| Research Article |

Measuring Efficacy of Speaking English Chatbot NUMLINA: A User Study

1. (Corresponding Autor) Dr Zafar Ullah, Assistant Professor, Department of English, University of Education, Lahore (Vehari campus) , zafarullah@ue.edu.pk ORCID: https://orcid.org/0000-0003-3773-1467
2. Dr Saeed Ullah, Assistant Professor, Department of Education, University of Education, Lahore, (Vehari campus), saeedullah@ue.edu.pk
3. Ms Nauzaina Khalid, Research Associate, National University of Modern Languages, Islamabad, ibu.kk111@gmail.com
4. Muhammad Mansoor Abbas, PhD Scholar Department of Psychology, Riphah International University Faisalabad Campus, veharicollege@gmail.com

Abstract: *The utilisation of Natural Language Processing has enabled us to build conversational chatbots. In particular, pedagogical settings necessitate the utilisation of a speaking chatbot because writing chatbots are already functional in the industry. There was a need for such a conversational chatbot which could fulfil pedagogical needs. It was hypothesised that a chatbot could teach English speaking skills better than a conventional teacher. First, the research team built the NUMLINA chatbot with DialogFlow, a Google built-in automated infrastructure supported by Artificial Intelligence and machine learning. This study aimed to teach speaking English skills effectively to students with NUMLINA chatbot. This user study followed an experimental research paradigm to measure the effectiveness of the newly built NUMLINA chatbot while speaking English with a human being. Comparing the post-test of the controlled and experimental group's means, 1.4 and 1.9 in fluency, 1.1 and 1.8 in vocabulary, 1.3 and 1.8 in pronunciation, 1 and 1.4 in learning idioms, and 1.3 and 1.7 in communication. The experimental group outperformed in the five speaking English categories validating the hypothesis. Thus, it promoted autonomous learning by advocating the modern teaching method for learning English.*

Keywords: *Chatbot, English speaking, E-learning, human-computer interaction, CALL system.*

Introduction

This research introduces the novel idea of chatbot utilisation in English language learning, especially English-speaking skill. There have always been new ways of learning and gaining knowledge in education (Akram et al., 2021), and this study promotes Computer Assisted Language Learning. It is evident that CALL systems are familiar and have been used for decades. Still, one should know that Computer Assisted Language Instruction was in use before CALL (Abdelrady & Akram, 2022). CALI has been replaced by CALL systems, which have become more popular and useful in academia while promoting autonomous learning.

However, a chatbot is an advancement in Computer Assisted Language Learning, and now it can replace traditional computer programs used in the past. Chatbots can be used for different tasks and educational purposes; for instance, it has been seen that multiple tasks, such as customer care, have been dealt with by chatbots. It can be said that customer care has always been a task for humans, but now it is taken over by a virtual conversational agent, virtual assistant or chatbot; therefore, one must realise that chatbots are now capable of conversing like human beings. Thus, English speaking can also be

learned with chatbots, and it has been proved in this research that a chatbot can aid a human teacher. The study also explores how other studies have advocated CALL systems and chatbots for learning a new language. Furthermore, educational institutes are inclined to use chatbots for teaching different subjects or topics.

Teachers and institutes have used the CALL (Computer Assisted Language Learning) systems for several years, and it is time to revolutionise this field. The best way to revolutionise CALL systems is to expand their scope by introducing virtual tutors and chatbots trained according to individuals' needs. There has not been a better time for this study since the world is adapting to new technologically-enriched ways of learning and teaching.

However, the study specifically aims at learning English speaking with a chatbot because most English language learners do not get the chance to talk to a native speaker while learning the language, and that's why a chatbot speaks just like a native speaker or a substitute for a teacher. This study has proven that English language learners can quickly enhance their fluency, vocabulary and pronunciation. In addition, the chatbot also enables learners to ask a question unhesitatingly multiple times. This chatbot feature can also allow the learners to practise more, so learners can speak eloquently. This study also reviews other studies that have advocated using virtual tutors or chatbots for educational purposes. Since new English language learning methods are introduced and adopted worldwide, it is a suitable time for this study to show the importance and benefits of a chatbot for learning the English language. Now, education is one of the fields where chatbots have not been used extensively, and it is time to revolutionise education by introducing chatbots, heralding an AI-based pedagogical future.

Problem Statement

Hiring new teachers costs money, and there are some hard-hit places like Gwadar where teachers are reluctant to go. This is a significant problem because it needs more teachers to teach the English language correctly. Moreover, teachers cannot teach appropriately sometimes because they are humans. Introducing chatbots for language learning can be effective in English language learning, and anyone can count on the chatbots because they can be upgraded with time as well.

Furthermore, the learners do not learn the true expression of the English language, which is possible if the learners interact with native speakers. It is only possible for some language learners. The learners who learn the English language, specifically English speaking, cannot utter correct expressions like native speakers. Moreover, learning English speaking takes a long time, and they just focus on receptive language skills. Chatbots can successfully teach English speaking because learners can interact with the chatbot through speech. A chatbot can be a substitute for a native teacher.

Significance of the Study

This study is significant since it shows a new method for learning English. Furthermore, this study discovers new language learning methods, replacing old ideas. In addition, the least research on pedagogical and English-speaking chatbots has been carried out in Pakistan, necessitating this study.

Research Objectives

This study aims to accomplish the following objectives:

i. To find out the effectiveness and scope of the chatbot in learning speaking English skill.

ii. To find an authentic substitute to replace old pedagogic methods for teaching speaking English skills.

Hypothesis

This experimental research hypothesises that using the newly-built NUMLINA chatbot enhances speaking English proficiency in terms of fluency, vocabulary, pronunciation, use of idioms and phrasal verbs and communicative proficiency among BS English non-native learners at the National University of Modern Languages, Islamabad, Pakistan. In addition to it, chatbot promotes autonomous learning among learners.

Delimitations

The study is restricted to English-speaking skill. Elucidation of the programming of the chatbot used in this research is beyond the scope of this study; it only shows the usage of the chatbot for language learning purposes. NUMLINA chatbot (Ullah, 2023) is built with Google DialogFlow (a chatbot-building tool) to conduct this user study.

Literature Review

Natural Language Input activates the chatbots, and this input can be in textual or voice form. Now, chatbots are considered one class of intelligent and pedagogical tools. Chatbots have been in service since the 1960s. Chatbots can be trained easily, and their design is easier than before. The chatbots are capable of providing conversational output, and they can also perform tasks when they are asked or commanded. The evolution of chatbots has opened new ways of communication. It has become one of any corporate company's basic needs since chatbots are used for online queries at run time. However, this study aimed to provide evidence of how chatbots can be used for language learning. Ramzan et al. (2023) have claimed that by identifying the influence of social media use, educators and policymakers may create plans to maximize its potential for raising academic engagement and performance among ESL students. Kerly, Hall, & Bull (2007) discussed two chatbots, A.L.I.C.E and The Jabberwacky. Both chatbots were tested to determine if they were appropriate for education. The study explored the use of these two chatbots, how they have been successfully taught, and how they also facilitated students. However, the study concluded that students' learning experience was enhanced when two distinct chatbots were used one responded to learners' emotions, and the other only responded to students' questions; and it also assigned tasks. As a result, students learnt various things easily. Ramzan et al. (2023) by recognizing the interconnectedness of various writing facets educators can more foster more holistic development of writing skills among ESL undergraduates.

Fei & Petrina (2013) stated, "Chatbot is least explored regarding its efficacy in second language learning because the technology in this function is still under development and has not been widely applied yet". Moreover, a chatbot named Lucy was used as a language tutor. Lucy was shown to be capable of teaching the English language. The study also advocated use of chatbots in educational institutes, showing that Lucy chatbot was successful in learning English.

Now, Human-machine interaction has been evolving, and Human-Computer Speech is becoming

more popular since there are many platforms available on the internet where one can avail of this facility. A study showed that there had been an expansion in speech-based search engines and assistants such as Siri, Google Chrome and Cortana. Natural Language Processing (NLP) techniques such as Natural Language Toolkit (NLTK) for Python can be applied to analyse speech. Intelligent responses can be found by designing an engine to provide appropriate human-like responses; hence this programme is called a Chatbot (A. & John, 2015).Ramzan et al.(2023) have confirmed about motivation of ESL learners that this research would be a useful insight for English teachers and students in an intracultural environment infused with ethnicity and multi-gender student population

While studying other studies about the usage of chatbots, it became clear that chatbots have evolved enough to interact with humans through speech. A. & John (2015) discussed that speech interaction with computing devices had received increasing interest in the past few years with contributions from Google, Android and iPhone Operating Systems. The research article also showed that spoken dialogue systems successfully formed the primary interaction method with a machine. Thus, speech interaction will play a significant role in humanizing machines in the near future. Bhutto and Ramzan (2021) have said that the state must form a policy to curb the roots of such inhuman crimes by producing the remedies.

In addition, the study found that some of the most cited research articles considered that chatbots could influence users' behaviour by asking and responding to the users' questions. "Universally, a Chatbot is a computer programme that mimics intelligent conversation; the input to this programme is natural language text" (A. & John, 2015). Ramzan et al. (2023)'s findings have implications for ESL teachers in designing effective language instruction that can help students improve their writing skills and avoid common subject-verb agreement errors.

"Speech is considered as a signal, and it happens at different levels: "semantic, linguistic, articulatory, and acoustic" (A. & John, 2015). It also claimed that speech was the most natural among the aspects of human communication. The study results showed that spoken words always have a different impact on any person than written words, speech is more complex, and it takes time to understand the language and its particular style of speech.

Khanna, Pandey, Vashishta, Kalia, Pradeepkumar, & Das (2015) showed how chatbots could be intelligent enough to converse with a human being. The Turing Test helped to determine the success ratio of chatbots during their conversation with humans. However, the study did not deal with actual chatbot building, yet it mentioned building intelligent chatbots and their successful performance in the past.

Previous studies discussing the usage of chatbots for learning a language revealed that the current research studied the roots of chatbots, CALL and technologies that helped in second language learning. In addition, Hattem & Lomicka (2016) stated that for more than two decades, technological development had played a vital role in language learning. The significant findings of this study showed that advanced technological tools shaped the research in language and education. However, these advanced or innovative tools included email, chat, wikis, and social media. Hattem and Lomicka (2016) considered Twitter one of the most important social media platforms or tools for education and learning. Moreover, since it opened new teaching and learning methods, Twitter has been vital in academic and

non-academic settings.

Studying Twitter was to understand the nature of the conversation between humans and a pedagogical agent. The responses by Twitter autonomous accounts were considered human-like or natural. Hattem & Lomicka (2016) showed seventeen studies, and their results claimed that Twitter had been used between classes, teachers and students, students and students, or even celebrities or well-known tweeters. Furthermore, the use of Twitter in language learning has progressed significantly since 2009. "Twitter serves as a strong pedagogical tool to allow students to use the timeline inside and outside of the classroom, to interact with classmates, the teacher" (Hattem & Lomicka, 2016).

One study was about utilising Twitter to understand the importance of conversational systems; making conversational systems more natural and human-like for learning English speaking is one of the purposes of conducting this study. However, this research was made possible with the collaboration of IBM Research Centre, having twelve labs on six continents, created a conversational system that could automatically generate responses to users' requests on social media (Xu, Liu, Guo, Sinha, & Akkiraju, 2017). Furthermore, Xu et al. (2017) discussed that users positively responded to interacting with a chatbot on social media. The key findings suggested that the chatbots, which were used on social media, provided individualised attention to the users and made the users more positive about the interactions between users and brands. In addition, the study suggested that future studies could be about understanding how chatbots affect the relationship between users and the brand. Xu et al. (2017) showed people's interest in interacting with chatbots. It can be essential for the scope of the research since more educational institutes use chatbots for learning a language or learning the English language in Pakistan.

However, chatbots have been used for different types of work or tasks. A study discussed the usage of chatbots claimed that during the Presidential election of US in 2016, a fifth of the comments, answers or responses on Twitter were driven by autonomous Twitter accounts; the chatbots or pedagogical agents were used for the service, and they controlled the communication with the users (Radziwill & Benton, 2017).

Huang, Lee, Kwon and Kim (2017) discussed two chatbots, GenieTutor and GenieTutor Plus. GenieTutor was a task-oriented chatbot; on the other hand, GenieTutor allowed free conversations. The study found that both chatbots successfully taught learners a second language. Even though GenieTutor Plus was more successful than GenieTutor in teaching the language, but still, GenieTutor helped the students to learn. This study also helped to understand different types of chatbots were used to accomplish one purpose. Teachers' attitude towards the use of chatbots in routing teaching was analysed. P. K, et al. (2018) discussed the usage of chatbots in education. The study concluded that the majority of the teachers favoured the use of chatbots for teaching purposes.

Ruan et al. (2019)'s work was presented at ACM Conference on Learning, and it helped to find some authentic evidence to support the study. The study presented an experiment where five 6-year-old native Chinese-speaking children studied English as a foreign language. The study recorded the facial expressions and emotions of the children when they were interacting with the system for learning the English language. The results showed that children were having fun interacting with the machine and were learning English in a fun way. Moreover, it also stated that they were eager to talk to a chatbot and were learning naturally that they could not have learnt otherwise.

The voice user interface was used via Google Assistant in a TESOL class for speaking skill development. This study assessed the pedagogical benefit for native and non-native speakers. Its results highlighted comfort, comprehension, worthwhile, user friendly and enjoyable (Kent, 2021). The study conducted a meta-analysis of a conversational bot for second language practice. Thus, Morris and DeShon's (2002) formulas were utilised to measure effect size (Bibauw, Noortgate, François, & Desmet, 2022). The latest language learning scenario had been built with chatbots, augmented reality, and theories of sociomaterialism and 4E cognition (Embodied, Embedded, Enactive, Extended) supported them. Conversation partners and chatbots will build a metaverse where language encompasses all gestures, facial expressions and ambience (Godwin-Jones, 2023). Generative AI has revolutionised the entire social and communicative fabric with ChatGPT and Dall-E2 (Roose, 2022).

The aforementioned studies were significant since they provided enough data to support the claim that language learning was possible with chatbots. Moreover, solid evidence showed that language learners were learning the language positively, specifically English speaking.

Research Methodology

Research Design

Thirty students from BS English 8th semester participated in this user study from an undergraduate course taught at the National University of Modern Languages, Islamabad, Pakistan. A systematic sampling technique was followed for this experimental study. Out of thirty participants, an even number of fifteen participants were included in the control group, and the remaining odd number of fifteen students became part of the experimental group. The control and experimental groups were taught the same contents but through different methods; for instance, the control group learned English speaking through the conventional method as they were already learning at the National University of Modern Languages, Islamabad, Pakistan. On the other hand, participants in the experimental group learned English speaking with the NUMLINA chatbot (Ullah, 2023), which was developed and trained through DialogFlow, one of Google's platforms for building conversational agents or chatbots (Google, 2020). Chatbot is a recent addition to Human-Computer Interaction; therefore, chatbots are supposed to revolutionise HCI since they are smarter than other conventional HCI systems (Asbjørn, & Bae, 2017).

Furthermore, the participants were taught English speaking particularly, and the study's goal was to check whether teaching English through a chatbot was better than the conventional teaching method. Therefore, the experimental and control groups were taught the same contents; for instance, both learned fluency in spoken English by practising pronunciation, and they learned new words to enhance their vocabulary for better speaking. However, five lessons were organised for the study, and each session consisted of sixty minutes, and both groups received an equal amount of time during the treatment phase. Before starting the treatment phase, all of the participants took the test, and a similar test was taken after the treatment phase. The pre-test and post-test determine the efficacy of the treatment.

Data Collection and Analysis

The study organised the pre-test to learn about the competency of the individuals in English speaking before starting the treatment phase. The pre-test contained five questions, and each question carried ten points. The study asked questions, and the participants had to answer them. Then the research team

assigned points to each participant's answers by keeping these five things in mind: their ability to speak fluently, the usage of appropriate words, the correct pronunciation of the words, the usage of idioms and phrasal verbs, and their ability to communicate in English unhesitatingly. In conclusion, there were five questions, each with ten marks. The examiner also assigned extra ten marks based on fluency, pronunciation, vocabulary, use of idioms and phrasal verbs and ability to communicate in English. However, five elements were considered: fluency, pronunciation, vocabulary, use of idioms and phrasal verbs, and ability to communicate in English. Each contained two points (ten points collectively), and the research team assigned those separately. After the pre-test, their obtained marks were communicated; in this case, the total marks were sixty.

After the treatment phase, the participants took a post-test containing sixty marks; however, the questions asked in the test were not similar to the pre-test but were more complex. After taking the pre-test and post-test, the data were collected and analysed with Statistical Package for Social Sciences (SPSS) software.

Results and Discussion

Pre-test and post-test (1) evaluated participants' English speaking in general; (2) evaluated participants' fluency, vocabulary, pronunciation, use of idioms and phrasal verbs, and ability to communicate in English coherently. The research team asked five questions to each participant and then assigned marks to each question separately. However, the questions asked in the pre-test and post-test were similar, but post-test's questions were more difficult than the pre-test. For instance, in a pre-test, the research team asked the participants a question about them like this "What is your name; tell me a bit about yourself, like where are you from and what do your parents do?" On the other hand, the post-test question was, "Tell me about yourself in detail and tell me who you are and what you want to be?" However, when participants answered the questions, the research team assigned marks to those answers separately.

Five questions were asked in the pre-test are as follows:

1. What is your name; tell me a bit about yourself, like where are you from and what do your parents do?
2. Are you satisfied with this university, what's your experience so far?
3. Who is your favourite personality; tell me about them in detail like why you do like them and what do you like about them the most?
4. Do you watch movies, or are you more of a reader?
   4.1. Which is your favourite movie? Tell me its plot.
   4.2. Which book is your favourite? Tell me about it.
5. What do you think of feminism? Tell me about your thoughts on that.

Every participant in the experimental and control was asked the same questions, whether it was control or experimental. The participants were marked by noticing how participants answered the questions with the use of synonyms and phrasal verbs. Moreover, the fluency and pronunciation of the participant also played a vital role in scoring the marks. However, the results of each group are shown in

Table I to table 9. These tables show five questions, each question's marks are assigned separately, and the mean is calculated.

Table I Results of the pre-test of the control group

Participants	Question# I	Question#2	Question# 3	Question# 4	Question# 5	Obtained Total
01	5	4	3	3	2.5	17.5
03	4	3	4	5	4	20
05	4	3.5	3.5	3.5	2.5	17
07	5	4	4.5	5	3.5	22
09	4	4	4.5	5	4.5	22
11	6	6.5	7	6	5	30.5
13	5.5	5	4.5	4	3.5	22.5
15	4	5.5	4.5	3.5	2	19.5
17	3	3	2.5	3	2.5	14
19	6	5.5	5	6	4.5	27
21	6.5	5	4	4.5	3	23
23	5	5	5	5	5	25
25	7	6	6.5	7	5	31.5
27	5	5	4.5	4	4	22.5
29	6	4.5	4	4.5	4	23
Mean	**5.0**	**4.6**	**4.4**	**4.6**	**3.7**	**22.4**

These results showed that most of the participants scored average. They did not face any problem answering those questions. However, it was noticed that they did not use synonyms and needed to be more fluent, so they were assigned below 50% marks. Participants also did not score well on the fifth question. While answering the fifth question, participants lost coherence and could not connect their thoughts in words smoothly.

In the treatment phase, participants of the control group learnt English through the conventional method, where the tutor had to ask every student individually to speak on a particular topic. Afterwards, students were given a home task. In that home task, students had to learn synonyms, phrasal verbs and idioms. Furthermore, they were asked to practice speaking English as well. After five sessions of sixty minutes each, the post-test. Similar questions were asked in the post-test or post-treatment test, but they were more difficult than pre-tests.

Five questions were asked in the post-treatment test are as follows:

1. Tell me about yourself in detail and tell me who you really are and what do you want to be?
2. How did you end up in this university, tell me about your journey as a student like how you have planned to study in this particular course and how is your experience so far?
3. Who do you see as an inspiration, do you look up to someone as a perfect example? Elaborate their personality.

4. How do you spend your time when you have completed your tasks, how do you make yourself productive and how do you entertain yourself?
5. Since you are a university student, and everyone expects you to be a literary and well-educated person, tell me what is your stance on feminism as a modern movement?

These five questions were asked to both groups, and the research team assigned marks to every individual separately. The results of the post-treatment test of the control group are shown in Table 2.

Table 2 Results of the post-test of the control group

Participants	Question# I	Question#2	Question# 3	Question# 4	Question# 5	Obtained Marks
01	6.5	5	4	4	3	22.5
03	5	3.5	4	5	5	22.5
05	5	4	4	4.5	3	20.5
0	5.5	4	5	5.5	4	24
09	5	5	5	5.5	5	25.5
11	6.5	7	7.5	6.5	5.5	33
13	6	5.5	5	5	4	25.5
15	5	6	5	5.5	4.5	26
17	4	4	3.5	4	3	18.5
19	6.5	6	5.5	6.5	5	29.5
21	7	6	5	5.5	5	28.5
23	6	6	6	6	6.5	30.5
25	8	7	7.5	8	6	36.5
27	6	6	5.5	5	5	27.5
29	7	6	6	6.5	6.5	32
Mean	5.9	5.4	5.2	5.5	4.7	26.8

The results of the post-treatment test of the control group showed that individuals improved a little. They lacked multiple things, such as coherence in their thoughts and words. In two of the questions where participants were asked to describe something, they did not use descriptive language, and that is why they could not score higher in those questions, such as question no. 3 and 4. The control group only improved a little after five sessions since the tutor needed to engage more students in the discussion so they could improve their English speaking. Thus, some of the participants scored higher than others.

On the other hand, the experimental group also had five sessions/classes of sixty minutes each and took the pre-treatment test. They were asked the same questions which were asked of the control group. The results of the pre-treatment tests of both groups were almost identical; thus, their post-treatment tests' results did show the efficacy of the treatment.

Table 3 shows that the experimental group participants were asked five questions, and each question's marks were assigned separately.

Table 3 Pre-treatment test of the experimental group

Participants	Question# I	Question#2	Question# 3	Question# 4	Question# 5	Obtained Total
02	5	4.5	3.5	3.5	3	19.5
04	6	4	6	6	5	27
06	5.5	5	4	3.5	4	22
08	3	3	2.5	4	3	15.5
10	5	5	5	4.5	4	23.5
12	4	4	3.5	4	2	17.5
14	6	5	4	4.5	3.5	23
16	5	4.5	3.5	3	3	19
18	4.5	4.5	4	3.5	3	19.5
20	5	5.5	5	4.5	4.5	24.5
22	4.5	4	3.5	3	2	17
24	6	5.5	5	4.5	4	25
26	6.5	6	6	6.5	6.5	31.5
28	4.5	4	4	3.5	3	19
30	6.5	5.5	5	5	5	27
Mean	**5.1**	**4.6**	**4.3**	**4.2**	**3.7**	**22**

Same as the control group, the experimental group participants did not score high, showing that both groups were at the same level. Thus, this study's ultimate goal was achieved when the results of the pre-treatment test of both groups were almost the same. Therefore, the treatment became crucial. The experimental group also had five sessions of sixty minutes, but instead of relying on the teacher, every participant interacted with the chatbot. The chatbot answered participants' questions, and then it gave some exercises. Meanwhile, every participant in the experimental group spoke simultaneously, unlike the other group.

Table 4 Results of the post-treatment test of the experimental group

Participants	Question# I	Question#2	Question# 3	Question# 4	Question# 5	Total Marks
02	7	6	5	5.5	5.5	29
04	5	5.5	6.5	6.5	5.5	29
06	7	6.5	5.5	5.5	6	30.5
08	5.5	5	5	5.5	5	26
10	7.5	7	6.5	6.5	6	33.5
12	5.5	5	5.5	5.5	5	26
14	7.5	6.5	5.5	6	6.5	32
16	6.5	5.5	5.5	4.5	4.5	26.5
18	6	6	6	6	5	29
20	8	7.5	7.5	8	7	38
22	6	5.5	5.5	5	4.5	26.5
24	7	6.5	6.5	6	5.5	31.5
26	9	8.5	8.5	9	8	43
28	6	5	6	5	5	27
30	8	7.5	6.5	7	6.5	35.5
Mean	**6.7**	**6.2**	**6.1**	**6.1**	**5.7**	**30.8**

The results clearly showed that experimental group did better as compared to the control group. Since every participant had enough time to interact with the chatbot and all participants conversed for sixty minutes, the experimental group's collective score was higher than 60%. It was noticed that participants of the experimental group were more fluent in language than the other group, and they shaped their thoughts into words intelligently. Moreover, it was also observed that the experimental group learnt more synonyms, and their pronunciation improved. The participants of the experimental group talked to the chatbot as they would do to a native speaker; thus, their speech was more eloquent as well.

The results of the tests clearly showed that learning English speaking with NUMLINA chatbot was more successful than the traditional method. Moreover, it also showed that language learners learnt many things in just five sessions, which is a short time for honing their speaking skills. The research also evaluated more details about participants' speaking while answering the questions enhances the significance. Table 5 shows how marks were assigned to each participant's fluency, pronunciation, vocabulary, ability to communicate in the English language eloquently and use of idioms and phrasal verbs:

Table 5 Marks categories

Skills	Marks/Points
Fluency	2
Vocabulary	2
Pronunciation	2
Use of idioms & phrasal verbs	2
Ability to communicate in English Eloquently	2
Total Marks	10

Table 6 Results of the control group

Participants	Fluency	Vocabulary	Pronunciation	Use of idioms and phrasal verbs	Ability to Communicate in English Eloquently	Total
01	2	1.5	1.5	0	1	6
03	1	2	2	1.5	0	6.5
05	1	1	1.5	0	1	4.5
07	1	2	1	1	1	6
09	2	2	2	0	1	7
11	1	0	1	0	0	2
13	1.5	1	2	1	1	6.5
15	0.5	1	0.5	0	2	4
17	2	0.5	1	0.5	1	5
19	1	1.5	1	0	1.5	5
21	1	1	1.5	0.5	1	6
23	2	1	1	1	2	7
25	1	0.5	1	1	1.5	5
27	2	0	1	0	1.5	4.5
29	2	1	2	1	2	8
Mean	**1.4**	**1**	**1.3**	**0.5**	**1.1**	**5.5**

In table 5, the results showed that participants lacked fluency, vocabulary, and pronunciation and did not use phrasal verbs or idioms in their speech.

Table 7 Results of the post-treatment test of the control group

Participants	Fluency	Vocabulary	Pronunciation	Use of Idioms and phrasal verbs	Ability to communicate in English Eloquently	Total
01	2	1	1.5	1	1	6.5
03	1	2	2	2	1	8
05	1	1	1.5	1	1	5.5
07	1	2	1.5	1	1	6.5
09	2	2	2	1	1	8
11	1	0.5	1	0.5	1	4
13	1.5	1	2	1	1.5	7
15	0.5	1	0.5	0.5	1	3.5
17	2	1	1	1	1.5	6.5
19	1	1.5	1	1	2	6.5
21	1	1	1.5	1	1.5	6
23	2	1	1	1	2	7
25	1	1	1	1	1.5	5.5
27	2	0.5	1	1	1.5	6
29	2	1	2	1.5	2	8.5
Mean	**1.4**	**1.1**	**1.3**	**1**	**1.3**	**6.3**

Even after the treatment phase, the control group did not improve enough, and the results remained almost the same. It showed that the traditional method did not even improve those minor skills in the participants in that short period.

Table 8 Results of the pre-treatment test of the experimental group

Participants	Fluency	Vocabulary	Pronunciation	Use of idioms and phrasal verbs	Ability to Communicate in English Eloquently	Total
02	1	1.5	2	0	1	5.5
04	2	1	1	0	0	4
06	2	1.5	2	0	2	7.5
08	1	1.5	1.5	1	1	6
10	1.5	1	2	0	2	6.5
12	2	0	2	0	2	6
14	1.5	1.5	1	0.5	1	5.5
16	1	1	1	1	1.5	5.5
18	1	1	0.5	0.5	1	4
20	0.5	0	1	0	0.5	2
22	0.5	0	0.5	0	1	2
24	1.5	2	2	1.5	2	9
26	2	1	1.5	0	1	5.5
28	1	1	0.5	0	1	3.5
30	1	0.5	1	0.5	1	4
Mean	**1.3**	**0.9**	**1.3**	**0.3**	**1.2**	**5.1**

The results of the pre-treatment test of the experimental group are almost identical to the control group's results. It helped prove that participants improved significantly compared to the control group's since the results were significantly better than before.

Table 9 Results of the post-treatment test of the experimental group

Participants	Fluency	Vocabulary	Pronunciation	Use of idioms and phrasal verbs	Ability to communicate in English Eloquently	Total
02	2	2	2	1.5	1.5	9
04	2	2	2	1.5	2	9.5
06	2	2	2	1.5	2	9.5
08	2	1.5	2	1.5	1.5	8.5
10	2	1.5	2	1.5	2	9
12	2	2	2	1.5	2	9.5
14	2	2	2	1.5	2	9.5
16	2	2	1.5	1.5	2	9
18	2	1.5	1.5	1	1.5	7.5
20	1.5	1	1.5	1	1	6
22	1.5	1.5	2	1	1.5	7.5
24	2	2	2	2	2	10
26	2	2	2	1.5	1.5	9
28	2	2	1.5	1.5	2	9
30	1.5	2	1.5	1.5	1.5	8
Mean	**1.9**	**1.8**	**1.8**	**1.4**	**1.7**	**8.7**

The results of the post-treatment test of the experimental group showed a significant difference. It showed how participants improved their English speaking quite drastically. The experimental group participants had more time to practice speaking and held conversations with a chatbot as they conversed with a native speaker. The experiment also proved that a chatbot could substitute for a native speaker, and the language learners can practice their English speaking as much as they want.

However, the data shown in the research shows that the traditional method needs to be more effective in teaching individuals English speaking while the chatbot has done a great job. As other research articles have shown the importance of chatbots for language learning, this study has also helped to determine that a chatbot can act as a language tutor. Furthermore, the study has also been successful in proving that there are downsides to the traditional method of teaching the English language (Akram & Abdelrady, 2023). For instance, in a conventional language learning class, a tutor cannot engage everyone in the class at the same time. It also shows that some students learn better than their other class fellows, but the chatbot minimises that disadvantage.

While conducting research, it was also noticed that participants who interacted with the chatbot were speaking and interacting unhesitatingly as compared to the others. For instance, if a participant needed guidance about a word, they would ask the chatbot the meaning multiple times. Unlike a human teacher, the chatbot would answer without getting tired of the question. Since English language learning requires so much practice, it must be remembered that chatbots facilitate learners to perfect their

pronunciation of difficult words by repeating them more. A human teacher cannot give their full attention to everyone in the class, but a chatbot can be used by everyone in the class simultaneously. P. K, Too, & Mukwa (2018) have shown that most teachers have positive attitudes towards chatbots in education. Therefore, TEFL teachers can also use chatbots to teach their students. Moreover, they can rely on such chatbots to do multiple tasks. For instance, they can ask students to interact with the chatbot and learn new words and their pronunciation.

Conclusion

The study has explored the right way to teach English language speaking; it found data about Twitter Chatbots and how teachers have used them previously for language learning. The study found positive results regarding the behaviour of teachers towards the use of chatbots for language learning and other subjects as well in the study (P. K et al., 2018, p. 9). Some institutes are promoting the inculcation of chatbots for learning language skills. Ruan, Willis, Xu, Davis, Jiang, Brunskill, & Landay (2019) provided evidence about the application of chatbots in language learning.

The hypothesis was valid, and the experiment conducted by the research team proved that NUMLINA chatbot (Ullah, 2023) was better than the human teacher since it enabled participants to speak more fluently and made them eloquent in their discourse. Furthermore, the study concluded that chatbots could teach more things than a human tutor in a brief time. However, the study did not explain the training process of the chatbot since it was not the purpose of the study. The chatbot training played an important role since the chatbot was trained in this manner that it could teach multiple things to the participants. Furthermore, the chatbot was capable of answering closed-ended questions. It could not answer like any other human teacher, but it provided enough knowledge to the participants as any other human teacher. There was ease and comfort in using the NUMLINA chatbot, and it did not require any special knowledge or skill to interact with it; thus, it was considered efficient enough to enable students to speak. The participants had fun while learning new things, and it was their first experience where they learnt something new with the chatbot. Therefore, the participants' response towards the chatbot was quite positive, and they welcomed the idea of learning more with a chatbot in future. In conclusion, the study has opened new ways for future research works. Future research projects can explore more details in this field; for instance, they can build chatbots for specific purposes and include more participants to strengthen the claim. Moreover, future research projects can explore the chatbot's capabilities more minutely, and work on an AI chatbot that can revolutionise English language learning.

References

Abdul-Kader, S. A., & Woods, J. C. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications, 6*(7). doi:10.14569/ijacsa.2015.060712

Abdelrady, A. H., & Akram, H. (2022). An empirical study of ClassPoint tool application in enhancing EFL students' online learning satisfaction. *Systems, 10*(5), 154.

Akram, H., Yingxiu, Y., Al-Adwan, A. S., & Alkhalifah, A. (2021). Technology integration in higher education during COVID-19: An assessment of online teaching competencies through technological pedagogical content knowledge model. *Frontiers in psychology, 12*, 736522.

Akram, H., & Abdelrady, A. H. (2023). Application of ClassPoint tool in reducing EFL learners' test

anxiety: an empirical evidence from Saudi Arabia. *Journal of Computers in Education*, 1-19.

Asbjørn, F., & Bae, B. P. (2017). Chatbots and the new world of HCI. Interactions, 24(4), 38-42.

Bhutto, J and Ramzan. M. (2021). "ENGLISH: Verses of Quran, Gender Issues, Feminine Injustice, and Media Transmission - CDA of Pakistani Press Reports". *Rahatulquloob* 5 (2), 111-26.

Bii, P. K., Too, J. K., & Mukwa, C. W. (2018). Teacher Attitude towards Use of Chatbots in Routine Teaching. Universal Journal of Educational Research, 6(7), 1586-1597. doi:10.13189/ujer.2018.

Bibauw, S., Van den Noortgate, W., François, T., & Desmet, P. (2022). Dialogue systems for language learning: a meta-analysis. Language Learning & Technology, 26(1). 1–24. https://hdl.handle.net/10125/73488

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. Language Learning & Technology, 27(2). 6–27. https://hdl.handle.net/10125/73501

Google. (2020, June 18). Dialogflow. Google Cloud. Retrieved June 18, 2020, from https://cloud.google.com/dialogflow/docs

Hattem, D., & Lomicka, L. (2016). What the Tweets say: A critical analysis of Twitter research in language learning from 2009 to 2016. E-learning and Digital Media, 13(1-2), 5-23.

Huang, J. X., Lee, K. S., Kwon, O. W., & Kim, Y. K. (2017). A chatbot for a dialogue-based second language learning system. CALL in a climate of change: adapting to turbulent global conditions–short papers from EUROCALL, 151-156. doi:10.14705/rpnet.2017.eurocall2017.705

Kent, D. (2021). Voice-user interfaces for TESOL: Potential and receptiveness among native and non-native English speaking instructors. Language Learning & Technology, 25(3), 27-39. http://hdl.handle.net/10125/73444

Kerly, A., Hall, P., & Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. Knowledge-based systems, 20(2), 177-185. doi:10.1016/j.knosys.2006.11.014

Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Pradeepkumar, B., & Das, T. (2015). A study of today's AI through chatbots and rediscovery of machine intelligence. International Journal of u-and e-Service, Science and Technology, 8(7), 277-284. doi:10.14257/ijunesst.2015.8.7.28

Raj, S., & Raj, S. (2019). Building Chatbots the Hard Way. Building Chatbots with Python: Using Natural Language Processing and Machine Learning, 105-154. doi:10.1007/978-1-4842-4096-0_4060719

Ramzan, M., Bibi, R., & Khunsa, N. (2023). Unravelling the Link between Social Media Usage and Academic Achievement among ESL Learners: A Quantitative Analysis. *Global Educational Studies Review, VIII*(II), 407-421. https://doi.org/10.31703/gesr.2023(VIII-II).37

Ramzan, M., Azmat, Z., Khan. M.A. & Nisa, Z. (2023) Subject-Verb Agreement Errors in ESL

Students' Academic Writing: A Surface Taxonomy Approach, MARS Publishers, Linguistic Forum, Volume 5, Issue 2, 2023 Pages 16-21

Ramzan, M., Mushtaq, A. Ashraf, Z. (2023). Evacuation of Difficulties and Challenges for Academic Writing in ESL Learning. University of Chitral Journal of Linguistics and Literature Volume 7, Issue 1, Pages 42-49.

Ramzan, M., Oteir, I., Khan, M. A., Al-Otaibi, A., & Malik, S. (2023). English learning motivation of ESL learners from ethnic, gender, and cultural perspectives in sustainable development goals. *International Journal of English Language and Literature Studies*, *12*(3), 195-212.

Roose, K. (2022). A coming-out party for generative AI, Silicon Valley's new craze. The New York Times. https://www.nytimes.com/2022/10/21/technology/generative-ai.html

Ruan, S., Willis, A., Xu, Q., Davis, G. M., Jiang, L., Brunskill, E., & Landay, J. A. (2019, June). Bookbuddy: Turning digital materials into interactive foreign language lessons through a voice chatbot. In Proceedings of the sixth (2019) ACM conference on learning@ scale (pp. 1-4). Hattem, D., & Lomicka, L. (2016). doi:10.1145/3330430.3333643

Ullah, Z. (2023, March 26). Scholarly work 05. Building a chatbot NUMLINA for teaching speaking skill [Video]. YouTube. https://www.youtube.com/watch?v=c-bogB27rcc

Wang, Y. F., & Petrina, S. (2013). Using learning analytics to understand the design of an intelligent language tutor–Chatbot lucy. Editorial Preface, 4(11), 124-131. doi:10.14569/ijacsa.2013.041117

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017, May). A new chatbot for customer service on social media. In Proceedings of the 2017 CHI conference on human factors in computing systems (pp. 3506-3510). doi:10.1145/ 3025453.3025496

Appendix

Pre-test

Time: 50 minutes **Total Marks: 60**

Each question contains equal marks. All of the questions must be asked to each participant.

1. What is your name; tell me a bit about yourself like where are you from and what do your parents do? (10)
2. Are you satisfied with this university, what's your experience so far? (10)
3. Who is your favourite personality; tell me about them in detail like why you do like them and what do you like about them the most? (10)
4. Do you watch movies or are you more of a reader? (10)
   - 4.1. Which is your favourite movie? Tell me its plot.
   - 4.2. Which book is your favourite? Tell me about it.

5. What do you think of feminism, tell me about your thoughts on that? (10)
6. Analysis of fluency, pronunciation, vocabulary, use of idioms and ability to speak English eloquently. (10)

Post-test

Time: 50 minutes Total Marks: 60

Each question contains equal marks. All of the questions must be asked to each participant.

1. Tell me about yourself in detail and tell me who you really are and what do you want to be in life? (10)
2. How did you end up in this university, tell me about your journey as a student like how you have planned to study in this particular course and how is your experience so far? (10)
3. Who do you see as an inspiration, do you look up to someone as a perfect example? Elaborate their personality. (10)
4. How do you spend your time when you have completed your tasks, how do you make yourself productive and how do you entertain yourself? (10)
5. Since you are a university student, and everyone expects you to be a well-read and well-educated person, tell me what is your stance on feminism as a modern movement? (10)
6. Analysis of fluency, pronunciation, vocabulary, use of idioms and ability to speak English eloquently. (10)