Research in English Language Pedagogy (2022) 10(4): 599-624

 DOI: 10.30486/RELP.2022.1946281.1327

Original Article

Using a Messenger Bot as a Tool for Providing Written Corrective Feedback: Examining L2 Development and Learners' Attitudes

Maryam Abdollahzadeh[1], *Mohammad Golshan*[1,*], *Hamid Allami*[2]

[1]Department of English, Maybod Branch, Islamic Azad University, Maybod, Iran.

[2]Department of English, Tarbiat Modarres University, Tehran, Iran.

Submission date: 04 December, 2021 Acceptance date: 18 January, 2022

Abstract

The current study aimed to examine the effect of written corrective feedback (WCF) provided via a messenger bot on Iranian English as a foreign language (EFL) learners' use of definite/indefinite articles. It also explored learners' attitudes toward using the messenger bot, which worked as a medium for providing WCF on the TOEFL iBT speaking task II. Seventy male and female TOEFL candidates (mean age: 26, range: 18-34) chosen through convenience sampling participated in the study. Before the treatment, the participants in all groups were given a proficiency test and a speaking pre-test. Then, they were randomly assigned to one of the three groups: one received WCF on their errors in the use of articles through the bot (MB group, n=30); the second one received WCF in traditional classes (TF group, n=20); and the participants in the third group received no feedback and followed the normal course of instruction (NF group, n=20). Subsequently, all groups participated in a post-test. The results of the mixed between-within subjects, ANOVA, showed that the MB group significantly outperformed the other two groups and there was not a significant difference between the TF and NF groups. The participants' attitudes toward receiving WCF through the bot were elicited through in-depth semi-structured interviews. The transcribed interviews were then analyzed using the first two phases of the grounded theory, and two main themes of pedagogical applications and technological issues emerged. The implications stemming from the findings concerned the efficacy of using the messenger bot in providing corrective feedback.

Keywords: CALL, Definite/Indefinite Articles, Grounded Theory, Messenger Bot, Written Corrective Feedback

Corresponding Author's E- mail: mohammadgolshann@gmail.com

1. Introduction

In recent decades, technology has developed rapidly and become the expected means of communication and education. The realm of language teaching and learning is no exception. Computers, cellphones, the Internet, and social media can greatly assist teachers who can make the most efficient use of technology to enhance second language teaching and learning. Computer-assisted language learning (CALL) is considered as one of the processes in which the technology is used to improve or facilitate learning a language (Beatty, 2003). Egbert (2005) defines CALL as "learners learning a language in any context with, through, and around computer technologies" (p. 4). This definition, as Kern (2006) asserts, not only prioritizes learners learning languages, but also broadens the potential types of relationships between computer technologies and language learning.

Learning in an engaging digital environment could have beneficial effects on language pedagogy. It increases learners' motivation and concentration while carrying out the learning tasks (Malouf, 1988; Papastergiou, 2009). Besides, technology-based instructions could decrease the cognitive load and enhance the retention of learning materials (Mayer & Moreno, 2010; Ricci et al.,1996; Williams & Zahed, 1996). Another advantage of using technology is reducing the involvement of instructors and teachers and the burden of their responsibility (Andreev et al., 2009; Athanaselis et al., 2014). Integrating modern technology with language instruction could be to language learners' benefit as well.

In the CALL platform, bots are one of the recently used technological tools implemented for language teaching and learning. A bot is a generalized term used to describe any software that automates a task. They are designed to interact with users through text, voice, or image (Abushawar & Atwell, 2007; Bii, 2013). They can be implemented in different messaging frameworks, such as Facebook Messenger, Telegram, Slack, and so on (Schmulian & Coetzee, 2018). Bots that interact with users through Messenger, not a standalone platform, are called messenger bots (Schmulian & Coetzee, 2018). In other words, a messenger bot (chatbot) is a piece of software that communicates with users, understands what is being asked, and can formulate a reply humanly. Nowadays, messenger bots are used worldwide for different purposes, including e-commerce, medication and health, and education. Some chatbots have been designed and implemented for language teaching and learning, which are fully explained in the next section. According to Kim et al.

(2019), language instructors should be able to select suitable and appropriate chatbots for their learners in their teaching environment based on the learners' needs and learning styles.

As mentioned, some chatbots have been designed for various aspects of language teaching and learning. They are implemented to enhance learners' skills. Since providing learners with corrective feedback (CF) is a vital concept in most of the theories of language teaching and learning, and learners can reflect on their learning and improve their skills through focusing on CF, such innovative ways, i.e., messenger bots, can be used for this purpose. Given the fact that the effect of CF on learning has been documented empirically (Black & Wiliam, 1998; Hattie &Timperley, 2007; Shute, 2008), a question that can be of pedagogic interest is whether the use of chatbots for the provision of CF can lead to more positive outcomes in comparison to the traditional approaches.

Although different messenger bots have been designed for language learning, almost no messenger bot has been designed to provide students with WCF and act as a mediator between the teacher and students. Therefore, the present study aimed to investigate the effect of WCF provided via a newly designed messenger bot on learners' accurate use of the target language (i.e., definite/indefinite articles). It also explores the learners' attitudes toward using the messenger bot to improve it to be applied for various skills and different kinds of CF in language teaching and learning.

2. Literature Review

From Vygotskian socio-cultural perspective (1978), cognitive development takes place in social interaction, whereby an expert member of the society such as parents, teachers, or more knowledgeable peer provides appropriate forms of assistance to a novice. However, we consider that not all forms of assistance can be useful and cause development. A key aspect of the socio-cultural theory is mediating tools. Vygotsky asserts that human cognitive development is mediated by culturally constructed means or tools that can be material (e.g., computers) or symbolic (e.g., language, gestures). In the case of WCF, tools can impact the provision of feedback and the learners' processing of the feedback.

The use of messaging apps as a learning tool can improve the quality of learning (Chuang & Tsao, 2013; Rambe & Bere, 2013; So, 2016; Sun et al., 2018). Some researchers (Bowman et al., 2010; Junco & Cotten, 2011) believe that using such apps may cause distraction due to multitasking and dealing with unrelated messages. Others believe that

these apps can be advantageous due to being portable and popular among young people (Klimova & Maresova, 2017). However, a few computer apps were designed to provide CF as one of the main aspects of language teaching. A study with CF via computers was conducted. There were 200 participants to whom elaborated feedback was provided. The results indicated that the use of technology affects motivation and improvement of learning outcomes. Concerning mobile apps, there are almost no mobile apps with an immediate CF that allows language learners to identify and correct their errors in different skills (Guanuche et al., 2020).

Besides mobile apps designed and implemented in education, some chatbots have been recently used for teaching and learning. Although bots offer differentiated instruction (Pokatilo, 2016; Riel, 2016), a few messenger bots have been designed for educational purposes, and a few studies have been conducted to examine the effect of messenger bots and chatbots on language learning (Akcora et al., 2018; Bii et al., 2013). Although there were some smart and developed chatbots, they did not provide pre-programmed answers. Some of them, such as Cleverbot, stimulated human-like conversation by comparing the received input to the possible responses and sending a text reply.

According to Jia and Ruan (2008), bots could be more motivating to language learners and they allow learners to feel more comfortable while using the new language, which can lead to more effective learning. Some studies investigated student perceptions such as motivation, involvement, confidence, enjoyment, and interest in using bots in English classes (Bii et al., 2013; Yoo, 2010). In Bii et al.'s study, students' attitudes toward using a chatbot named *Knowie* were explored, which seemed to be positive. The implication of this study was to demonstrate that chatbots could be implemented for various types of activities.

Two other chatbots, Alexa and Google assistant were used in a study by Kim et al. (2019). They investigated the difficulty of vocabulary use and the quality of the conversation with these chatbots. Both chatbots were highly satisfactory to the users and the researchers; however, Google assistant was easier to comprehend and more user-friendly.

Regarding the positive attitudes of learners toward using bots in language learning, all the bots are considered to possess pleasing features; however, none of them have had the role of a medium between teachers and their learners to provide WCF. They have just been chatbots with the capability of making oral or written productions and they have eliminated the presence of a teacher which has worked like a barrier to the simulation of the real

conditions of a learning environment. The main aim of the aforementioned bots has been to interact with learners as a partner. By considering these properties, the messenger bot utilized in the present article added to the capabilities of bots in language teaching and learning. Although there are a lot of studies that have indicated the effect of WCF on learning different aspects of language and there are some chatbots designed for language learning, to the researchers' knowledge, the effect of providing students with WCF through chatbots has not been investigated so far.

Therefore, the present study aimed to deal with the following research questions:

RQ1. To what extent does providing WCF through a messenger bot lead to learners' accurate use of definite and indefinite articles in their speaking?

RQ2. What do learners perceive as the benefits and disadvantages of using the designed messenger bot?

3. Method

3.1. Participants

The participants of the current study were 70 male (n=32) and female (n=38) Iranian EFL learners who were selected from the pool of 92 students through convenience sampling. The reason behind the procedure for selecting these participants was that they had some TOEFL speaking courses with one of the researchers, which made conducting the study practical. The participants' age range was from 18 to 34 (mean age: 26); they were all native speakers of Persian, and none of them had the experience of living in a foreign country. They intended to take the TOEFL iBT in some months. Some of them have had some experience of using applications for language learning, and some others have used certain apps to correct their writing mistakes.

3.2. Design and Context of the Study

A mixed-methods sequential explanatory design was applied in this study. The researchers used a pre-test-post-test control group design with seventy Iranian TOEFL iBT candidates who were randomly assigned to an experimental group (MB group), a comparison group (TF group), and a control group (NF group). NF group attended the traditional classes and received no CF. TF group received WCF in traditional form, and MB group received WCF through the newly designed messenger bot. In the quantitative phase of the study, the

independent variables were conditions regarding the medium for the provision of CF and time from the pre-test to the post-test. The dependent variable was the candidates' ability to use the target forms, which were definite and indefinite articles, from pre-test to post-test. In each treatment session, one TOEFL iBT speaking task II was answered by the candidates. In the qualitative phase of the study, the candidates in the MB group participated in an in-depth semi-structured interview and explained their ideas about the merits and drawbacks of the messenger bot and gave their opinions about how the messenger bot can be improved.

3. 3 Target Forms

To provide WCF in response to learners' errors, the English definite article "the" and indefinite article "a/an" were considered as target forms for the current study. The indefinite article "a/an" in this study was restricted to its function and referred to an object or a person for the first mention, while the function of the definite article "the" was limited to its function for referring to an object or a person for the second mention or the person the participants listened to in the listening section of the task. These target forms were selected because learners usually seem to have difficulty with their correct use in speech even up to advanced levels of proficiency (Nassaji, 2017). Moreover, from the researchers' experience, these target forms can be easily elicited in the TOEFL iBT speaking task II.

3.4. Instruments

To examine the potential effects of providing the learners with WCF through a messenger bot, the following instruments were employed.

3.4.1. Oxford Placement Test

The first instrument used in this study was the Oxford Placement Test (OPT) (Edwards, 2007). This test enabled the researchers to make sure the level of the participants was B2 and select those who were compatible with the conditions of the study. This test consisted of 50 multiple-choice language use items. Based on the results of the test, 26 students were excluded from the list of 92. Since 70 students were required to participate in the study, eight students were evaluated through OPT, and four with the same proficiency level were added to the previous ones.

3.4.2. Tests and Tasks

To investigate the learners' progress in using definite/indefinite articles and the role of the messenger bot in this regard, the researchers gave a pre-test, a post-test, and chose ten tasks from the TOEFL iBT speaking task II samples. All of these tests were taken from TOEFL Practice Online (TPO), which provides candidates with real past test questions.

3.4.3. Interview

The qualitative phase of the study was conducted through a semi-structured interview in which the participants in the MB group were asked some questions to explore their attitudes toward the messenger bot and their perceptions of its advantages and drawbacks of it. The interview questions were designed and later evaluated by a Ph.D. candidate of ELT and an assistant professor of Artificial Intelligence (See Appendix).

3.4.4. Messenger Bot

The main instrument of the study was the newly designed messenger bot. This bot is a user-initiative messenger bot that interacts with users in two modes concerning their roles, i.e., teacher and student. Since Telegram is one of the most popular messengers in Iran and the researchers and the students were in touch through this application, a finalized prototype using Telegram's bot API was implemented in the current study.

3.5. Data Collection Procedure

First, the learners took the Oxford Placement Test to guarantee the homogeneity of the participants (they were all at the B2 level of proficiency). Then, the TOEFL iBT speaking task II was explained to all the participants to make sure that they remembered how to provide the answer. Subsequently, they were assigned to three groups. Thirty students in the MB group were introduced to a newly designed messenger bot through which they were able to send their voices related to the tasks in the form of audio files to the researchers and receive WCF. The bot contained 10 TOEFL iBT speaking task II samples (TPO 45-54) which could be selected randomly to assure the variability of the tasks. Twenty learners in the TF group answered the tasks in a traditional classroom and received implicit and explicit WCF from the teacher. Similarly, twenty participants in the NF group experienced doing the tasks in a traditional classroom; nevertheless, they did not receive any kind of CF.

At the beginning of the study, a speaking pre-test was administered to assess learners' accurate use of definite and indefinite articles in TOEFL iBT speaking task II. In the same week, the treatment sessions began during five weeks (ten sessions), MB group utilized the messenger bot and they were asked to send TPO task IIs twice a week on specified days and received treatment in the form of WCF; whereas, TF group attended face-to-face classes twice a week and they were required to answer the tasks in class and received WCF from the teacher. What differentiated the NF group from the TF group was that the NF group received no feedback. All the feedback provided by the teachers was explicit and implicit in the written form. Immediately after the last treatment sessions, a speaking post-test was administered. The learners' performance on pre-test and post-test speaking tasks was later scored by two raters to assure inter-rater reliability.

In the qualitative phase of the current study, a semi-structured interview was employed to gain an insight into the attitudes of the MB group toward using the messenger bot as a tool for obtaining corrective feedback. They were asked to answer the pre-determined questions in an individual session with the researchers and to elaborate on their answers when needed to enrich the data collected. Each interview lasted approximately 10 minutes. The transcriptions of the interview were analyzed through the first two phases introduced in the grounded theory, which were called open coding and axial coding.

3.6. Data Collection Analysis

The scoring of the collected data was conducted by two raters. The correlation indicating inter-rater reliability was reported .86 which showed that reliability was confirmed. After obtaining the learners' scores as explained above for the pre-test and post-test, descriptive statistics were first obtained. Then, to measure the effects of the messenger bot, a mixed between-within subjects ANOVA with one within-subject factor (i.e., time) and one between-subjects factor (i.e., feedback modality) was conducted.

After the first phase, the participant's answers to the interview questions were transcribed and analyzed based on the framework of the grounded theory. In the first phase, i.e., the open coding, the interviews were transcribed and the researchers added any relevant points which were considered in the interview sessions. In the axial coding, the data were categorized and the themes of the data were identified. More data were collected and categorized until the researchers made sure that saturation in the data occurred.

4. Results

To answer the first research question, a mixed between-within subjects ANOVA was conducted to find out if there were differences between the performance of each group in the two aforementioned exams. The descriptive statistics help investigate the differences numerically.

Table 4.1

The Descriptive Statistics Regarding the Performance of Three Groups in Pre-test and Post-test

	Group	Mean	Std. Deviation	N
Pre-test	MB	70.0333	5.68412	30
	TF	69.8500	4.98709	20
	NF	69.6000	5.10315	20
	Total	69.8571	5.25656	70
Post-test	MB	84.5333	5.90577	30
	TF	79.0000	4.73509	20
	NF	71.2000	5.22746	20
	Total	80.0000	6.87992	70

As demonstrated in Table 4.1, the mean scores of the three groups in the pre-test were close to each other (MB group, *M=70, SD=5.7;* TF group, *M=69.9, SD=4.9*; NF group, *M=69.6, SD=5.1*). However, in the post-test, the mean scores of the three groups were different and a comparison between pre-test and post-test scores in each group showed an increase in the scores gained in the post-test (MB group, *M=84.5, SD=5.9*; TF group, *M=79, SD=4.7*; NF group, *M=71.2, SD=5.2*). To compare the differences statistically, in the first place, it is needed to ensure that the error variance was equal.

Table 4.2

Leven's Test of Equality of Error Variance

	F	df1	df2	Sig.
Pre-test	.509	2	67	.603
Post-test	.295	2	67	.745

In Table 4.2, the p-values for both pre-test and post-test scores were larger than .05 (pre-test, *p=.603*; post-test, *p=.745*) and it can safely be assumed that there was no violation of the homogeneity assumption. A further assumption that was required for mixed between-within subjects ANOVA was the similarity of the intercorrelation among the levels of the within-subject variables.

Table 4.3

The Equality of Covariate Matrices

Box's M	33.926
F	5.399
df1	6
df2	58331.622
Sig.	.021

Based on Table 4.3, the assumption of the equality of covariate was not violated as the p-value is higher than .01 (*p=.021*). The test of the between-subjects effect needed to be analyzed to compare the progress of the participants in the three groups.

Table 4.4

Multivariate Test of the Three Groups Based on Their Scores in Pre-test and Post-test

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
time	time Time Wilks' Lambda	.273	178.643	1.000	67.000	.000	.727
time *	Wilks' Lambda	.651	17.963	2.000	67.000	.000	.349

As the findings presented in Table 4.4 indicate, there was a substantial main effect of *F(1,67)=178.6, p< 001*. The partial eta squared was *.727*, indicating a large effect. It shows that the participants performed much better in the post-test. There was also a significant interaction between group and time (*Wilk's Lambda=.651, F(2,67)=17.96*).

Table 4.5

The Pairwise Comparisons for the Pre-test and Post-test Scores

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-9.417*	.705	.000	-10.823	-8.010
2	1	9.417*	.705	.000	8.010	10.823

The pairwise comparisons of the pre-test and post-test scores reveal that, in general, the groups performed significantly differently in the two exams. To compare the difference in the performance of the groups, a test of the between-subjects effect needs to be investigated.

Table 4.6

Tests of Between-Subjects Effect for the Three Groups

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	750010.301	1	750010.301	18349.323	.000	.996
Group	709.727	2	354.864	8.682	.000	.206
Error	2738.558	67	40.874			

In Table 4.6, a substantial main effect was observed, showing that the overall mean scores of the groups significantly differed from each other. The mentioned differences were obvious in the means plot as well.

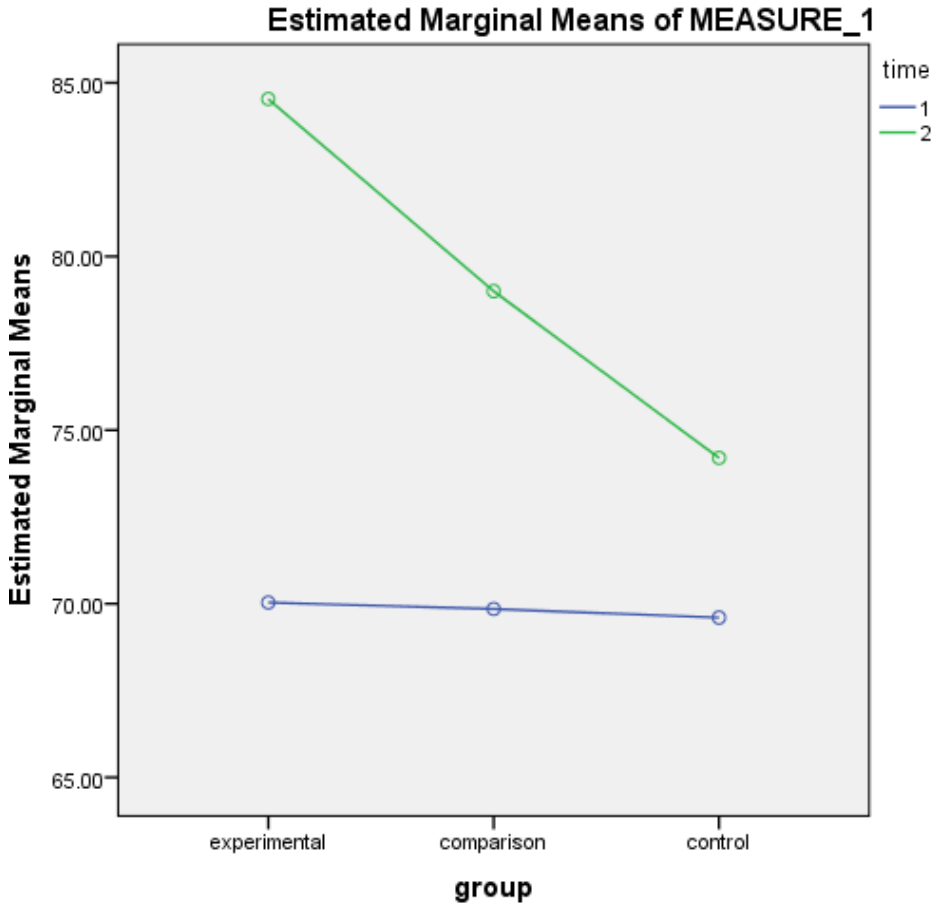

Figure 4.1. *The Estimated Marginal Means of Scores in Pre-test and Post-test for the Three Groups*

As illustrated in Figure 4.1, the groups were the same in the pre-test, which shows the homogeneity of the participants assigned into different groups before the treatment. However, the MB group outperformed TF and NF groups in the post-test. To investigate the progress of each group, a one-way between-groups ANOVA was conducted.

Table 4.7

The Multivariate Tests of Pre-test and Post-test Scores for MB Group

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.824	135.948	1.000	29.000	.000	.824
Wilks' lambda	.176	135.948	1.000	29.000	.000	.824
Hotelling's trace	4.688	135.948	1.000	29.000	.000	.824
Roy's largest root	4.688	135.948	1.000	29.000	.000	.824

As the multivariate test of pre-test and post-test scores for the MB group indicates, the scores in the two tests were significantly different (*Wilk's Lambda=.176, F(1,29)=135.95*). Partial eta squared (=*.824*) shows that there was a large effect size. The results for the TF group are presented in Table 4.8.

Table 4.8

The Multivariate Tests of Pre-test and Post-test Scores for TF Group

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.665	37.760	1.000	19.000	.000	.665
Wilks' lambda	.335	37.760	1.000	19.000	.000	.665
Hotelling's trace	1.987	37.760	1.000	19.000	.000	.665
Roy's largest root	1.987	37.760	1.000	19.000	.000	.665

The results presented in Table 4.8 for the TF group show that the scores in the two tests were significantly different (*Wilk's Lambda=.335, F(1,19)=37.76*). There was a large effect size as the partial eta squared was reported *.665*. The results for the NF group are as follows.

Table 4.9

The Multivariate Tests of Pre-test and Post-test Scores for NF Group

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.282	21.563	1.000	19.000	.060	.382
Wilks' lambda	.818	21.563	1.000	19.000	.060	.382
Hotelling's trace	.451	21.563	1.000	19.000	.060	.382
Roy's largest root	.451	21.563	1.000	19.000	.060	.382

However, as Table 4.9 presents, the NF group did not significantly progress from pre-test to post-test results (*Wilk's Lambda=.818, F(1,19)=21.56, p=.06*). To compare the improvement of groups, a bar graph is presented below.

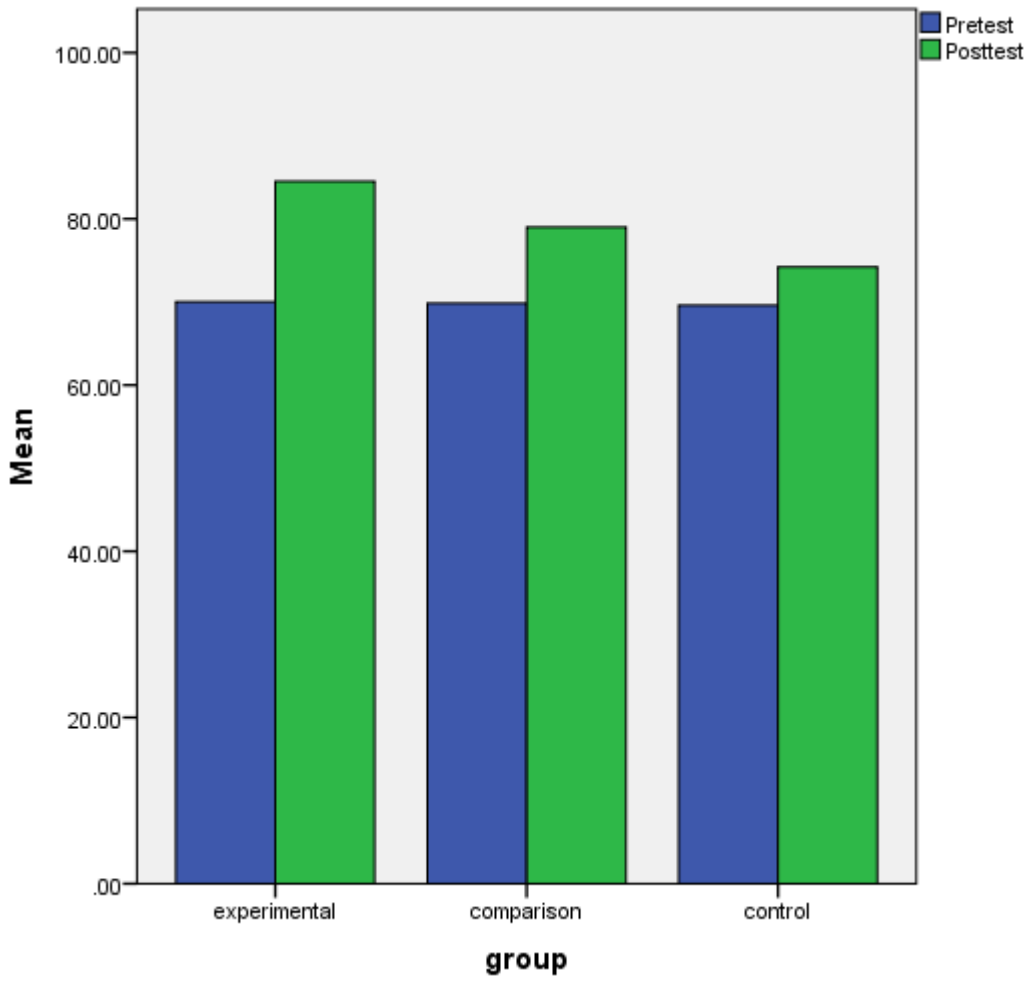

Figure 4.2. *The Bar Graph Representing the Scores on MB, TF, and NF Groups in Pre-test and Post-test*

As the figure depicts, all the three groups improved from pre-test to post-test. However, it seems that the MB group underwent the most changes compared to other groups. TF and NF groups experienced changes as well, although the changes were not statistically significant for the NF group. The graph shows that the MB group outperformed the other two groups.

To answer the second research question, it was required to analyze the interviews the learners participated in. In the first phase, the data collected from the participants through the interview were transcribed and the results of the first phase were launched into the second phase of coding to feed the themes and subtheme introduction. The coding phases of the grounded theory revealed that the participants in this study showed positive attitudes toward the use of the bot, which was directly mentioned in their interview. In addition, they suggested some practical uses for the bot in other skills, which could be considered as

evidence that they found the bot quite useful. Based on the information obtained from the first two phases, the learners' attitudes toward receiving feedback through the messenger bot were discussed below.

5. The Trustworthiness of the Study Findings

To gain a deeper insight into each participant's description of their attitudes toward working with the bot, the researchers initially read transcripts many times to thicken the description. The scoring and coding procedures were done by two raters to ensure the scores were given and the themes were chosen as accurately as possible. The procedure used in coding was in NVivo, in which the researchers would decrease the sentences into comprehensible phrases (Ary, 2012). To validate the themes, all identified themes were submitted to the participants for member checking to ensure that they agreed with the findings of the study. Ultimately, all the participants agreed that the results were compatible with their attitudes.

6. Discussion

Regarding the first research question, the statistical analyses in the previous section indicate that the MB group for whom the feedback was provided through the messenger bot outperformed the TF group, who received the feedback traditionally, and the NF group, who did not receive any feedback. In comparing MB and TF groups, the difference in their performance could be associated with the use of the bot. The findings of the study are in line with the studies which considered the use of applications advantageous in the realm of language learning (Klimova & Maresova, 2017). However, the findings are particularly interesting in that they are a rebuttal to the claim that the use of apps could work against learners' concentration and work like a distracter (Bowman et al., 2010; Junco & Cotten, 2011). In addition, it is claimed that the use of messaging apps as a learning tool can improve the quality of learning (Chuang & Tsao, 2013; Rambe & Bere, 2013; So, 2016; Sun et al., 2018). Some researchers (Bowman et al., 2010; Junco & Cotten, 2011) believe that using such apps may cause distraction due to multitasking and dealing with unrelated messages. Others believe that these apps can be advantageous due to being portable and popular among young people (Klimova & Maresova, 2017).

The outperformance of MB and TF groups, who received feedback, over NF group was evidence for the facilitative effect of feedback on language acquisition (Black & Wiliam, 1998; Shute, 2008), either immediate or delayed (Thornbury, 2006) or either explicit (Carroll & Swain, 1993) or implicit (Sanz & Morgan-Short, 2004).

The second phase of the current study, which was qualitative, revealed the emerging themes that were based on the participants' perceptions of the messenger bot, which were discussed in the second research question. Based on the interviews, the attitudes of the participants toward the use of the bot in fulfilling the defined tasks were elicited. In addition, most of the participants commented on the use of the bot in other fields or for other skills.

Theme One: Pedagogical Application

In the qualitative phase of the study, the first theme revealed speaks to the pedagogical application of the bot. From the participants' responses and the interaction that the researchers had with them, besides the follow-up interviews with the participants, it was concluded that the participants had positive attitudes toward the bot and its use to practice speaking. They commonly mentioned that, as seen in the following reflections provided by Ali and Ayda.

Extract 1. *I used to send my speaking samples to my teacher practicing for the TOEFL exam via Whatsapp and Telegram. My experience was awful because sometimes she was in class, another time I was in a meeting. Synchronizing our schedule for free time slots was challenging, and in many cases, I didn't receive my feedback on time. Since the day she has introduced this bot to me, my life quality improves ten steps ahead. Why? Because I am a postdoctoral researcher working in eight different research centers and I spend most of my working hours in meetings. By using this bot, I can send my answers and receive feedbacks in every time slot of the day or night. I haven't seen any other bot which is similar to this one. It is well-customized for its application*. (Ali)

Extract 2. *Well, I have never seen such an application that can help me with grammar in speaking. Concerning the purpose of bot maker, all students who don't have access to TOEFL speaking classes can take advantage of this bot. They can save time. Additionally, the students who participate in classes can answer their exercises in this bot and get their teacher's feedback.* (Ayda)

In Extracts 1 and 2, Ali and Ayda described how, although they had had previous experiences related to the use of other applications, those experiences did not help them correct their mistakes in speaking. Their accounts allude to how the messenger bot can be useful to them and how it would be different from similar technological tools.

The unique features of the bot were mentioned to be appealing to the learners. They maintained that the traditional form of receiving feedback had many disadvantages, such as being mocked by a classmate. Working with the messenger bot gives them more confidence to perform the task and reflect on the feedback provided, which is in line with the findings of Jia and Ruan (2008), who state that learners feel more comfortable while using the bot, which can promote more effective learning.

The findings of the study are in contrast to those found by Bowman et al. (2010) who believe that the use of the messenger bot works like a distracter for the learners and reduces the profitability of the mediating tool. In addition, Junco and Cotten (2011) believe that unrelated messages can work as an obstacle to the performance of the learners, reduce their speed, and decrease the quality of their work.

Another issue that raised the participants' interest was the user-friendly nature of the bot. The participants appreciated the fact that they did not need to install any other software or application to use the bot and made it easy to work with. The reflection provided by Saeed and Tara is an example which can be considered in this regard.

Extract 3. *I haven't seen any other bot which is similar to this one. It is well-customized for its application. As a computer science professional, I certify that this design decreases the user's cognitive load and can be used by all-level learners who know how to work with messenger bots. This UI/UX design surprised me just as I finished my first round of practice. If I were in your shoes, I'd add a progressing plot that shows some positive and negative feedback weekly*. (Saeed)

Extract 4. *I've always been afraid of using technology in language learning. You know, you try to install something and then you click something and you feel that you are lost. Each time I wanted to try an application, I would remember this and decided to continue reading books. But using this bot is so simple. I liked it.* (Tara)

Extracts 3 and 4 reflect the important issue of the hindrance learners face while trying to combine technology into language learning. For most of the learners, the experience of working with a newly designed application or software has been reported to be satisfactory.

As Kim et al. (2019) mention in their study on Alexa and Google assistant chatbots, both of them were satisfactory to the users; however, a factor which plays a role here is the better comprehensibility and hence user-friendliness of the latter, which made it more popular among users.

In Extract 4, Tara mentioned the fact that her lack of skills in working with applications and software or perhaps generally with computers made it difficult for her to integrate technology with language learning. Hence, it is facilitated by the user-friendly nature of the bot. This finding is in line with the second subtheme of the study and highlights the importance of the first phase of working with any software, which is the preparation phase. The fewer the steps in preparing the bot, the more popular it would be among users.

The comfort that the use of bots brings to learners is one of the main reasons that they are among popular gadgets nowadays. Time limitation is one of the annoying factors with the homework in traditional classes. The accessibility of bots is an astounding feature that enables the users to have their learning plan any time of the day and night and reflect on the feedback the teacher provides them with through the bot with no time restriction. This feature is perfect for occupied, busy people who cannot attend classes like others.

Extract 5. *Previously I used to use some applications to learn English and improve my speaking proficiency. They helped me but not as much as this bot. This one helped me more because there was a real teacher who could… help me with my problems and whenever I send her my answer, she corrected me. Sometimes she corrected me directly and sometimes she gave … uh, some hints to me and let me correct myself.* (Ahmad)

Extract 6. *The idea of receiving feedback whenever you send the audio to the teachers is something interesting. I still can remember why I said the sentence like that and I can understand my problem. In addition, the teacher doesn't always give feedback the same way and sometimes you have to think about what she meant*. (Parisa)

In the aforementioned extracts, the participants mentioned the advantage of the bot regarding the immediate feedback it provides, compared to the traditional methods in which learners have to wait until the next session to be provided with the feedback. According to Heller (2017), users can make the most out of bots because they do not need to spend a lot of time going to classes and sifting through the feedback provided by the teacher to find the one which suits their own mistake.

The participants mentioned that the use of the bot helped them improve their accuracy. The reason, they stated, was the feedback provided for each learner which s/he could reflect on to avoid repeating that mistake. The participants would receive the feedback on their speaking report right after sending it through the bot when the reasoning based on which they spoke was fresh in their mind. This way they could make a clear image of the mistakes and the possible reasons behind them; gradually this procedure would help omit the mistakes from their interlanguage. According to Ellis and Barkhuzien (2005), appropriate language as a mediating tool is one of the main factors affecting language development. In addition, bots are considered as tools that increase the quality of learning (Chuang & Tsao, 2013; Rambe & Bere, 2013; So, 2016; Sun et al., 2018).

Theme Two: Technological Issues

The second theme was regarding technological issues. Participants mentioned different issues which were raised in the previous section. The suggestions concerned the use of a bot to show the learners' progress.

Extract 7. *Personally speaking, if you could add a high score example after evaluating each question, it would help students to try to improve their speaking and in addition, I think it would be beneficial if you could add a score for each speaking and by this means students can be familiar with their progress over time by Ahhh… analyzing their scores over time.* (Arezoo)

Extract 8. *I would like the bot more if I could know about my progress. Like it could show me how I improved by comparing my scores in the first and last speaking tasks, for example.* (Hossein) It is worth mentioning that this option is active for the researchers in the bot, but not for the common users. The solution to this problem is not hard and can be regarded as further research into the realm.

As mentioned before, the learners would find the bot more advantageous if it were not for the time interval between each report they sent for every question. The reason behind this possibility is educational due to the reason that the learners need to have some cognitive load of information regarding world knowledge, vocabulary, and grammar when they want to speak or write about a topic. Therefore, having the topics successively without any pauses causes them to forget the reason they made a mistake in the previous file and minimizes the chance of internalizing the grammatical rule they missed to apply. These findings are in line

with Karaagac and Threlfall (2004), who claim that if the meditational tools are used purposefully and flexibly, they can cause dramatic changes to the learner's interlanguage.

The satisfaction the participants experienced with using the bot rose the issue of manipulating the bot for other skills in addition to speaking.

Extract 9. *This bot was something interesting and really easy to use. I like it and I think every student from any part of the world can use it. But it can be also used for writing. I think we can make improvements in our skills.* (Shahin)

Extract 10. *I need such an app for other skills and they can improve it. If we can send our correct answers after the teacher corrects us, it will be good. I want to have this bot on my computer, but there are no teachers to correct me.* (Sahar)

The participants believe that the bot is perfect for practicing writing, as they will be able to receive feedback individually and they do not have to wait for the next session of the classes when they cannot remember what their logic behind making the mistakes had been. On the other hand, practicing listening and reading skills does not need the online presence of the teacher in most cases and it can reduce learners' expenses on private tutors.

Based on the above-discussed findings, two main themes emerged, which were pedagogical applications and technological issues. These themes included various aspects, which were mainly positive in nature, ranging from enhanced accuracy to better accessibility and comprehensibility. Furthermore, the use of the bot boosted the students' confidence and motivation. As a suggestion, the students proposed that the bot could be improved to be used for other skills and other kinds of feedback, such as oral feedback.

7. Conclusion

This study intended to provide language learners with WCF on their responses to the TOEFL speaking task II via a messenger bot. It studied the effect of using the messenger bot on students' improvement in using definite and indefinite articles more accurately. Moreover, it investigated the users' attitudes toward making use of this bot and developed a grounded theory based on an in-depth semi-structured interview.

The numerical analysis conducted to investigate the impact of the bot on the use of articles by the learners revealed that the MB group outperformed the other two groups. The learners in the MB group not only improved significantly from the pre-test to the post-test, but also had the highest average of the post-test. The trend can be explained by referring to

the impact of technology use and its influence on learning. Despite the distracting aspect of technology use reported before, it transpired to be helpful in most of the reported cases.

The learners' perspectives presented in two main themes in the article show that all the participants had positive attitudes toward using the bot. The reason behind this positive attitude can be the user-friendliness of the bot for the participants and some other issues which were raised in the previous section.

The second theme, moreover, presented the practical use of the bot the learners proposed. Looking through the practical use of the bot they had in mind gives new perspectives on the use of the bot in other fields or for other skills. Seeking the reason which made the participants think of new ways of using the bot reveals that their pleasant experience with the bot for the speaking skill pushed them to expand their imagination to have it for other skills or even in other realms of practicing languages. In other words, the positive attitude toward using the bot was a motivating factor for opening new windows to the researchers to contrive/devise ways of improving the bot. The improvement the learners who worked with the messenger bot experienced and the positive attitude they expressed toward the bot demonstrates the fact that the numeric and qualitative phases of the study coordinate with each other.

The limitations of this study lie in some parts. The first limitation goes back to the bot architecture. To conduct the present study, the bot was designed to merely provide WCF and focus on only one task and one skill. The limitation could be overcome if we design a bot that includes different skills and different tasks, so that it will be more convenient for students to use it. Furthermore, this bot could be improved to be intelligent and attached to a corpus glossary to be able to correct some parts of the errors. It is costly and arduous to design such a bot. It is worth mentioning that the lack of control over the gender variable is another limitation related to the design of the study. There are a few suggestions for further research. Since the post-test was given as soon as the treatment sessions had finished, it can be suggested that a delayed post-test could be given to investigating the long-term effect of the treatment. Furthermore, the explicit and implicit knowledge of using definite/indefinite articles can be investigated in future studies. Besides, this bot can be improved to be implicated for different skills and different kinds of feedback.

There is no doubt that in this era, we have to utilize technology to make the process of language learning and teaching more effective and motivating. Therefore, messenger bots,

which are considered innovative in this realm, could be used for instructional purposes and have the potential to be a rewarding area for further research. Teachers should be helped to gain insights into the most appropriate and efficient bots for their students' language development based on the context and environment.

References

Abushawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *Ldv Forum*, *22*(1), 29-49.

Akcora, D. E., Belli, A., Berardi, M., Casola, S., Di Blas, N., Falletta, S., Faraotti, A., Lodi, L., Diaz, D. N., Paolini, P., & Renzi, F. (2018). *Conversational support for education*. International Conference on Artificial Intelligence in Educatio*n*.

Alavi, M., & Leidner, D.E. (2001) Knowledge management and knowledge management systems: Conceptual foundations and research issues. *3 Quarterly, 25*, 107-136. http://dx.doi.org/10.2307/3250961

Aljaafreh, A. L. I., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal*, *75*(4), 465-483.

Andreev, R., Terzieva, V., & Kademova-Katzarova, P. (2009). An approach to the development of personalized e-learning environment for dyslexic pupils' acquisition of reading competence. *Proceedings of the International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing - CompSysTech '09*, 1. https://doi.org/10.1145/1731740.1731829

Athanaselis, T., Bakamidis, S., Dologlou, I., Argyriou, E. N., & Symvonis, A. (2014). Making assistive reading tools user friendly: A new platform for Greek dyslexic students empowered by automatic speech recognition. *Multimedia Tools and Applications*, *68*(3), 681-699. https://doi.org/10.1007/s11042-012-1073-5

Beatty, K. (2003). *Teaching and researching computer-assisted language learning*. Longman.

Bii, P. (2013). Chatbot technology: A possible means of unlocking student potential to learn how to learn. *Educational Research*, *4*(2), 218-221.

Bii, P., Too, J., & Langat, R. (2013). An investigation of student's attitude towards use of chatbot technology in instruction: The case of Knowie in a selected high school. *Education Research*, *4*(10), 710-716.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Bonk, C. J., & Cunningham, D. J. (1998). Searching for learner-centered, constructivist, & sociocultural components of collab educ lrng tools. In C. J. Bonk, & K. S. King (Eds.), *Elec collaborators: Learner-centered tech's for literacy, apprenticeship, & discourse* (pp. 25-50). Erlbaum.

Bowman, L. L., Levine, L. E., Waite, B. M., & Gendron, M. (2010). Can students multitask? An experimental study of instant messaging while reading. *Computers & Education*, *54*(4), 927-931.

Brinton, C. G., Rill, R., Ha, S., Chiang, M., Smith, R., & Ju, W. (2015). Individualization for education at scale: MIIC design and preliminary evaluation. IEEE Transactions on Learning Technologies, *8*(1), 136-148.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Study in Second Language Acquisition, 15*, 357-386.

Chuang, Y. H., & Tsao, C. W. (2013). Enhancing nursing students' medication knowledge: the effect of learning materials delivered by short message service. *Computers & Education, 61*, 168-175.

Doff, A. (1998). *Teach English: A training course for the teacher*. Cambridge University Press.

Egbert, J. L. (2005). Conducting CALL research. In J. L. Egbert & G. M. Petrie (Eds.), *CALL research perspectives* (pp. 3-8). Lawrence Erlbaum.

Ellis, R., & Barkhuzien, G. (2005). *Analyzing learner language*. Oxford University Press.

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research.* Routledge.

Gehl, R. W. (2014). Teaching to the Turing Test with Cleverbot. *Transformations: The Journal of Inclusive Scholarship and Pedagogy, 24*(1-2), 56-66.

Guanuche, A., Eiriz, O., & Espi, R. (2020). Corrective feedback through mobile apps for English learning: A review. In Narváez, F. R., Vallejo, D. F., Morillo, P. A., & Proaño, J. R. (Eds.), *Smart technologies, systems and applications: First international conference, SmartTech-IC* (pp. 229–242). Springer.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. Computers & Education, *98*, 157-168.

Jia, J., & Ruan, M. (2008). Use chatbot CSIEC to facilitate the individual learning in English instruction: A case study. *Intelligent Tutoring Systems*, 70a6-708.

Junco, R., & Cotten, S. R. (2011). Perceived academic effects of instant messaging use. *Computers & Education*, *56*(2), 370-378.

Karaagac, K., &Threlfall, J. (2004). The tension between teacher beliefs and teacher practice: The impact of work setting. *Proceedings of 28th annual conference for the Psychology of Mathematics Education (PME), 3*, 137-144.

Kern, R. (2006). Perspectives on technology in learning and teaching languages. *TESOL Quarterly, 40*(1), 183-210. https://doi.org/10.2307/40264516

Kim, N. Y. (2017). Effects of different types of text-based chat on Korean EFL students' writing performance. *Korean Journal of Linguistics, 42*(3), 277-301.

Kim, H., Shin, D. K., Yang, H., & Lee, J. H. (2019). A study of AI chatbot as an assistant tool for school English curriculum. *Journal of Learner-Centered Curriculum and Instruction*, *19*(1), 89-110.

Kim, N. Y., Cha, Y., & Kim, H. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning, 22*(3), 32-53.

Klimova, B., & Maresova, P. (2017). Computer-based training programs for older people with mild cognitive impairment and/or dementia. *Frontiers in Human Neuroscience, 11,* https://doi.org/10.3389/fnhum.2017.00262

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Lantolf, J. P. (2000). Second language learning as a mediated process. *Language Teaching, 33*(2), 79-96.

Malouf, D. B. (1988). The effect of instructional computer games on continuing student motivation. *The Journal of Special Education*, *21*(4), 27-38. https://doi.org/10.1177/002246698802100406

Mayer, R. E., & Moreno, R. (2010). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52. https://doi.org/10.1207/S15326985EP3801

Nassaji, H. (2017). The effectiveness of extensive versus intensive recasts for learning L2 grammar. *The Modern Language Journal, 101*(2), 353-368.

Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking. ESL & Applied Linguistics Professional Series*. Routledge Taylor & Francis Group.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in higher education, *31*(2), 199-218.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Newbury House Publishers.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, *52*(1), 1-12. https://doi.org/10.1016/j.compedu.2008.06.004

Pokatilo, A. (2016). Chatbots take education to the next level. *Chatbot News Daily*. Retrieved from https://chatbotnewsdaily.com/chatbots-take-education-to-the-next-level-23bc02cdbccf

Rambe, P., & Bere, A. (2013). Using mobile instant messaging to leverage learner participation and transform pedagogy at a South African University of Technology. *British Journal of Educational Technology*, *44*(4), 544-561.

Ricci, K. E., Salas, E., & Cannon-bowers, J. A. (1996). Do computer-based games facilitate knowledge acquisition and retention? *Military Psychology*, *8*(4), 295-307. https://doi.org/10.1207/s15327876mp0804_3

Riel, J. (2016). How chatbots can help with learning. University of Illinois Chicago (UIC). *College of Education*. https://education.uic.edu/academics-admissions/student-life/how-chatbotscan-help-learning

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning*, *54*, 35-p78.

Schmulian, A., & Coetzee, S. A. (2018). The development of Messenger bots for teaching and learning and accounting students' experience of the use thereof. *British Journal of Educational Technology*, 1-27. doi:10.1111/bjet.12723

Shumin, K. (2002). Factors to consider: Developing adult EFL students speaking abilities. In J. C. Richards, & W. A. Renandya (Eds.), *Methodology in Language Teaching* (pp. 204-211). Cambridge University Press.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. https://doi.org/10.3102/0034654307313795

So, S. (2016). Mobile instant messaging support for teaching and learning in higher education. *The Internet and Higher Education*, *31*, 32-42.

Sun, Z., Lin, C. H., Wu, M., Zhou, J., & Luo, L. (2018). A tale of two communication tools: discussion-forum and mobile instant-messaging apps in collaborative learning. *British Journal of Educational Technology*, *49*(2), 248–261.

Thornbury, S. (2006). *An A-Z of ELT: A dictionary of terms and concepts used in English language teaching*. Macmillan Education.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

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

Williams, M., & Burden, R. L. (1997). *Psychology for language teachers*. Cambridge University Press.

Williams, T. C., & Zahed, H. (1996). Computer-based training versus traditional lecture: Effect on learning and retention. *Journal of Business and Psychology*, *11*(2), 297-310. https://doi.org/10.1007/BF02193865

Winkler, R., & Sollner, M. (2018). *Unleashing the potential of chatbots in education: A state-of-the-art analysis*. Academy of Management Annual Meeting (AOM), Chicago, USA. https://doi.org/10.5465/AMBPP.2018.15903

Winne, P. H., & Butler, D. L. (1994). Student cognition in learning from teaching. In T. Husen & T. Postlewaite (Eds.), *International encyclopedia of education* (2nd ed., pp.5738-5745). Pergamon.

Yoo, Y. (2010). Computing in everyday life: A call for research on experiential computing. *MIS Quarterly*, *34*(2), 213-231. https://doi.org/10.2307/20721425

Appendix

The interview questions were as follows:

1. What is your main goal when using this bot?
2. What improvements could be made to make this bot better?
3. What is your impression of this bot?

4. When do you think someone can use this bot?
5. Do you think this bot is similar to another one?
6. Who do you think can take advantage of this bot?
7. Was there anything surprising to you while using this bot?