

Chatting with AI Bot: Vocabulary Learning Assistant for Saudi EFL Learners

Reem Alsadoon¹

¹ College of Languages and Translation, Imam Mohammed Ibn Saud Islamic University, Riyadh, Saudi Arabia

Correspondence: Reem Alsadoon, Imam Mohammed Ibn Saud Islamic University, Unnamed Road, Riyadh 13318, Saudi Arabia.

Received: March 18, 2021

Accepted: May 27, 2021

Online Published: May 31, 2021

doi: 10.5539/elt.v14n6p135

URL: <https://doi.org/10.5539/elt.v14n6p135>

Abstract

In the AI field of language learning, chatterbots are an interesting area for language learning and practice. This research investigates Arabic EFL vocabulary learning using an interactive storytelling chatterbot. A chatterbot was created and equipped with four vocabulary tools: a dictionary, images, an L1 translation tool, and a concordancer. The target words were enhanced by these tools to provide the learners with interactive comprehensible input. This project seeks to identify which tools are mostly used when EFL learners are practicing English with a chatterbot. It also seeks to determine which tool could help most in vocabulary learning as well as retention. The results of the study indicate that the dictionary is the most favoured and effective tool for vocabulary learning. For retention, the findings uncover that L1 translation is slightly (but insignificantly) higher than the dictionary.

Keywords: AI bots, Arabic EFL learners, vocabulary learning, interactive help tools

1. Introduction

1.1 Research Problem

A chatterbot is one of the potential resources for language learners to have authentic and human-like practice. Brennan (2006) defines a chatterbot as “an artificial construct that is designed to converse with human beings using natural language as input and output.” In the early 1960s, the first chatterbot, “Eliza,” was programmed by Joseph Weizenbaum. Eliza played the role of a psychotherapist and would simply rearrange a submitted statement or question into a new question. The software was designed to be very human-like by operating through a pattern-matching technique. This technique works by parsing “input looking for keywords (‘family’, ‘mother’, ‘job,’ etc.) with the output consisting of rephrasing elements of the input sentence into the output or pre-set automatic responses” (Coniam, 2008). Over the past few decades, a range of chatterbots has grown in quality and quantity.

Hubbard (2009) emphasizes that chatterbots are a promising area for future research on second language acquisition (SLA) and computer-assisted language learning (CALL). Furthermore, Coniam (2008) reports that “it’s apparent that a chatbot’s ability to respond in English has interesting language potentials issues”. However, this area has been neglected by CALL researchers due to the technological limitations of these conversation practice machines, which make the goal of using it for language learning almost too far (Atwell, 1999). However, with the advances in speech understanding technologies, Fryer, Nakao, and Thompson (2019) emphasize that language learning and practice has improved for second language learners especially for those at beginning and intermediate levels. They note that chatterbots usually have large lexicons which can make them excellent tools for conversation practice and vocabulary acquisition.

The current research aims to improve EFL learners’ vocabulary learning through a chatterbot enhanced with vocabulary learning tools, i.e. a dictionary, images, L1 translation and a concordance, to facilitate vocabulary understanding in conversations.

1.2 Objectives

The study has several objectives regarding the field of ICALL and ESL/EFL learning. First, it examines how an online conversational model enhanced with vocabulary learning tools can help EFL beginners to learn vocabulary. Second, it analyses EFL learners’ preferences for the vocabulary learning tools while conversing

with the chatterbot. Third, the study draws the attention of ESL/EFL teachers to the potentials of chatterbots in improving conversation practice and vocabulary learning, especially in the early language learning stages. Finally, it draws the attention of CALL researchers and designers to the tools that could enhance the learning opportunities provided by such artificially intelligent software. To fulfil these objectives, the study first discusses the relevant previous research and theories underpinning this research. Then, it asks the research questions and describes the research methodology and analysis. Finally, it presents the study's findings and their interpretations based on previous research.

2. Literature Review

The literature review discusses the theoretical framework underpinning this research as well as the previous literature on vocabulary learning and online vocabulary learning tools, and chatterbots and language learning.

2.1 Theoretical Framework

The study will be mainly based on SLA interactionist theory and input enhancement hypothesis. Interactionist theory focuses on conversational interaction. Long (1996) argues for the importance of comprehensible input as a major factor in second language acquisition; however, he also believes that interactive input is more important than non-interactive input. The present study uses a chatterbot to provide EFL learners with interactive comprehensible input. Chapelle (1998) proposes input-enhancement hypothesis as one of the principles relevant to developing multimedia CALL. This hypothesis states that input is the "potentially processible language data which are made available, by chance or by design, to the language learner" (Sharwood Smith, 1993). In order for the input to be learnt, especially in a CALL environment, it should be enhanced or modified "through simplification, elaboration, added redundancy, or sequencing to make it 'comprehensible'" (Chapelle, 1997). In this study, the chatterbot was enhanced with a dictionary, a concordancer, images and L1 translation.

2.2 Vocabulary and Online Learning Tools

The Internet has made it easy for L2 learners to learn vocabulary with a variety of learning aids. They can look up the meaning of unknown words through online dictionaries, L1 translation tools, image search engines or concordancers. Also, they can practice the new words using online materials, interactive exercises or computer-mediated communication (CMC). However, it is challenging for developers and researchers to evaluate the effectiveness of these materials and tools (Horst, Cobb, & Nicolae, 2005). Sökmen (1997) argues that:

There is a need for programs which specialize on a useful corpus, provide expanded rehearsal, and engage the learner on deeper levels and in a variety of ways as they practice vocabulary. There is also the fairly uncharted world of the Internet as a source for meaningful vocabulary activities for the classroom and for the independent learner. (p. 257)

Horst, Cobb, and Nicolae (2005) found this challenge so interesting that they designed an experimental ESL course with a set of existing and purpose-built on-line tools for vocabulary learning. These free online resources, available at www.lex tutor.ca, include a concordancer, a dictionary, a cloze-builder and hypertext, as well as a database with an interactive self-quizzing feature. Their course was based on Sökmen's (1997) criteria for designing computerized vocabulary activities. The study aimed to specialize in a 'useful corpus' to engage the learners at a deeper level of cognitive processes; hence, fostering retention of the vocabulary. The participants' entries into the database were analysed to determine whether their example sentences supported the word meanings adequately. Also, their responses in a pre-test and post-test were compared to establish learning gain and retention. The study found that the quality of their example sentences was high with some cues of the meaning of the word. From all the tools, the online dictionary, as well as the word bank, were found to be used more than the other tools. In terms of gain, however, no significant relationship was found between the tools and the gain.

Laufer and Hill (2000) have discovered that the use of an electronic dictionary has a positive effect on incidental vocabulary learning. The reason is that students may not use a physical dictionary because of the time involved or the disruption of the flow of reading. Laufer and Hill argue that if a pedagogical tool is popular with students, the chances are it will also be beneficial for learning. Furthermore, they reveal that the use of electronic dictionaries during reading, especially L1 translation dictionaries, helps in retention. On the other hand, Chen (2017) found that collocations looked up in the dictionary are not remembered after one week. Furthermore, he emphasized that the number of lookups does not have an impact on retention.

On a separate but related research area, Chapelle (1998) argues for the input-enhancement hypothesis in a multimedia CALL environment. Vocabulary input can be enhanced by the various online tools such as dictionary definitions, concordancer examples, images, videos, L1 translation, and a variety of other useful tools.

In this context, Gürkan (2019) explored the effect of vocabulary enhancement through multimedia annotations as compared to paper-based annotations. The study was conducted on 122 elementary students in Turkey learning English as a foreign language. They study employed a pre, post and a delayed-test model to measure the students' vocabulary recall and retention. The findings show significant impact for the online enhanced vocabulary learning through multimedia annotation when compared to paper-based annotations and no treatment group.

In another context, Yoshii (2006) examined the effect of L1 and L2 glosses on incidental vocabulary learning in a multimedia environment. However, this study examined the effect of additional pictorial cues to the L1 and L2 glosses, as compared to text only, in a computerized reading context. The results showed no significant differences between L1 and L2 glosses. However, the study did find a significant difference between picture (text-plus-picture) and no-picture (text only) glosses. Consequently, the researcher suggested that long-term retention may differ between the two types of glosses.

Yun (2011) attempted to synthesize the literature on this subject. He conducted a systematic analysis of the studies focusing on the effect of hypertext glosses of either text only, visual or a combination of both modes on L2 vocabulary learning in an online reading context. The meta-analysis revealed that a combination of more than one type of hypertext gloss is more beneficial than a single mode representation.

A concordancer is another tool with a powerful impact on L2 vocabulary acquisition (Horst, Cobb, & Nicolae, 2005). In 2001, John examined whether a translation corpus and a concordancer would be beneficial tools to supplement language learning. The main goal of the study was a trial use of the translation corpus and the concordancer to decide if they were useful for a teaching program of German at the beginners' level in an unsupervised environment. Only one beginner student of German was recruited for this study. The participant was asked to find satisfactory answers to unknown vocabulary and formulate the appropriate grammar rules for himself using only the translation corpus and the concordance. The results show that these tools can be greatly beneficial for beginners learning vocabulary and grammar.

Another concordancer study was conducted by Lee and Liou (2003), who investigated the feasibility of incorporating a web-based monolingual English concordancer as an electronic referencing tool into traditional senior high school English classes for vocabulary learning. The findings indicated that students with lower vocabulary proficiency seemed to catch up with students at the high vocabulary level after concordance learning. In other words, concordancing has the potential for scaffolding weak learners to accelerate their vocabulary acquisition. Moreover, students with inductive learning styles benefited most from the concordancing learning experiences.

Lee, Warschauer, and Lee (2019) conducted a meta-analysis study on corpus and concordancer's use for vocabulary learning of a second language. Based on 29 studies, they indicated that the effectiveness of the corpus and concordancer's use on vocabulary learning of L2 depends on several factors such as proficiency level, interaction types, corpus types, training and duration.

2.3 Chatterbots and Language Learning

Nowadays, chatterbot technology opens up new opportunities for language learners to practice anytime and as much as they want. This area is relatively new in the field of ICALL. Harless et al. (1999) created a speech-activated multimedia system (Conversim) for learners to have lengthy conversations with virtual native speakers of Arabic. Four conversation programs were designed for the learners to move from one level to another. The series was rigorously tested by the Defense Advanced Research Projects Agency and in association with the Defense Language Institute and the U.S. Army Research Institute. The results revealed that the participants were motivated to study and learn through the opportunity to speak directly with a native-like speaker, control the conversation and compare their pronunciations of Arabic phrases to those of a native speaker.

In 2004, Jia created the Computer Simulator in Educational Communication (CSIEC) system. The goal of the system is to function as a chatting partner for foreign language learners. In his paper, Jia discusses the computational linguistic side in detail. The system is intelligent in terms of grammar and syntax. It deals with all kinds of sentences and phrases. The sentence types are classified according to the complexity and their representing objects in Natural Language Object Modal in Java (NLOMJ). NLOMJ is a technique that transfers the parsing result to the objects representing the grammar elements in the rules using Java. In 2008, Jia

conducted a study on the CSIEC system after developing additional features such as assessment and fill in the blank grammar exercises. The system was integrated into an English course in a middle school in China. A comparison of the participants' scores from the pre-test to the post-test shows great improvement in their performance. Moreover, the questionnaire of attitudes indicated that the students found the system beneficial.

Fryer and Carpenter (2006) drew attention to how chatterbot programs can give the appearance of an interactive listener by responding to users with follow-up comments or questions. However, this interactive listener is, in fact, a learner who continues to benefit from the huge pool of conversation data received from users. Over time, the chatterbot becomes more intelligent in its outputs. Fryer and Carpenter (2006) surveyed 211 students who were asked to use well-known chatterbots in class such as Alice and Jabberwacky. The results of the questionnaire indicated that the students felt more comfortable chatting with bots than with their teacher or partners. It also showed that the technology is more beneficial for advanced learners than lower levels learners.

Coniam (2008) studied the effectiveness of a chatterbot for ESL learners from another side. He focused on examining the linguistic potential of the current chatterbot programs. The goal of his research was to examine the potential of chatterbots as an ESL learning resource. Five popular chatterbots were evaluated in terms of accuracy, word level, range of vocabulary, sentence level, spelling part of speech and question forms. The results showed there were many limitations with chatterbots. However, their ability to converse in native-like English has interesting potential for language learning.

From the technical side, C-H. Lu et al. (2006) reported on the Tutorbot, which was designed to operate by instant messaging software. The system was directed at ESL learners to provide them with an online environment that was enhanced by reference materials, dictionaries, authorized conversation material and a question-answering function. Unfortunately, they did not study the system on ESL learners. Kim (2017) studied the effect of voice-based chatterbot on EFLs' negotiation of meaning according to their proficiency levels. The participants were 123 Korean students of English. Evidence of negotiation of meaning were coded for confirmation check, comprehension check, clarification requests, repetition and reformulation, and was measured by counting the number of meaning negotiation moves. The results indicated a significant difference between the first and the last chat, indicating that the learners benefited from the chatterbot over the time of the study, which was 16 weeks.

Recently, Kim et al. (2020) explored the impact of using a mobile-based AI agent on EFL learners' writing performance and attitude. They reported significant improvement on the writing performance, especially on grammar and vocabulary. Also, they found that learners' anxiety declined over the course of the study and they became more confident, with positive attitudes towards their learning experience through the AI agent. Furthermore, Alm and Nkomo (2020) analysed learners' experiences in an informal setting through various chatterbots online (e.g., Duolingo forum, Memrise community, Reddit). The study revealed that learners are interested and willing to carry out an informal conversation with an AI agent to practice language learning outside the classroom. However, they feel frustrated if the dialogues do not meet their language learning needs and goals.

2.4 Literature Gaps and the Current Study

The current study is based on interactionist theory and the input-enhancement hypothesis. Therefore, a chatterbot system was created to give EFL learners interactive input that was enhanced by online vocabulary learning tools, namely, a dictionary, images, a concordancer and L1 translation. C-H. Lu et al.'s Tutorbot (2006) included learning tools such as a dictionary to enhance the process of learning English in general. However, the system was not tested on ESL/EFL learners. The current study's chatterbot looked at the effectiveness of these tools for vocabulary learning when they are built into such interactive system. Another attempt to enhance the inputs for ESL/EFL in an AI chatterbot was made by Jia (2008). She created a chatterbot system that gives feedback on grammar. In a web-based environment, Horst, Cobb, and Nicolae (2005) designed an experimental ESL course with a set of existing and purpose-built on-line tools for vocabulary learning, including a dictionary, concordance, L1 translation, and interactive exercises and database. The current study implements the same idea but in an AI chatterbot environment, with more interactivity through conversational stories, and more tools such as images. Peters (2007) reported that enhancement techniques such as definition glosses and pictures help in retention. However, it is more helpful if the word is relevant to a comprehension task. The current study combines four different enhancement tools for learning vocabulary through a chatterbot to examine the effectiveness of these tools in an interactive AI environment.

3. Method

3.1 Research Questions

The present study addresses the following questions:

- 1) What are the participants' preferences regarding the available vocabulary learning tools?
- 2) What is the effect of the tools on vocabulary learning and retention?
- 3) What are the participants' attitudes towards the vocabulary learning chatterbot before and after the experiment?

3.2 Participants

The participants were all EFL Saudi students at a high-beginner to low-intermediate level at the British Council in Riyadh, Saudi Arabia. Their level was determined by their IELTS scores, which were from 4 to 4.5. According to the official IELTS website, these scores indicate that the learners were still at a very basic level. The total number of participants was 20: 16 females and 4 males. Their ages ranged from 18 to 33.

3.3 Data Collection

The data was collected through the following instruments:

Screen recording: the students' look up behaviour and use of the provided vocabulary tools was recorded for analysis.

Pre- and post-questionnaire: a pre- and post-questionnaire was administered before and after the study to have some background information and find out their attitudes towards the chatterbot.

Pre-test and post-test: the participants took a pre-test on the vocabulary that they might encounter in their conversations with the chatterbot. After the experiment, they took a post-test on the words they looked up using the vocabulary tools to check the learning outcomes from using the chatterbot. The tests included distractors to avoid the priming effect.

Delayed post-test: the participants were given another post-test after one week to check their retention.

3.4 Procedures

The study was divided into two phases over a period of seven days. In the first phase, the design of the study was a 4x4 randomized block, such that subjects were randomly assigned to groups of four, with each group reading all four stories and receiving all four treatments, but in different orders. The chatterbot was designed to tell four stories. However, each story was connected to only one vocabulary learning tool. In this way, the participants only had one option to find out the meaning of the new words.

At the beginning of the experiment, the participants were given an attitude questionnaire about the chatterbot, conversational stories, background information and the vocabulary learning tools. The questionnaire began with the consent form. By clicking on the word "next" and filling out the questionnaire, they agreed to take part in the study. After that, the participants were given a pre-test of 132 target words from the four stories of the first phase to rate their knowledge. This test was adopted from Horst, Cobb, and Nicolae's Rating Measure Test, in which the participants are asked to rate their knowledge of a list of words that may occur in their readings. In the test, they have a list of words they may encounter in the four stories and they have to choose from three options: YES (sure I know it), NS (not sure), and NO (I do not know it), as shown in Appendix 1.

By the fifth day, phase 2 began with the post-test. The participants first took a post-test of all the words they looked up to test their vocabulary learning. The test was adapted from Horst, Cobb, and Nicolae's Demonstration Test (2005). The test involved identifying the words they looked up using one of the tools during reading. In the test, they were required to choose "No" if they did not know the word yet or "Yes," with a definition in Arabic or English, and, if possible, to incorporate it in a meaningful sentence.

After the post-test, the participants chatted with the chatterbot about another two stories in two sessions. In these sessions, they were given the option to choose from all four tools whenever they wanted to look up a word. By the end of the experiment, another attitude questionnaire was given to find out the participants' attitudes towards the chatterbot after the experiment. Finally, the researcher gave the participants a delayed-post-test after one week to check their retention.

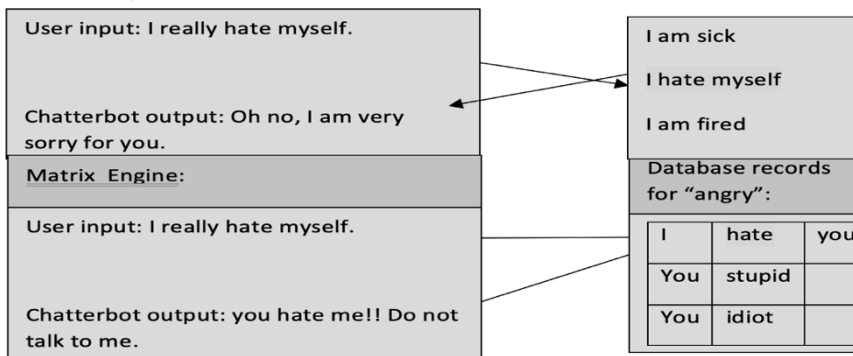


Figure 1. The difference between the special and the matrix engine

The special engine was more accurate at the sentence level, but the matrix engine was also important because it covers more possibilities than the special engine. The two engines actually complemented each other, as the matrix engine searched at the word level in all the entries and then sent the information to the special engine to find an exact match of a sentence. A third engine was the stories engine, which searched for all the stories stored in the local database. The chatterbot displayed a list of all the stories in the local database to enable the user to choose from them. Figure 2 shows behind the scenes operations for this engine.

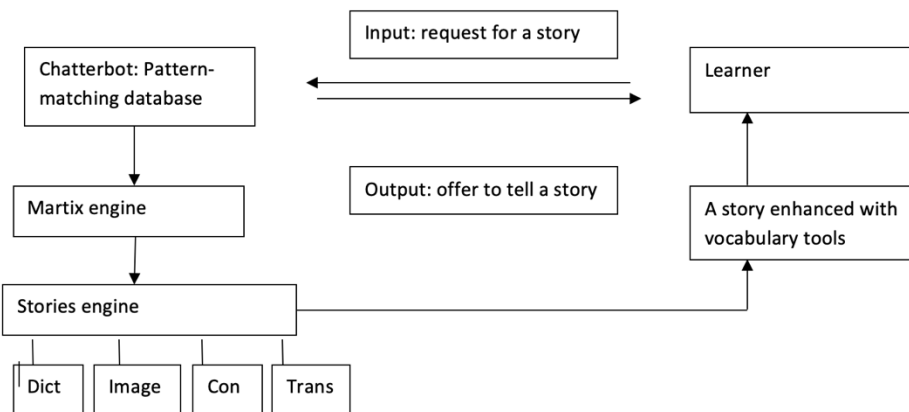


Figure 2. Behind the scenes operations for the stories' engine

From each story, 33 target words are chosen in the first phase of the study while in the second phase, 50 words for each story are selected and connected with all of the tools (see Figure 3 & 4).

Storyteller: Nice to meet you Redrose. I am a storyteller bot. I have just been born but incredibly I know how to speak. Mainly, I like to tell stories. I have hear one?
You: yes, please
Storyteller: I have many stories to tell. They happened to me or my friends. Choose from the following stories:
 1- The Cost of Hope
 2- The Scar
 3- Flying A Kite
 4- Chuleta
 5- At the Foot of the Bed
 6- Mother's Hands

Figure 3. List of stories from the local database of stories

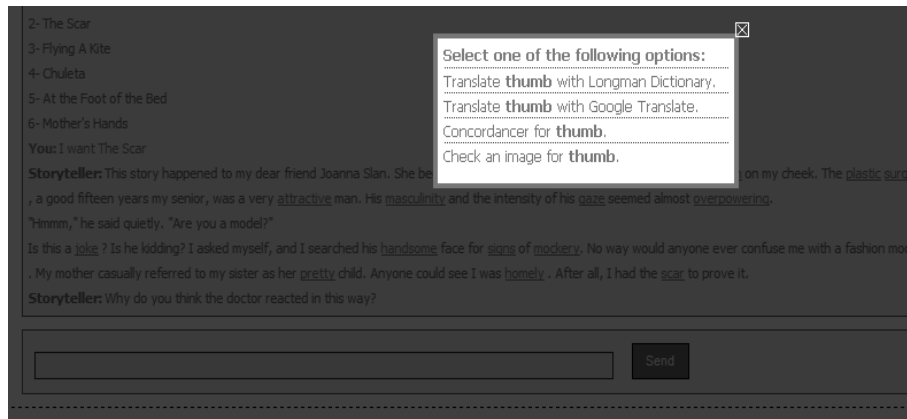


Figure 4. The vocabulary learning tools connected to a target word

4. Results

4.1 RQ1: What are the Participants' Preferences Regarding the Available Vocabulary Learning Tools

The participants' preferences were analysed from their screen recordings plus the number of clicks. The statistics meter in the chatterbot counted the number of clicks for the participants for the second phase of the study only. From the number of clicks, the dictionary had the highest number of clicks (1152). The images seemed to be the learners' second preference with 823 clicks. However, their screen recordings showed that the participants usually looked up the meaning of the word in the dictionary, and then checked the image option for the same word. The L1 translation came in third place with 741 clicks. Yet again, screen recordings showed that the participants tended to first look up in the dictionary, and then use the L1 translation tool. Sometimes, they tried out the dictionary, L1 translation and images. The concordancer seemed to be the least favoured tool with 83 clicks. Screen recordings showed that the participants always used one of the other tools when using the concordancer.



Figure 5. Learners' preferences of the tools by number of clicks

4.2 RQ2: What is the Effect of the Tools on Vocabulary Learning and Retention

The data set were statically manipulated in different ways to answer the research question about vocabulary learning and retention from different angles. First, a general comparison of the participants' performance from the pre-test to the post-test was carried out by performing a paired-samples t-test. The results revealed that the participants scored significantly higher on the post-test ($M=65.65$, $SD=10.57$) than they did on the pre-test ($M=36.85$, $SD=10.98$). The paired-samples t-test for this difference of 28.80 was significant, $t(19)=21.14$, $p<.001$. This suggests significant improvement from pre-test to post-test, suggesting that, as a group, the tools are effective agents of learning.

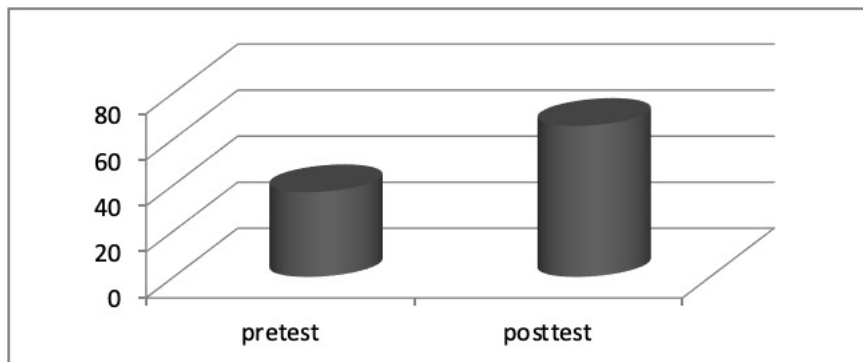


Figure 6. General performance from the pretest to the posttest

Then, repeated measures ANOVA was carried out to analyse the participants’ performance in the post-test with regards to the four tools. The overall repeated-measures ANOVA were significant, $F(3, 57)=73.91, p<.001$. Least Significant Difference (LSD) pairwise comparisons revealed significant group mean differences between all comparisons, except between dictionary and translation. The results revealed the following: the image tool ($M=13.65$) was significantly lower than the dictionary ($M=22.85$); the image tool was significantly higher than the concordancer ($M=8.65$); the image tool was significantly lower than translation ($M=20.50$); the dictionary was significantly higher than the concordancer; and translation was significantly higher than the concordancer. Thus, the results revealed that the dictionary and translation are the most effective tools; that the concordancer is the least effective; and that the image tool is in the middle.

Table 1. The participants’ performance in the post-test with regards to the four tools

Tools	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Image	13.650	.847	11.877	15.423
Dict.	22.850	1.141	20.462	25.238
Con.	8.650	.642	7.307	9.993
Trans.	20.500	.819	18.785	22.215

Dict.=Dictionary; Con.=Concordancer; Trans.=Translation

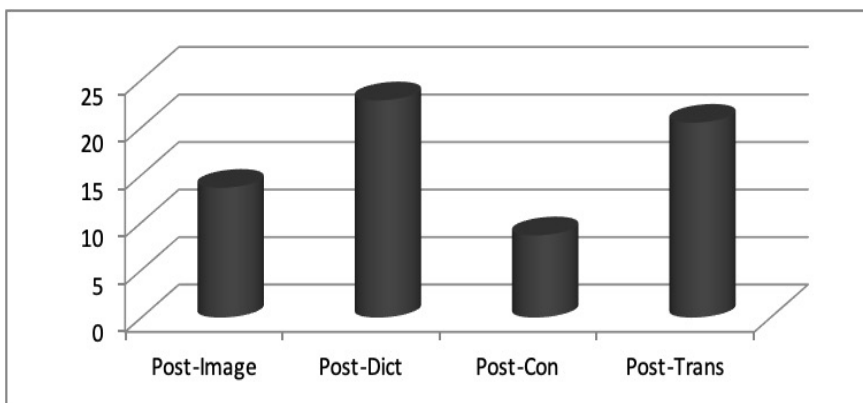


Figure 7. The participants’ performance in the post-test with regards to the four tools

To look at vocabulary retention, the participants’ general performance from the post-test to the delayed-test was compared using a paired-samples t-test. The results indicate that the participants scored significantly higher on the post-test ($M=65.65, SD=10.57$) than they did on the delayed test ($M=55.00, SD=9.55$). The paired-samples t-test for this difference of 10.65 was significant, $t(19)=11.22, p<.001$. This suggests significant knowledge loss after one week.

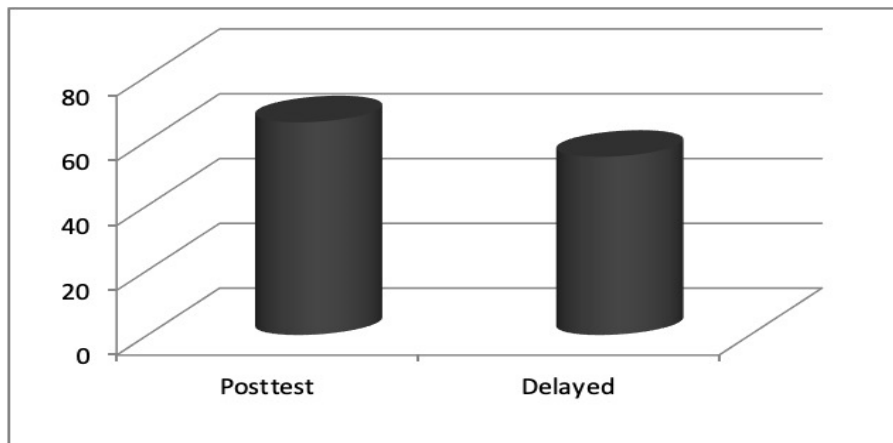


Figure 8. The participants’ general performance from the post-test to the delayed test

However, to answer the question about which tool helps more in retention, a repeated measures ANOVA was carried out. The results show that the overall repeated-measures ANOVA was significant, $F(3, 57)=66.69, p<.001$. Further, LSD pairwise comparisons revealed significant group mean differences between all comparisons, except between dictionary and translation. The study found that for the delayed test, image ($M=11.85$) was significantly lower than dictionary ($M=17.05$); that the image tool was significantly higher than concordancer ($M=7.75$); that the image tool was significantly lower than translation ($M=18.35$); that dictionary was significantly higher than concordancer; and that translation was significantly higher than concordancer. Thus, the study revealed that dictionary and translation are the most effective tools; that concordancer is the least effective; and that image is in the middle.

Table 2. The participants’ performance in the delayed test with regards to the four tools

Tool	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Image	11.850	.741	10.299	13.401
Dict.	17.050	.972	15.016	19.084
Con.	7.750	.517	6.667	8.833
Trans.	18.350	.674	16.940	19.760

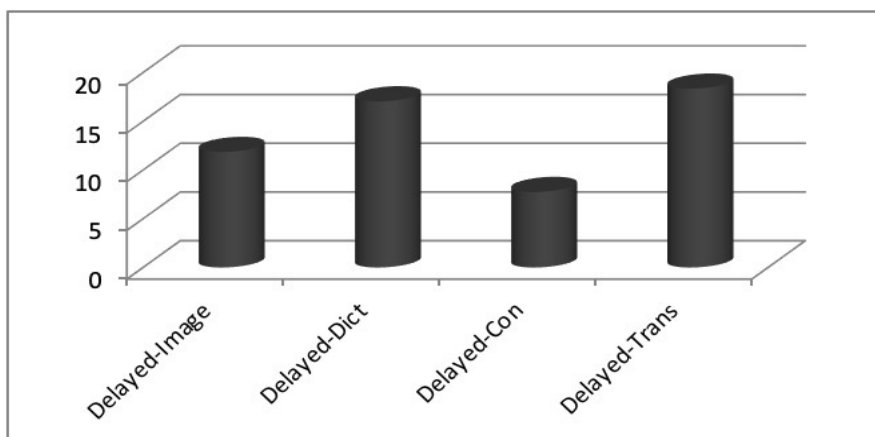


Figure 9. The participants’ performance in the delayed test with regards to the four tools

Finally, a repeated measures ANOVA was performed to compare participants’ gain of vocabulary in the pre-test, post-test and delayed test. The overall repeated-measures ANOVA was significant, $F(2, 38)=270.35, p<.001$.

Further, LSD pairwise comparisons revealed significant group mean differences between all comparisons. The study found that the pre-test (M=36.85) was significantly lower than both the post-test (M=65.65) and the delayed test (M=55.00), and that the post-test was significantly higher than the delayed test. This suggests that: (1) the tools significantly improve performance from pre-test to post-test; (2) that between the post-test and delayed test, a significant amount of information is lost; but that (3) even after a week, a significant improvement from pre-test is observed.

Table 3. The participants' performance in the pre-test, post-test and delayed test

Time	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Pre-test	36.850	2.454	31.714	41.986
Post-test	65.650	2.363	60.703	70.597
Delayed	55.000	2.136	50.529	59.471

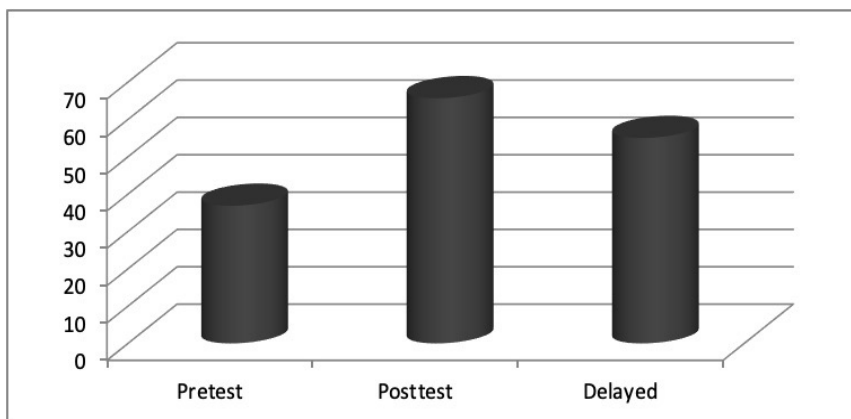


Figure 10. The participants' performance in the pre-test, post-test and delayed test

The results indicate that the participants in both tests improved from the post-test; however, the improvement from the post-test to delayed test declined to some extent. Figure 11 is a spaghetti plot showing the individual trajectories from pre-test to post-test to delayed test.

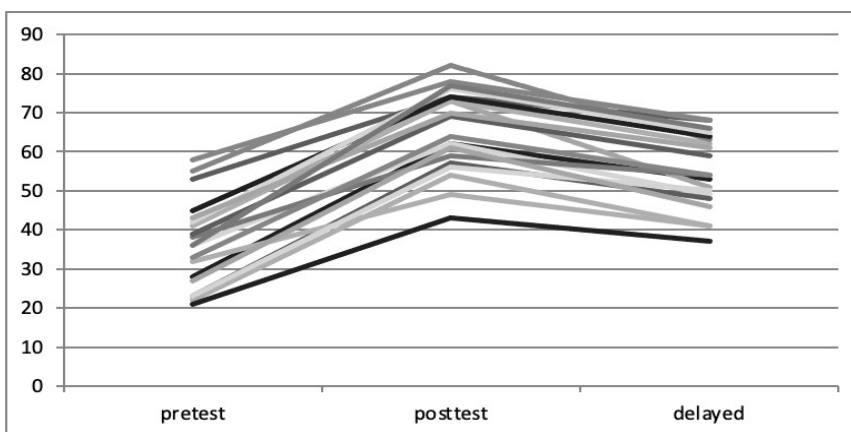


Figure 11. The individual trajectories from pre-test to post-test to delayed test

4.3 RQ3: What are the Participants' Attitudes towards the Vocabulary Learning Chatterbot Before and After the Experiment

A pre- and post-questionnaire were given to participants. The pre-questionnaire indicated that 85% of the participants had not used the chatterbot before. The chatterbot's function was explained to the participants prior to the experiment with examples such as Alice and Eliza. From the pre-questionnaire, 75% of the participants expected the chatterbot to be beneficial for language learning and 65% believed that the chatterbot would give the same experience as chatting with native speakers. Moreover, 60% of the participants expected to have a fun time with the chatterbot. The post-questionnaire shows that 80% of the participants believed that the chatterbot was beneficial for language learning and 85% found that the chatterbot gave an authentic native-like conversational practice. Furthermore, 70% of the participants believed that they could remember and use the vocabulary they learned from the chatterbot. The post-questionnaire also shows that 70% enjoyed their time while learning from the chatterbot.

For the design of the chatterbot, 90% of the participants reported that they liked the background colour of the page. On the other hand, 75% of the participants found the red colour of the text was uncomfortable to the eyes. Finally, the participants reported that the hyperlinking to the dictionary, concordance and the translation was not as easy for them to use as for the images. The reason was that it opened on the same page of the chatterbot, which created some technical problems when they clicked on the "back" button.

5. Discussion

The results indicate that the participants' preference for the dictionary turned out to be beneficial for their learning. The results support Laufer and Hill's claim (2000) that if a pedagogical tool is favoured by students, the chances are that it will also be beneficial for learning. Also, the results are consistent with those of Horst, Cobb, and Nicolae (2005), who reported that the dictionary is found to be the most used for looking up the meaning of new vocabulary. Even though the dictionary was found to be more effective than images and concordance for vocabulary learning, it was still slightly insignificantly higher than L1 translation. Conversely, L1 translation was found to be slightly higher than the dictionary for retention. The results are consistent with those of Laufer and Hill (2000). They reported that the dictionary, generally, is very effective for vocabulary learning; however, L1 translation is more effective for retention. On the other hand, it disagrees with Chen (2017), who emphasized that dictionaries have no effect on vocabulary retention. Peters (2007) and Gürkan (2019) reported that enhancement techniques such as definition glosses and pictures help in retention.

The current study's analysis of screen recordings of the second phase support these findings in the sense that the participants mostly used the dictionary first, then they checked the meaning again using images. In fact, they tended to make their own combinations of choosing between the dictionary, images and L1 translation. Nonetheless, they tried to avoid using the concordancer and, if they did, they always used it with another tool. In the post-questionnaire, 70% of the participants reported that they prefer the dictionary because their teachers recommend using it. One of the participants stated that, "my teacher likes us to use the dictionary because translation is not good for learning".

For the concordancer, the results of the study are inconsistent with the findings of John (2001) and Lee and Liou (2003). John (2001) encouraged the use of a concordancer, especially for beginners, while Lee and Liou (2003) emphasized that a concordancer helps low-vocabulary proficient learners to catch-up with advanced learners. The current study revealed that the concordancer is the least effective tool for vocabulary learning and retention for high-beginners. In the second phase, their screen recordings plus the number of clicks showed that the participants tended to avoid using it. In the first phase, the results indicate that it is the least beneficial tool. Moreover, the post-questionnaire showed that 80% of the participants did not find the concordancer easy-to-use and hence, they reported that it is not beneficial for learning. One of the participants commented that "the concordancer is really confusing for me". Lee, Warschauer, and Lee (2019) indicated that learners benefited from the concordance only after they had used it for a while over the course of the study.

In the attitude questionnaires, the participants believed that the chatterbot provided authentic and native-like conversational practice for them. They enjoyed chatting and learning at the same time. Harless et al. (1999) reported that that the participants were motivated to study and learn through the opportunity of speaking directly with a native-like speaker. Also, the findings of the questionnaires support Fryer and Carpenter's argument (2006) and that of Alm and Nkomo (2020), that chatterbots help language learners with fluency, as chances in the classroom are not always available. One of the participants stated that, "I love this chatterbot. I feel shy in class and I cannot speak. Here, I can speak anything I want".

6. Conclusion

The study revealed that the dictionary was the most favoured tool, which also happens to be the most effective tool for vocabulary learning. In fact, there was not a significant difference between the dictionary and translation. For retention, the findings show that L1 translation was slightly (but insignificantly) higher than the dictionary. Generally speaking, the results indicate that the learners' vocabulary knowledge improved by using the vocabulary learning chatterbot. However, a significant amount of information was lost between the post-test and the delayed test. Yet, a significant improvement after two weeks from the pre-test to delayed test was still observed. There are two main limitations that can be avoided in future research to obtain more valid results. First, future research needs to be carried out on larger sample of participants so the results can be generalized. Second, it should be applied over a longer period of time to test the treatment effect, as well as the retention level over a longer stretch of time. Furthermore, it would be interesting to measure the impact of conversing with the chatterbot on speaking with real native speakers or other language learners.

References

- Alm, A., & Nkomo, L. M. (2020). Chatbot Experiences of Informal Language Learners: A Sentiment Analysis. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 10(4), 51-65. <https://doi.org/10.4018/IJCALLT.2020100104>
- Atwell, E. (1999). *The Language Machine: The impact of speech and language technologies on English language teaching*. London: British Council.
- Brennan, K. (2006). The Managed Teacher: Emotional Labour, Education, and Technology. *Educational Insights*, 10(2), 55-65.
- Lu, C. H., Chiou, G. F., Day, M. Y., Ong, C. S., & Hsu, W. L. (2006). Using Instant Messaging to Provide an Intelligent Learning Environment. In Ikeda M., Ashley, K. D., & Chan, T. W. (Eds.), *Intelligent Tutoring Systems* (pp. 575-583). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11774303_57
- Chapelle, C. (1997). Call in the year 2000: Still in search of research paradigms? *Language Learning and Technology*, 1, 19-43.
- Chapelle, C. (1998). Multimedia Call: Lessons to be learned from research on instructed SLA. *Language Learning and Technology*, 2, 22-34.
- Coniam, D. (2008). An evaluation of chatbots as software aids to learning English as a second language. *EuroCALL Review*, 13.
- Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*, 93, 279-289. <https://doi.org/10.1016/j.chb.2018.12.023>
- Fryer, L., & Carpenter, R. (2006). Emerging Technologies: Bots as Language Learning Tools. *Language Learning and Technology*, 10(3), 8-14.
- Gürkan, S. (2019). Effect of annotation preferences of the EFL students' on their level of vocabulary recall and retention. *Journal of Educational Computing Research*, 57(6), 1436-1467. <https://doi.org/10.1177/0735633118796843>
- Harless, W. G., Zier, M. A., & Duncan, R. C. (1999). Virtual Dialogues with Native Speakers: The Evaluation of an Interactive Multimedia Method. *CALICO Journal*, 16(3), 313-337.
- Horst, M., Cobb, T., & Nicolae, I., (2005). Expanding academic vocabulary with an interactive on-line database. *Language Learning and Technology*, 9(2), 90-110.
- Hubbard, P. (Ed.) (2009). *Computer Assisted Language Learning: Critical Concepts in Linguistics*. London & New York: Routledge.
- Jiyoun Jia. (2008). An Intelligent Web-based Human-Computer Dialogue System for Language Instruction: CSIEC. *Asia Research Network*, 5(2), 42-45.
- Jiyoun Jia. (2004). CSIEC (Computer Simulator in Educational Communication): An Intelligent Web-Based Teaching System for Foreign Language Learning. In P. Kommers & G. Richards (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 4147-4152). Chesapeake, VA: AACE.

- Kim, H. S., Cha, Y., & Kim, N. Y. (2020). Impact of Mobile Interactions With AI on Writing Performance. *Modern English Education*, 21(2), 1-13. <https://doi.org/10.18095/meeso.2020.21.2.1>
- Kim, N. Y. (2017). Effects of Types of Voice-Based Chat on EFL Students' Negotiation of Meaning According to Proficiency Levels. *English Teaching*, 72(1). <https://doi.org/10.18095/meeso.2017.18.1.03>
- Laufer, B., & M. M. Hill. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning and Technology*, 3(2), 58-76.
- Lee, C. Y., & Liou, H. C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching & Learning*, 27(3), 35-56.
- Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5), 721-753. <https://doi.org/10.1093/applin/amy012>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie & T.K. Bathia (Eds.), *Handbook of research on second language*. San Diego, CA: Academic Press. <https://doi.org/10.1016/B978-012589042-7/50015-3>
- Oxford, R. (2001). Language learning styles and strategies. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed.). Boston, MA: Heinle & Heinle.
- Peters, E. (2007). Manipulating L2 learners' online dictionary use and its effect on L2 word retention. *Language Learning & Technology*, 11, 36-58.
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15(2), 165-179. <https://doi.org/10.1017/S0272263100011943>
- Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 237-257). Cambridge, England: Cambridge University Press.
- St. John, E. (2001) A Case For Using a Parallel Corpus and Concordancer For Beginners of a Foreign Language. *Language Learning & Technology*, 5, 185-203.
- Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning and Technology*, 3(10), 85-101.

Appendix A

Questionnaire

Consent Form

Investigators: ***** , assistant professor at Imam Mohammed Ibn Saud Islamic University in Riyadh, Saudi Arabia. The purpose of this study is to find out more about the use of chatterbot to learn English. Benefits of the study include better understanding of language phenomenon as well as adding more data to the domain of Language learning. You will not be required to perform any activities other than chatting with the chatterbot and answering the questions. Any information that is obtained during this study will be kept confidential to the full extent permitted by law. The experiment will be conducted on a secured server. Your real name will not be reported in any resulting publications. If at any point during or after the performance you decide that you wish to withdraw your participation I will erase the data gained. Having been asked to participate in the research study named above, I certify that I have read the procedures specified in the Study Information Document describing the study. I understand the procedure to be used in this study and the personal risks to me in taking part in the study as described. I can obtain copies of the results of this study, upon its completion by contacting the investigator at *****By filling out the survey, you are agreeing to participate.

Q1. Background Information:

Name (Nickname) : *

Age : *

Gender : *

E-mail : *

How long have been learning English? : *

Q2. How do you evaluate your English Vocabulary?

*

Chatterbot

According to Chatbots.org, it is a term coined by Michael Mauldin (1994) who created a bot named Julia in an online MUD environment. He coined the term “chatterbot” to describe his conversational programs. Simple Definition is : a server application that can chat as if it were human. Synonyms: virtual agents, interactive virtual agents (IVA), brand agents, branded agents, virtual assistants, artificial conversational entities, virtual employees, conversational agents, embodied conversational agents (ECA), chatterbots, conversational systems, chatterbox, virtual chat agents, or chat bots. Example: <http://alice.pandorabots.com/>

Q3. Choose one level of rating for each question.

	Very much	Somewhat	Not sure	Not really
Have you ever heard of a chatterbot? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
From your understanding of what is a chatterbot, do you think it will be beneficial for Language learner? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you expect to have fun time with the storyteller bot? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you think that a chatterbot will help learner practice English? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you think that the chatterbot will give the same experience as chatting with a native speaker? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q4. Choose one level of rating for each question.

	Always	Sometimes	Every once in a while	Rarely	Never
Have you ever used a chatterbot? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q5. Choose one level of rating for each question.

	Exciting	Fun	Neutral	Stupid	Boring
If yes, choose the word that best describes your experience.:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Conversational Stories:

Conversational stories: Personal experience stories in natural conversation. Example: A: Yesterday, a really scary funny thing happen to me! B: What happened? A: I was about to use hair removal product on my hair. can you imagine? B: WHAT? A: Yah, I went to the supermarket and I was in hurry so, I took the hair removal foam instead of the hair curly foam

Q6. Choose one level of rating for each question

	Yes	Maybe	No
Do like conversational stories? Why? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vocabulary Learning Tools

This is to explain a term you will find in this section: Concordancer: It is a piece of software, either installed on a computer or accessed through a website, which can be used to search, access and analyse language from a corpus. They can be particularly useful in exploring the relationships between words and can give us very accurate information about the way language is authentically used.

Q7. Vocabulary Learning Tools

	Very much	Somewhat	Not sure	Not really
Would you like if the chatterbot is equipped with a dictionary :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Would like to use a dictionary while chatting in a synchronous conversation? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you usually use a dictionary while chatting online? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you find it easy to use a dictionary while chatting online? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you prefer to check the new vocabulary through images? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you check a meaning of new word through images search engines? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do find it easy to search for the new word by images search engines while chatting? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Would you like it if the chatterbot has hyperlinks to images for the new words? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Have ever tried a concordancer to check the word in different contexts? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you think a concordancer is easy to use? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you prefer using a concordancer to figure out the meaning of new words from context? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Would you prefer using a translation tool while chatting online? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you like translating new words? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q8. If you have the option to choose between the followings, which one will you choose?

Select One...
*

Q9. Tell me why you chose this option?

*

Thank you for your feedback, your opinion is important to us.

Appendix B**Sample from pre-test's questions****Pre-test****Instructions:**

Choose YES if you are sure you know the meaning of the word. Circle NS if you have an idea about the meaning but you are not sure. Circle NO if you do not know the word. Do not worry if do not know the word. Just answer as honestly as possible.

Pre-test**Q1. Childhood** YES NS NO**Q2. Longstanding** YES NS NO**Q3. Custom** YES NS NO**Q4. Forehead** YES NS NO**Q5. Annoying** YES

NS

NO

Q6. worn

YES

NS

NO

Appendix C

Post-test

Instructions

What do you know about these words? Please choose 1, 2, or 3 and complete:

Childhood

I don't know what this word means.

I am not sure. I think it means (you can write the meaning in English or Arabic)

I know this word. It means.....and I can use it in a sentence (write a sentence).....

Longstanding

I don't know what this word means.

I am not sure. I think it means (you can write the meaning in English or Arabic)

I know this word. It means.....and I can use it in a sentence (write a sentence).....

Custom

I don't know what this word means.

I am not sure. I think it means (you can write the meaning in English or Arabic)

I know this word. It means.....and I can use it in a sentence (write a

sentence).....

Forehead

I don't know what this word means.

I am not sure. I think it means (you can write the meaning in English or Arabic)

I know this word. It means.....and I can use it in a sentence (write a sentence).....

Annoying

I don't know what this word means.

I am not sure. I think it means (you can write the meaning in English or Arabic)

I know this word. It means.....and I can use it in a sentence (write a sentence).....

Worn

I don't know what this word means.

I am not sure. I think it means (you can write the meaning in English or Arabic)

I know this word. It means.....and I can use it in a sentence (write a sentence).....

Rough

I don't know what this word means.

I am not sure. I think it means (you can write the meaning in English or Arabic)

I know this word. It means.....and I can use it in a sentence (write a sentence).....

Skin

I don't know what this word means.

I am not sure. I think it means (you can write the meaning in English or Arabic)

I know this word. It means.....and I can use it in a sentence (write a sentence).....

Lashed out

I don't know what this word means.

I am not sure. I think it means (you can write the meaning in English or Arabic)

I know this word. It means.....and I can use it in a sentence (write a sentence).....

Appendix D

Post-Questionnaire

Q1. Contact Information

Name (Nickname): *

E-mail: *

Chatterbot

In this section, you will be asked to evaluate you experience with the chatterbot.

Q2. Choose one level of rating for each question.

	Very interesting	Somewhat interesting	Neutral	Not very interesting	Not at all interesting
How did you find your experience with the chatterbot? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q3. Choose one level of rating for each question.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
From your expereince with the storyteller bot, do you think it is benefinical to ESL learners? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you have a good practice with the storyteller bot? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you think that the chatterbot gives the same experience as chatting with a native speaker? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do you think the storyteller helps you to learn vocabulary? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Do you remember some of vocabulary you learnt while chatting with the storyteller bot? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Was it a fun learning experience for you? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Chatterbot Design

In this section, you will be asked to evaluate the design of the storyteller bot.

Q4. Choose one level of rating for each question.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Did you find the colour of the page comfortable to your eyes? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you find the hyperlinked words easily? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you like the colour of the font? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you find the images clear? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q5. What do you suggest to add, change, or modify to improve the storyteller bot?

Conversational stories

In this section, you will be asked to evaluate the stories of the storyteller bot.

Q6. Choose one level of rating.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Did you like chatting about stories? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you like the stories? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did the storyteller bot succeed in making the stories interactive for you? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q7. Evaluating the stories.

Very easy Easy Medium Difficult Very Difficult

Evaluate the level of the stories. :

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Q8. If you have any suggestion about the stories, please write it down.

Vocabulary Learning Tools

In this section, you will be asked about these tools: Dictionary Concordancer Translation tool Images

Q9. Please rate the following:

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Did you like using the dictionary while chatting? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Was it easy for you to use the dictionary while chatting? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Was it useful for you to use the dictionary while chatting? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did the images convey the meaning clearly? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Was it easy for you to check the meaning through images? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you find the concordancer helpful in learning the new words? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you find it easy to use the concordancer? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you find it useful to check the words in different contexts? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you like using the translation tool? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you find the translation tool easy to use? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Did you find it useful for learning new words? :	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q10. After using the storyteller bot, choose your favorite tool.

*

Q11. Could you please explain why?

*

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).