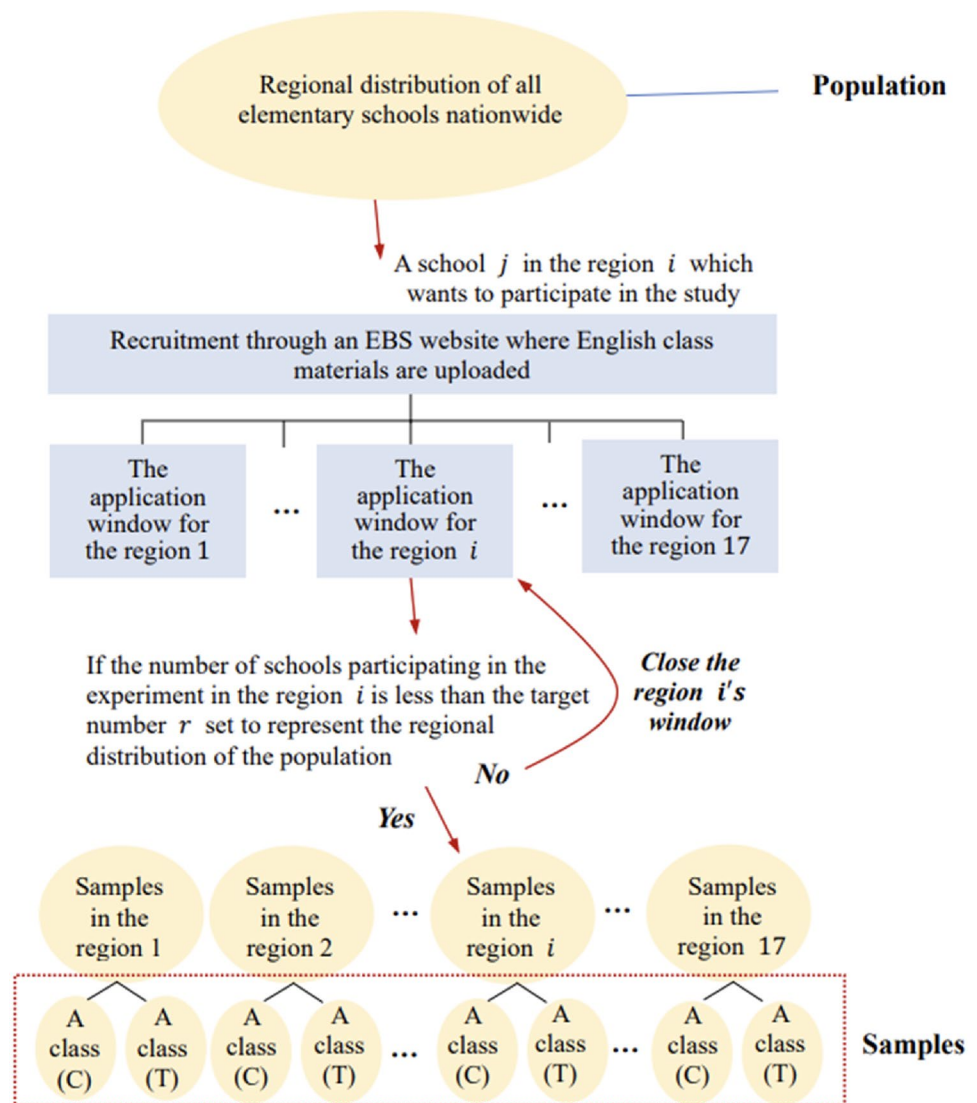
regions. There are 17 provinces in Korea, but we assigned 2 regions to Seoul given the high level of heterogeneity between the Gangnam and non-Gangnam areas of Seoul. With the help of the Ministry of Education and the use of EBS's school network, we recruited schools nationwide to voluntarily participate in the study. A school that was interested in this study could join the experiment through the EBS English website. When the target number of schools for a region was met, we closed the application window in the region. We ultimately found a total of 54 schools interested in taking part. In each school, two fourth-grade elementary classes taught by an English teacher were selected, one for the treatment group and the other for the control group. A total of 940 fourth graders across the country participated in our *AI-Pengtalk* experiment. Students in the treatment group (T) used *AI-Pengtalk* in English conversation classes. Students in the control group (C) used EBS's traditional online content in English conversation classes. This separation was

maintained during the study period from April 27 to May 22 in 2020. After the study period, *AI-Pengtalk* became available to every fourth-grade student in Korea.

Two sets of English tests were taken before and after the experiment² to measure students' level of improvement in English proficiency. In addition to test scores, responses obtained from the two surveys were used in our analysis. Surveys were conducted using Google Online Survey. The pre hoc survey included questions about attitudes towards English (such as confidence in English, subjective evaluation of English skills, interest, motivation, and desire for English skills), self-evaluated level of English proficiency, and previous personal efforts for improvement (such as study hours and English learning methods). The questionnaire was made by EBS with the

² Before the experiment began, the guardians (parents) of all students in both groups submitted consent for collection of students' personal information.

Fig. 6 Setting samples for this study



assistance of teachers participating in this study. The post hoc survey contained identical questions. In the follow-up survey, questions on the experience of using *AI-Pengtalk* were added for students in the treatment group. As shown in Table 1, not all participants completed the two sets of surveys and tests. In the control group, the proportion of first-test participants who took the second test was only 60.2%, and the proportion that completed the two surveys was even lower (51.2%). In the treatment group, the response rate was higher: the proportion of participants that took both tests was 70.6%, and the proportion that completed both surveys was 58.4%.

3.3 Sample characteristics and tests for randomization balance

The first plot of Fig. 7 compares the distributions of English conversation test scores for students in the control and the treatment groups before the experiment and shows

Table 1 The number of students who took part in surveys and tests

	Survey		Test	
	Before	After	Before	After
C group	241	234	457	275
T group	311	282	483	341

Note: The proportion of students who took part is shown in parenthesis

insignificant levels of difference in the test scores across the two groups. Table 2 summarizes the learning-related aptitudes and personal characteristics of students in both groups. The last column of Table 2 shows the t test results and confirms that the randomization was successful since there was no statistically significant difference in the attributes of participants' personalities across the two groups at baseline.

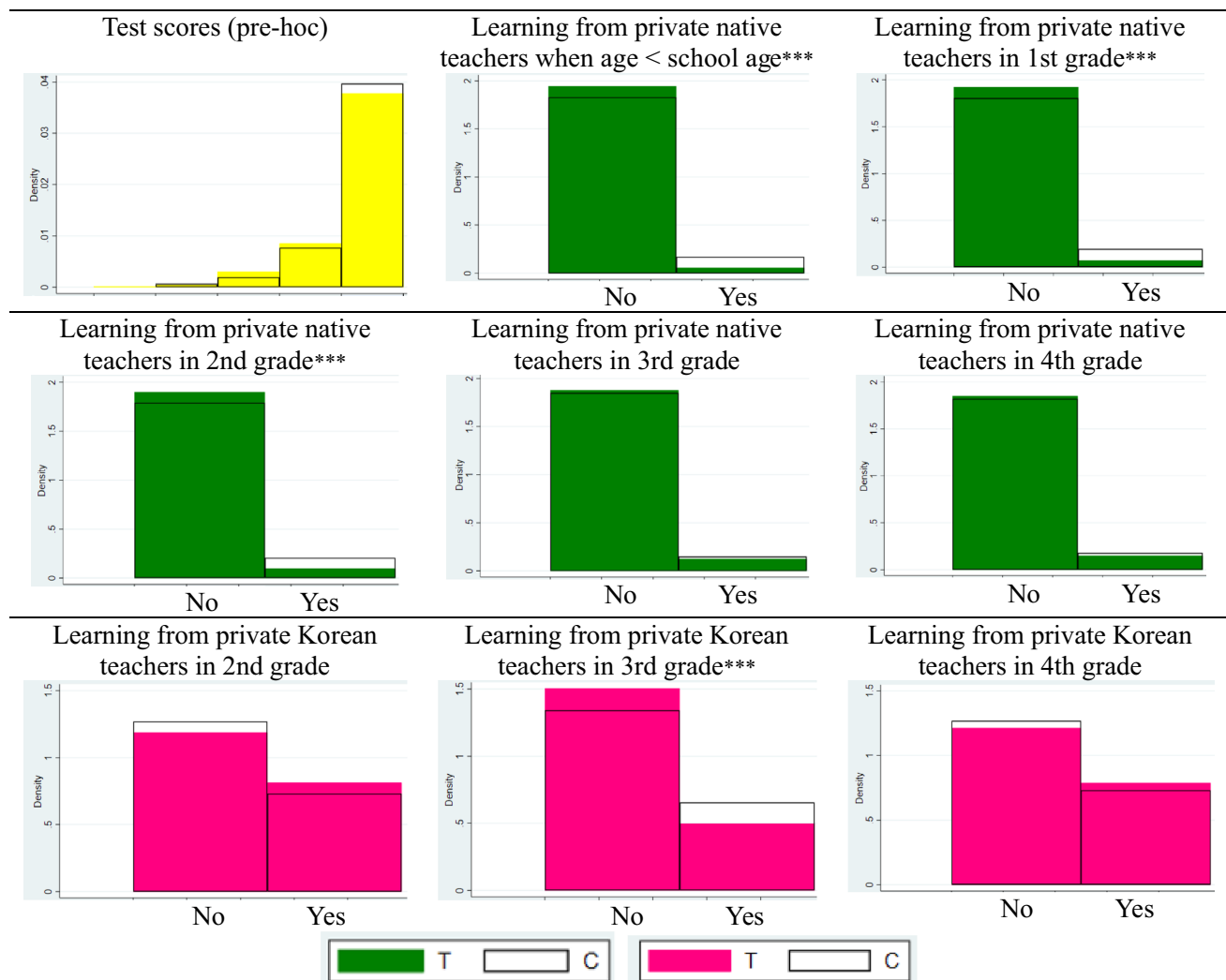


Fig. 7 Distribution of samples across treatment and control groups at baseline

Table 3 also summarizes the responses of the survey concerning whether participants had learned English conversation at school or outside the school before the experiment. As shown in the shaded area of Table 3, the proportion of students who learned English conversation with native or Korean English teachers either at school or at private institutes was not statistically the same across the two groups over the entire education period up until fourth grade. The last column of Table 3 shows the corresponding t test outcomes. With the exception of five cases, there were no significant differences in English-speaking learning experiences between the two groups. It would not be desirable to conduct an experiment on two groups with different prior learning experiences. However, the students' experiences with conversational English education immediately before the experiment (the 1st semester of fourth grade) were not significantly different between the two groups (see Fig. 7). Therefore,

if students with poor endowments, such as unfavourable parental SES, physical disability, or introverted personalities, showed improvement in their conversational English abilities after the use of *AI-Pengtalk* for four weeks, this would suggest that there is potential to overcome the gap in English proficiency associated with the presence of unequal endowment. The second null hypothesis was established to test this differentiated impact of *AI-Pengtalk*.

3.4 Testing hypothesis and estimation model

If *AI-Pengtalk* is an effective tool for teaching conversational English, students in the treatment group should show improvements in their test scores and in their attitudes towards English after the experimental period when compared to the control group. Therefore, changes in test scores, attitudes towards English, and self-confidence in English

Table 2 T test results: learning-related aptitude and personality

Items	C Group (C_0)	T Group (T_0)	Difference ($T_0 - C_0$)
I am interested in learning something or studying	3.59	3.50	−0.09
I like to talk	3.87	4.00	0.13
I am good at handling computers	3.14	3.10	−0.04
I am good at handling smart devices	3.72	3.67	−0.05
I am familiar to using apps for learning	3.42	3.46	0.04
I always try something new when I think it will be fun	4.01	3.99	−0.01
I am afraid of making a mistake	3.20	3.28	0.08
I tend to be shy	3.01	2.93	−0.08
I feel uncomfortable talking with my teacher for more than 20 min	2.36	2.23	−0.13
I do not want to do what I'm not good at in the presence of other people	3.23	3.17	−0.06
I feel uncomfortable meeting foreigners	2.46	2.59	0.12

Note: Responses were obtained on a 5-point scale (1=Strongly Disagree 2=Disagree 3=Unsure 4=Agree 5=Strongly Agree). The difference between groups is the mean of the experimental group minus the mean of the comparison group. The statistical significance level of the difference is * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. A difference without a significance sign (*) means that there is no significant difference between the two groups (this applies throughout this paper)

were set as variables of interest. We evaluated the effect of *AI-Pengtalk* in improving a student's English conversation skills. In Table 4, C_0 and C_1 are the respective expected test scores of a student in the control group before and after the experimental period. T_0 and T_1 are the counterparts of a student in the treatment group. A difference between the two periods, ($T_1 - T_0$), is attributed by two compound learning effects: the effect of conventional English classes and the effect of studying with *AI-Pengtalk* for a fourth-grade student in the treatment group. In other words, the net effect of *AI-Pengtalk* has not yet been isolated. A group difference after the experimental period, ($T_1 - C_1$), includes not only the effect of *AI-Pengtalk* but also the students' characteristics that can influence test scores.

The difference-in-differences (DID) estimator is frequently used to account for these confounders. DID is calculated through the formula $(T_1 - C_1) - (T_0 - C_0)$, as shown in Table 4, after obtaining the mean values of the variables of interest for two time periods (0 and 1) in the control and the treatment groups (C and T). Through this method, the net learning effect of *AI-Pengtalk* can be extracted. The null hypothesis (H_0) assumes that the DID value, $(T_1 - C_1) - (T_0 - C_0)$, is equal to zero. If the null hypothesis is not rejected, we can conclude that the use of *AI-Pengtalk* does not offer additional effects in improving the English conversation skills of participating students in the treatment group. If the null hypothesis is rejected and the DID value is significantly positive, it implies that the use of *AI-Pengtalk* improves students' English conversation skills.

$$H_0 : E[\bar{Y}_{t=1} - \bar{Y}_{t=0} | T = 1] = E[\bar{Y}_{t=1} - \bar{Y}_{t=0} | T = 0] \\ \rightarrow (T_1 - C_1) - (T_0 - C_0) = 0$$

$$H_A : (T_1 - C_1) - (T_0 - C_0) \neq 0$$

4 Analysis of experimental outcomes

4.1 *AI-Pengtalk* and English test scores

The effectiveness of using *AI-Pengtalk* in improving English test scores was examined by comparing the scores of pre and post hoc tests between the control and treatment groups. As shown in Table 5, both groups showed a large decrease in average scores in the post hoc test. The pre hoc test was set to evaluate students' English proficiency at a third-grade level, whereas the post hoc test was set to evaluate proficiency at the level of a 1st-semester fourth-grader and was therefore substantially more difficult. Hence, it is possible to observe drops in the groups' average scores. As shown in Table 5, the average score of the control group decreased by 22.1, while the average score of the treatment group decreased by 19.7. The average score of the treatment group decreased by less than that of the control group by 2.4, although the better performance demonstrated by *AI-Pengtalk* users was statistically insignificant (Fig. 8).

Table 3 T test outcomes: the proportion of students with experience in conversational English education

	Item	C Group (C ₀)	T Group (T ₀)	Difference (T ₀ – C ₀)
Before entering elementary school (i.e. kindergarten)	Learning from Korean school teachers	0.15	0.15	– 0.01
	Learning from native school teachers	0.03	0.02	– 0.01
	Learning from Korean private teachers	0.31	0.31	0.00
	Learning from native private teachers	0.10	0.03	– 0.07 ***
In the 1st grade of elementary school	Learning from Korean school teachers,	0.12	0.16	0.05 *
	Learning from native school teachers	0.07	0.05	– 0.03
	Learning from Korean private teachers	0.29	0.26	– 0.03
	Learning from native private teachers	0.11	0.04	– 0.07 ***
In the 2nd grade of elementary school	Learning from Korean school teachers	0.07	0.11	0.04
	Learning from native school teachers	0.04	0.02	– 0.02
	Learning from Korean private teachers	0.41	0.42	0.01
	Learning from native private teachers	0.12	0.05	– 0.06 ***
In the 3rd grade of elementary school	Learning from Korean school teachers	0.37	0.53	0.16 ***
	Learning from native school teachers	0.14	0.12	– 0.01
	Learning from Korean private teachers	0.37	0.26	– 0.11 ***
	Learning from native private teachers	0.08	0.06	– 0.02
In the 1st semester of the 4th grade of elementary school	Learning from Korean school teachers	0.23	0.27	0.04
	Learning from native school teachers	0.08	0.06	– 0.02
	Learning from Korean private teachers	0.41	0.41	0.00
	Learning from native private teachers	0.10	0.08	– 0.02

Note: The statistical significance level of the difference is * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4 DID: pre- and post-comparison between two groups

	Control group (C)		Treatment group (T)		DID value
	Before	After	Before	After	
Variable	C ₀	C ₁	T ₀	T ₁	(T ₁ – C ₁) – (T ₀ – C ₀)

a lot). After the experimental period, self-assessed levels of improvement were 0.34 for the treatment group, which were much larger than the levels for the control group (0.13). The DID value of 0.21 was positive and statistically significant enough to reject the null hypothesis at the 1 per cent significance level ($p < 0.01$). This is depicted in Fig. 9

We further investigated students' responses to seven

Table 5 The DID value: English Test Scores

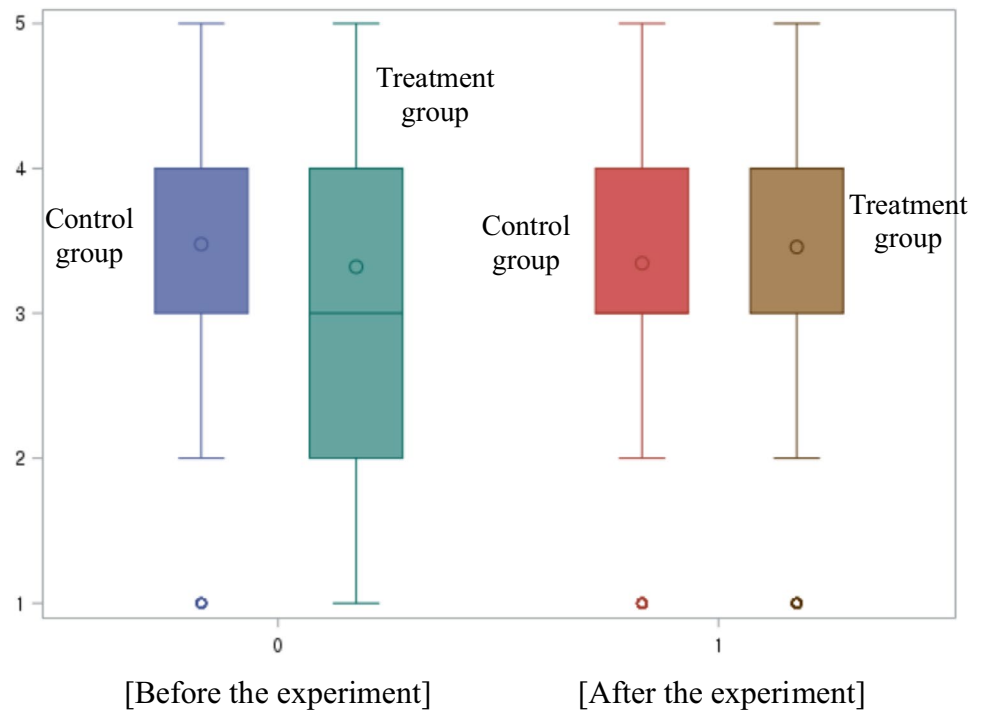
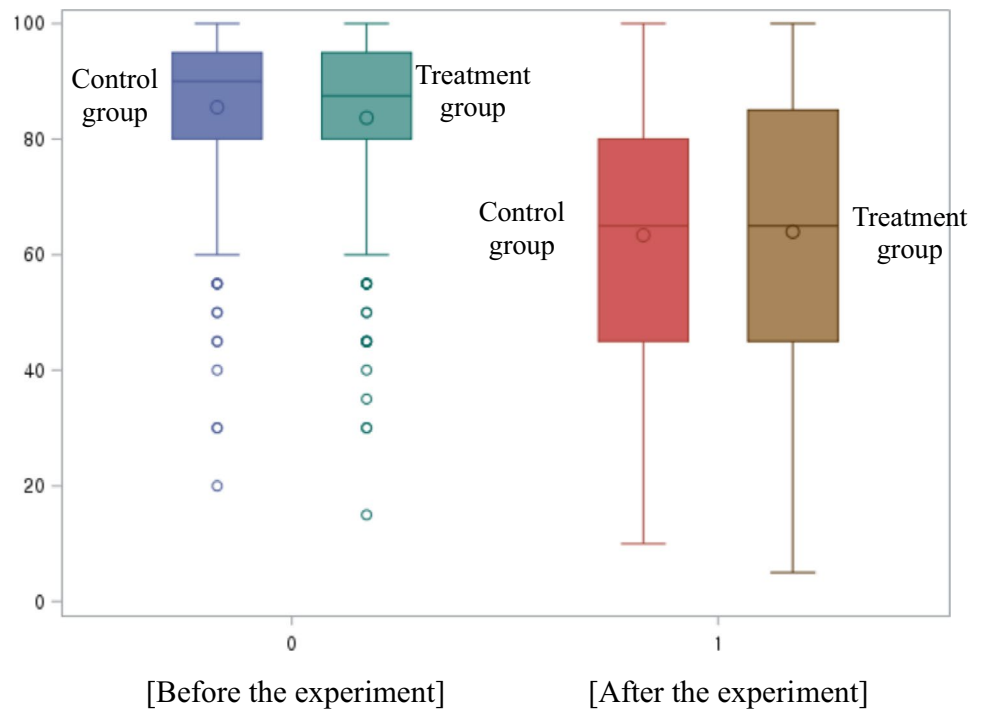
	Comparison group		Experimental group		DID value	
	Before	After	Before	After	Value ¹	P value
Test Scores: Assessment Result	85.49	63.40	83.69	64.00	2.41	0.411

Note: DID = (T₁ – C₁) – (T₀ – C₀)

4.2 AI-Pengtalk and self-evaluated improvement in English conversational skills

A DID approach was employed to test whether the use of *AI-Pengtalk* provided students with the feeling that their English conversational skills improved. To answer this question, data on students' self-assessments of their perceived levels of improvement in their English conversational skills obtained through surveys were used. The students' responses were measured on a 4-point scale (0 = not improved at all, 1 = slightly improved, 2 = somewhat improved, 3 = improved

questions asked in the pre- and post-surveys. Responses were obtained on a 5-point scale: 1—strongly disagree, 2—disagree, 3—uncertain, 4—agree, and 5—strongly agree. As shown in Table 6, which summarizes the questions and responses of the students, the degree of self-assessed improvement in conversational English skills was higher in the treatment group than in the control group. This finding is supported by the positive DID values shown in the last column of Table 7. This value was significantly positive, so we can reject the null hypothesis for three questions: “*I can*

Fig. 8 Box plots for English test scores**Fig. 9** Box plots for students' self-assessments of their English conversational skill improvement**Table 6** The DID value: Self-Evaluation of English Conversational Skill Improvement

	Comparison group		Experimental group		DID value	
	Before (C_0)	After (C_1)	Before (T_0)	After (T_1)	Value	P value
Self-evaluation on English Improvement over the Past Month	1.21	1.34	1.07	1.41	0.21	0.001

Note: The question with a 4-point scale was processed into a 0–3 point scale (0=Not Improved At All, 1=Slightly Improved, 3=Somewhat Improved, 4=Improved a Lot)

Table 7 DID values: self-assessed English-speaking ability

Item	Comparison (control) group		Experimental (treatment) group		DID value
	Before (C_0)	After (C_1)	Before (T_0)	After (T_1)	$(T_1 - C_1) - (T_0 - C_0)$
① I can listen and repeat simple English words	4.27	4.10	4.27	4.22	0.12
② I can listen and repeat simple English sentences	4.13	3.91	3.96	3.98	0.23 ***
③ I can introduce myself in three or more English sentences	3.64	3.52	3.54	3.55	0.14
④ I can say appropriate English greetings in the morning and at night	4.00	3.90	3.93	3.93	0.10
⑤ If I go to a grocery market in a foreign country, I can order and pay for five of my favourite fruits in English	3.41	3.26	3.23	3.31	0.22 **
⑥ If I go to the lost and found in a foreign country, I can ask if they have my lost blue hat in English	3.04	3.03	2.85	3.08	0.24 **
⑦ If I have a foreign friend, I can ask what sport he or she likes and how to play it in English	3.16	3.39	3.01	3.28	0.05

Note: Responses on a 5-point scale: 1 = Strongly Disagree 2 = Disagree 3 = Unsure 4 = Agree 5 = Strongly Agree

The statistical significance level of the difference is * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

listen and repeat simple English sentences"; "If I go to a grocery market in a foreign country, I can order and pay for five of my favourite fruits in English"; and "If I go to a lost and found in a foreign country, I can ask if they found my lost blue hat in English". It should be noted that the proportion of students who learned English from native teachers was lower for the treatment group than for the control group (see Table 3 and Fig. 7). Having students study with *AI-Pengtalk* may ameliorate this discrepancy.

4.3 *AI-Pengtalk* and English study hours

The pre and post hoc surveys asked students how many hours they spent studying English, including study hours at school or private institutions, online classes, and self-study hours. If the DID value is positive and statistically significant, we can reject the null and conclude that a student's English study hours will increase when he or she studies with *AI-Pengtalk*. Table 8 summarizes the average number of students' English study hours and the DID values. Since the number of study days included private tutoring days, which were assumed to have decreased in frequency due to COVID-19, a decrease in the number of English study days was expected. However, the use of *AI-Pengtalk* seems to have prevented a decrease in the number of days students spent studying English. While the average number of days the control group spent studying English decreased by 3.68, it decreased by only 2.34 for the treatment group. The average number of English study hours increased by 0.07 for the treatment group and decreased by 0.02 for the control group. Similarly, the average number of hours spent self-studying English increased by 0.09 for the treatment group but decreased by 0.02 for the control group. Every difference was statistically significant, so we rejected the corresponding

null hypothesis. These numbers demonstrate that the use of *AI-Pengtalk* is effective in increasing the time that students spend studying English.

Not all students completed the full four-week experiment. In this case, comparing the average test scores of the two groups was likely to obscure the effectiveness of *AI-Pengtalk*. Instead, we compared English study hours during the experimental period. We obtained these data by examining the log files of the participants in the control and treatment groups using EBS's conventional English class website and the *AI-Pengtalk* platform, respectively.³ As shown in Table 9, the number of hours spent studying was longer for the treatment group (4,123 s on average, 1.15 h) with *AI-Pengtalk* compared with users of conventional classes (363 s on average). Within the treatment group, the standard deviation of the students' *AI-Pengtalk* usage time was very large. There was a student who, with a minimum value of 10 s, did not actually use the service, and there was also a student with a maximum value of 42,118 s (11.7 h). Therefore, the effectiveness of *AI-Pengtalk* was remeasured using a regression model with value-added specification. As emphasized in [28], the concept of value added is highly useful in the modern economics of education because an increase in study hours implies increased input in the education production function. Table 9 shows the relationship between post hoc test scores and *AI-Pengtalk* usage time.⁴ When controlling for pre hoc test scores, the post-assessment scores increased by 2.34 points for each one-hour increase in *AI-Pengtalk*

³ Upon the request of teachers who participated in the experiment, regardless of whether students were in the experimental group or the control group, they were allowed to use *AI-Pengtalk* after the end of the experiment. We collected log files from May 25 to May 31.

⁴ For convenience of interpretation, the unit was converted from seconds to hours.

Table 8 Comparisons of English study time between the two groups

	Control group (prior)	Control group (post)	Treatment group (prior)	Treatment group (post)	DID value	P value
Average Number of Study Days	12.85	9.17	11.63	9.29	1.33	0.057
Average Hours of Study	0.93	0.91	0.81	0.88	0.08	0.066
Average Hours of Self-Study	0.65	0.63	0.59	0.68	0.11	0.004

Table 9 *AI-Pengtalk* Use Time (unit: second)

	Observed values	Mean	Standard deviation	Minimum	Maximum
Control group	457	363	1,572	0	23,103
Experimental group	369	4,124	5,358	10	42,118

usage (see Model (1)). Compared to Model (1), Models (2) and (3) employed additional explanatory variables such as learning-related aptitude.⁵ The estimation results from Models (2) and (3) also support the effectiveness of study time using *AI-Pengtalk* in improving English test scores Table 10.

4.4 *AI-Pengtalk* and students' attitudes towards English

A positive attitude towards English is important for students to study steadily and achieve long-term learning outcomes. In this regard, students responded to the following six items: “*I am good at English*”; “*Studying English will help me do what I want to do later*”; “*I am interested in studying English*”; “*I do my best when I study English*”; “*I like to speak English*”; and “*I want to improve my English skills a lot*”. Table 11 reports the summary statistics of the students' responses and the DID values. Before the experiment, the control group generally had more positive attitudes towards English than the treatment group. However, the positive attitudes of the control group became weaker for all questions after the experiment. Students in the treatment group showed similar changes for most questions, but the downwards slopes were smaller than those of the control group. The DID values were uniformly positive, and four were

statistically significant to reject the null hypothesis. These results imply that *AI-Pengtalk* helped students have positive attitudes towards English.

5 Conclusion

5.1 Summary

This paper examined whether *AI-Pengtalk*, an AI-based conversational English programme, is useful in learning conversational English and narrowing the English ability gap associated with differences in parental SES. To test this, 108 fourth-grade classes in 54 elementary schools voluntarily participated in an experiment over the course of four weeks from 27 April to 22 May 2020. One of the two classes in each school was designated as the treatment group, while the other was designated as the control group. For the treatment group, a tablet with *AI-Pengtalk* installed was provided. Two sets of pre and post hoc surveys and English tests were conducted. After four weeks of using *AI-Pengtalk*, the test scores, log files, and survey responses of the participants were analysed. A series of DID analyses demonstrated that the use of *AI-Pengtalk* improved students' self-confidence in their English skills and overall interest in English and increased their amount of English study time.

This study shows that *AI-Pengtalk* is worth applying to customized learning and/or remote education. *AI-Pengtalk* is designed to evaluate a student's English proficiency level, adjust the class accordingly, and enhance learning with a pedagogy customized to each student. If students feel something is lacking after class, repetitive learning is possible in the *AI-Pengtalk* app. This is even more valuable in abnormal situations such as the onset of COVID-19, which involves a

Table 10 *AI-Pengtalk* use time (unit: hour) and English test scores after the experiment

	(1)	(2)	(3)			
	Estimated Coefficient	t value	Estimated Coefficient	t value	Estimated Coefficient	t value
Intercept	− 10.961	− 2.22**	− 11.358	− 1.48	− 14.927	− 1.50
Prior Test Score	0.923	15.07***	0.801	11.71***	0.749	10.03***
<i>AI-Pengtalk</i> Use Time	2.335	5.13***	2.056	3.84***	2.153	4.29***
<i>R</i> ²	0.459		0.515		0.549	

⁵ Table 11 shows only part of the estimation outcomes of Models (2) and (3). Appendix A provides the entire results. The statistical significance level of the difference is *p < 0.1, **p < 0.05, ***p < 0.01

Table 11 Students' attitudes towards English

	Control group		Treatment group			DID value	
	Before	After	Before	After	Value	P value	
I am good at English	3.48	3.35	3.32	3.46	0.27	0.007	
Studying English will help me do what I want to do later	4.30	4.01	4.26	4.20	0.22	0.014	
I am interested in studying English	3.66	3.38	3.54	3.43	0.17	0.051	
I do my best when I study English	3.90	3.85	3.91	3.88	0.02	0.797	
I like to speak English	3.27	3.17	3.23	3.19	0.06	0.521	
I want to improve my English skills a lot	4.58	4.42	4.50	4.46	0.12	0.093	

considerable need for non-face-to-face education with social distancing policies. This smart education can help underperforming or introverted students learn the basics through a non-face-to-face method, and it can motivate students through immediate feedback and video game-like immersion. AI-based customized education can provide greater help, especially to academically low-achieving students and/or students with low levels of parental SES. Therefore, this study advocates the introduction of such a system in schools through public support⁶ to increase English education efficiency and narrow the educational gap.

5.2 Discussion and limitations of this study

Young students from low-income households or families living in less-developed rural counties are unlikely to have opportunities to learn English conversation with native English tutors. This results in an English conversation capacity gap between students with high parental SES and those with low SES. Our study finds that the effect of using *AI-Pengtalk* to improve English conversational learning is larger for students without native English-speaking tutor experience than for their counterparts. Given that education is one of the most representative public goods, it is recommended that the government expand the use of AI technology in English education to achieve one of the Sustainable Development Goals, “*Leaving No One Behind*”. It should also be noted that students' endowment includes not only their parents' SES but also their physical conditions and/or social skills. Compared to offline learning, obstacles faced by students with disabilities are quite marginal. Additionally, the participants of this study mentioned that they were more comfortable having English conversations with *AI-Pengtalk* than with a human [9, 23, 24]. The emotional benefits were specifically larger for students who identified themselves as being shy.

As suggested in a previous study [30], *AI-Pengtalk* can provide more personalized learning and guidance to students with special needs.

English conversation is one of the most difficult subjects to teach with a robot because there are many unstructured elements and no predetermined conclusions. The effectiveness of robotic teaching in this area means that there is great potential for robots to improve learning in other subjects as well. Especially in subjects with many predetermined outcomes, such as math, physics, and chemistry, robot teachers can enhance learning by visualizing or repeating existing knowledge at a pace appropriate to the student's level. However, this does not mean that a robot teacher can replace human teachers entirely. Robotic teachers are effective for students who are already self-regulated or self-motivated but are less effective for those without these characteristics. In addition, native language education is an area that does not show promise with the use of AI. This is because habit formation and repetitive practice, which are essential to language learning, are naturally supplied in everyday life in the case of the mother tongue.

This study has several limitations. The post-outcomes used in this study were test scores and responses to a post-survey conducted after the four-week experiment. Therefore, this study tested the impact of the use of *AI-Pengtalk* on short-term changes in students' English conversational performance and their attitudes towards English conversation. In other words, long-term effects were not included in the scope of this study. In addition, as *AI-Pengtalk* provides students with various learning modules, it is expected that its effectiveness depends on students' module choices. This study did not take this into account when evaluating the impact of *AI-Pengtalk*. Therefore, follow-up studies are needed to increase the credibility of the research findings by measuring whether the actual AI usage data and actual learning results match [12]. We plan to address these topics in future studies and to provide guidelines that can be used for developing AI-assisted language learning software.

⁶ The Ministry of Education [29] announced its plan to provide customized education in Korean, English, and math using artificial intelligence to elementary school students to prevent learning loss and gap widening due to COVID-19. AI-based adaptive learning will also be implemented in public education in Korea.

Appendix A

Improvement in English test scores on formative assessment according to *AI-Pengtalk* use time (unit: hours).

	(2)		(3)	
	Estimated coefficient	<i>T</i> value	Estimated coefficient	<i>T</i> value
Constant term	− 11.358	− 1.48	− 14.927	− 1.50
Scores of prior diagnostic assessment	0.801	11.78	0.749	10.03
AI-Pengtalk use time	2.056	3.84	2.153	4.29
Female students	1.835	1.00	2.028	0.96
I am interested in studying	1.829	1.66	1.582	1.46
I like to talk	1.236	0.98	1.138	0.85
I am good at handling computers	2.204	2.45	2.479	2.68
I am good at handling smart devices	− 1.059	− 1.10	− 1.193	− 1.11
I am familiar with using apps for learning	0.146	0.19	0.364	0.43
I always try something new when I think it will be fun	0.102	0.12	− 0.135	− 0.17
I am afraid of making a mistake	1.180	1.57	1.312	1.49
I tend to be shy	− 0.584	− 0.53	− 0.508	− 0.49

	(2)		(3)	
	Estimated coefficient	<i>T</i> value	Estimated coefficient	<i>T</i> value
I feel uncomfortable talking with my teacher for more than 20 min	− 0.473	− 0.69	− 0.129	− 0.17
I do not want to do what I'm not good at in the presence of others	− 0.637	− 0.85	− 0.641	− 0.79
I feel uncomfortable meeting foreigners	− 1.496	− 1.62	− 1.254	− 1.43
Learning English before entering an elementary school				
With school Korean teacher			− 0.280	− 0.11
With school native teachers			1.833	0.63
With private Korean teacher			4.436	1.62
With private native teacher			3.680	0.72
Learning English in the 1st grade of elementary school				
With school Korean teacher			0.224	0.07
With school native teachers			1.401	0.32
With private Korean teacher			1.044	0.38
With private native teacher			5.104	1.01

		(2)	(3)	
		Estimated coefficient	<i>T</i> value	Estimated coefficient
Learning English in the 2nd grade of elementary school	With school Korean teacher			1.042
	With school native teachers			1.822
	With private Korean teacher			2.457
	With private native teacher			2.254
Learning English in the 3rd grade of elementary school	With school Korean teacher			0.560
	With school native teachers			4.608
	With private Korean teacher			5.346
	With private native teacher			1.241
Learning English in the 1st semester of the 4th grade of elementary school	With school Korean teacher			3.939
	With school native teachers			−3.313
	With private Korean teacher			−0.437
	With private native teacher			1.994
Number of Observed Values	306		306	
Coefficient of Determination	0.515		0.549	

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Butler, Y., Le, V.: A longitudinal investigation of parental social-economic status and young students' learning of English as a foreign language. *System* **73**, 4–15 (2018)
- Oh, S., Kang, C., Jung, H., Joo, B.: Equality of opportunity for educational achievement in Korea: a study with the score of the college scholastic ability test. *Korean J Pub Finan* **9**(4), 1–32 (2016)
- Kawaguchi, D.: Fewer school days, more inequality. *J Jpn Int Econ* **39**, 35–52 (2016)
- Able-Kader, S., Woods, J.: Survey on chatbot design technique in speech conversation systems. *J Adv Comput Sci Appl* **6**(7), 72–80 (2015)
- Köse, U., Arslan, A.: E-learning experience with artificial intelligence supported software: an international application on English language courses. *GLOKALde* **1**, 61–75 (2015)
- Alhabbash, M., Mahdi, A., Abu Naser, S.: An intelligent tutoring system for teaching grammar english tenses. *Eur Acad Res* **4**(9), 7743–7757 (2016)
- Kim, H., Shin, D., Yang, H., Lee, J.: A study of AI chatbot as an assistant tool for school English curriculum. *J Learn Cent Curric Instrum* **19**(1), 89–110 (2019)
- Dodigovic, M.: Artificial intelligence and second language learning: an efficient approach to error remediation. *Lang. Aware.* **16**(2), 99–113 (2007)
- Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G. M., Brunskill, E. and Landay, J. A.: EnglishBot: an AI-powered conversational system for second language learning. *The 26th International Conference on Intelligent User Interface*: 434–444 (2021)
- Zhang, R., Zou, D.: Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Comput. Assist. Lang. Learn.* **35**(4), 696–742 (2022)
- Lin, C.J., Mubarak, H.: Learning analytics for investigating the mind map-guided AI Chatbot approach in an EFL flipped speaking classroom. *Educ. Technol. Soc.* **24**(4), 16–35 (2021)
- Randall, L.: A Survey of Robot-Assisted Language Learning. *ACM Trans Hum Robot Interact* **7**(1–7), 36 (2019)
- Fryer, L., Carpenter, R.: Bots as language learning tools. *Lang. Learn. Technol.* **10**(3), 8–14 (2006)
- Goda, Y., Yamada, M., Matsukawa, H., Hata, K.: Conversation with a chatbot before an online EFL group discussion and the effects on critical thinking. *Inf Syst Edu* **13**(1), 1–7 (2014)
- Kannan, J., Munday, P.: New trends in second language learning and teaching through the lens of ICT, networked learning, and artificial intelligence. *Circulo de Linguistica Aplicada a la Comunicacion* **76**, 13–30 (2018)
- Kim, T.: The effects of computer interest levels and chatting method (with AI chatting robot: chatterbot) on teaching and learning. *Eng Edu Res* **11**(4), 19–33 (2008)

17. Pokrivcakova, S.: Preparing teachers for the application of AI-powered technologies in foreign language education. *J. Lang. Cult. Educ.* **7**(3), 135–153 (2019)
18. Hill, J., Ford, R., Farreras, I.: Real conversations with artificial intelligence: a comparison between human-human online conversations and human-chatbot conversations. *Comput. Hum. Behav.* **49**, 245–250 (2015)
19. Shin, D.: Exploring the feasibility of AI Chatbots as a tool for improving learners' writing competence of English. *Korean J Teach Edu* **35**(1), 41–55 (2019)
20. Lee, D.: A developmental plan for an English conversation learning system through the application of a talking robot based on artificial intelligence. *Soc Teach Eng Through Med* **20**(1), 189–211 (2019)
21. Kim, N., Cha, Y., Kim, H.: Future english learning: chatbots and artificial intelligence. *Multimed Assist Lang Learn* **22**(3), 32–53 (2019)
22. Fryer, L.K., Ainley, M., Thompson, A., Gibson, A., Sherlock, Z.: Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Comput. Hum. Behav.* **75**, 461–468 (2017)
23. Wang, X., Pang, H., Wallace, M.P., Wang, Q., Chen, W.: Learners' perceived AI presences in AI-supported language learning: a study of AI as a humanized agent from community of inquiry. *Comput Assist Lang Learn* (2022). <https://doi.org/10.1080/09588221.2022.2056203>
24. Dizon, G.: Evaluating intelligent personal assistants for L2 listening and speaking development. *Lang. Learn. Technol.* **24**(1), 16–26 (2020)
25. Moussalli, S., Cardoso, W.: Intelligent personal assistants: can they understand and be understood by accented L2 learners? *Comput. Assist. Lang. Learn.* **33**(8), 865–890 (2020)
26. Tai, T.Y., Chen, H.H.J.: The impact of google assistant on adolescent EFL learners' willingness to communicate. *Interact Learn Environ* **31**(3), 1485–1502 (2023)
27. Dizon, G.: Using intelligent personal assistants for second language learning: a case study of Alexa. *Tesol Journal* **8**(4), 811–830 (2017)
28. Todd, P.E., Wolpin, K.I.: The production of cognitive achievement in children: home, school, and racial test score gaps. *J. Hum. Cap.* **1**(1), 91–136 (2007)
29. Ministry of Education: Plan to Strengthen Education Safety Network for All Students, Policy Brief on August 11, 2020 (2020)
30. Boccanfuso, L., Scarborough, S., Abramson, R.K., Hall, A.V., Wright, H.H., O'Kane, J.M.: A low-cost socially assistive robot and robot-assisted intervention for children with autism spectrum disorder: field trials and lessons learned. *Auton. Robot.* **41**, 637–655 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.