

# Evaluating the Effectiveness of AI-powered Chatbot in Enhancing Speaking Skills: An Experimental Study on Bangladeshi Undergraduate EFL Learners

Tabassum Mahmud Orna<sup>1</sup>

Md. Mazharul Islam<sup>2</sup>

Ashiful Islam<sup>3</sup>

## Abstract

Integrating artificial intelligence (AI) in language learning offers new ways to enhance speaking skills among English as a Foreign Language (EFL) learners. Particularly those facing challenges in traditional classroom environments. This study investigates the effectiveness of an AI-powered chatbot, ChatGPT, in improving speaking proficiency among Bangladeshi undergraduate EFL learners. The study employs an experimental pretest-posttest design. Fifty participants took part for four weeks, they engaged in conversational practice using ChatGPT. The practice mainly focused on four aspects, including fluency, accuracy, pronunciation, and confidence. Data were gathered through assessments conducted at the beginning and end of the study. Participants shared qualitative responses that illustrated their difficulties and views regarding the intervention. Based on the Interaction Hypothesis and Constructivist Learning Theory, the study discovered significant improvements in all measured areas. Challenges, including accent recognition limitations, highlighted the need for AI advancement in non-native language contexts. These findings suggest that AI-powered chatbots play an important role in developing spoken language skills. They also reveal useful information that can be beneficial for EFL teachers and AI developers.

**Keywords:** AI in language education, ChatGPT, EFL speaking skills, Bangladeshi undergraduate learners, language learning technology

Over recent years, Artificial Intelligence (AI) has brought a transformation to language learning through specialized and interactive educational tools for learners. Mastering spoken English is a challenge for a learner because of the educational settings in Bangladesh. Most of the time, learners lack the essential opportunity to maintain actual dialogue interactions. That is why their speaking fluency and self-confidence struggle to develop.

<sup>1</sup>Lecturer, Department of English, City University, Bangladesh

<sup>2</sup>Graduate Student, Department of English, Jahangirnagar University

<sup>3</sup>Graduate Student, Department of English, Jahangirnagar University

ChatGPT represents an effective solution to these issues. The system runs Natural Language Processing (NLP) algorithms to produce human-like dialogue. The environment enables EFL students to develop their speech under low-stress conditions. AI chatbots deliver quick feedback that checks both the grammatical structure of students' messages and their pronunciation, along with word choices. Essential precepts of language learning find strong support through this form of assistance. According to Long (1981), the interaction hypothesis stands as one important concept that suggests that the process of conversation enables better acquisition of knowledge. Chatbots allow learners to practice speaking by adjusting their speaking speed while practicing in various simulated conditions. According to hypotheses within the Constructivist Learning Theory by Piaget (1972) and Vygotsky (1978), learning becomes most effective through experiences that resemble real-life contexts and demonstrate practical value.

Language researchers have reviewed how well AI chatbots perform in teaching languages through existing investigations. However, they have given limited attention to studying the effectiveness of these methods in improving speaking abilities among undergraduate EFL learners in Bangladesh. This research seeks to address this missing piece of information. Bangladeshi undergraduate students' fluency, accuracy, pronunciation, and confidence in speaking skills are examined following regular communication with ChatGPT. The research evaluates both learner feedback regarding the AI chatbots and their distinct barriers to success. Educators and AI developers can attain beneficial lessons through these findings to enhance their chatbot-based language teaching technology.

### **Research Questions**

1. How does ChatGPT enhance speaking skills for Bangladeshi EFL learners?
2. How do learners perceive ChatGPT as a tool for speaking practice?
3. What are the benefits and challenges of using ChatGPT for speaking practice?
4. How does ChatGPT's feedback affect learners' self-correction and language use?

### **Literature Review**

AI in language learning has transformed traditional methods by making learning more personalized and flexible. Chatbots like ChatGPT use advanced technology to simulate real conversations. They give instant feedback and promote independent practice. This is especially helpful for EFL learners in Bangladesh, who often lack access to fluent English speakers. By offering conversational practice and instant feedback, AI chatbots support essential parts of language learning, like

regular practice, feedback, and learner autonomy (Zhai et al., 2021).

Studies show that AI chatbots help improve language skills. Fryer and Carpenter (2006) found that chatbots create a friendly and low-pressure space for speaking practice. They help learners boost fluency, accuracy, and pronunciation by giving real-time input and feedback (Xiao et al., 2023). In countries like Bangladesh, these tools are especially good for building fluency and confidence.

AI chatbots align with the Interaction Hypothesis, which says that language learning gets better through feedback. Long's idea (1981) suggests that talking with others and getting feedback helps improve language skills. ChatGPT supports this by letting learners adjust their speech based on instant corrections, which boosts self-correction. Also, the Constructivist Learning Theory highlights active learning, which ChatGPT promotes by offering real-life conversation practice.

Research shows the positive impact of AI chatbots on speaking skills. Mohamed (2024) found that using ChatGPT often improved learners' fluency and pronunciation. Wang et al. (2022) saw that chatbot interactions increased motivation and reduced anxiety, allowing EFL learners to practice more comfortably. Xiao et al. (2023) reported that EFL learners gained confidence and fluency from instant feedback on grammar and pronunciation.

With that being said, AI chatbots do have quite a lot of benefits for EFL education, while they also come with some challenges. There is one big problem that is people fail to recognize different accents. Thus, it is doubly difficult for Bangladeshi learners who have regional accents. According to Huang et al. (2022), talking about an accent difference can affect the accuracy of the feedback learners receive and can be very frustrating for them. Using chatbots is even harder when there is limited access to technology and the internet in a low-resource area like Bangladesh. This is where we need better AI technology and better infrastructure.

Learners' viewpoint of AI chatbots also plays a key role in determining the success of such bots. According to Wang and another authors (2022), learners who think of chatbots helping engage more, practising more and it translates in better progress. It alleviates speaking anxiety and increases motivation, as ChatGPT's friendly and nonjudgmental environment (Xiao et al., 2023). The focus of this study is to analyze whether Bangladeshi EFL students would rather use chatbots to exchange English because they will be less afraid of making mistakes in their efforts. Bangladeshi learners may find using chatbots helpful for practicing speaking, as they offer a high level of flexibility and independence.

In general, the research demonstrates that AI chatbots can facilitate language learning. Nevertheless, there are scarce studies of Bangladeshi EFL learners. This review shows us that AI chatbots can assist in developing speaking skills where there are few practice opportunities. In order to have better results, it has to address some issues such as accent recognition, cultural understanding, and access to technology. This study aims to investigate how ChatGPT may play a role in helping Bangladeshi undergraduate EFL learners enhance their speaking skills and their views on such a chatbot in the future use of chatbots in EFL education.

### **Theoretical Framework**

In this study, Interaction Hypothesis and Constructivist Learning Theory are applied to explore how ChatGPT assists Bangladeshi EFL learners to enhance their speaking skills. In the case of the Interaction Hypothesis (Long), language skills increase through interaction and feedback. This is what ChatGPT does by providing real-time feedback and enabling learners to work on fluency, accuracy, and pronunciation. Piaget and Vygotsky are two Constructivist Learning Theorists who teach that learning is better through real-life experiences. Instead of trying to memorize and remember language vocabulary without context, ChatGPT provides realistic language practice through which learners can discuss everyday topics, give and hear opinions, and this is more likely to make them remember and learn what they are learning. Theories about automaticity and surface repetition, which give both feedback and realistic practice, are supported by ChatGPT and cannot be done without native speakers.

### **Methodology**

The effectiveness of using ChatGPT to enhance the speaking skills of Bangladeshi undergraduate EFL learners is evaluated with the help of a pretest and posttest design. By blending numbers-based scoring of speaking skills with participant comments, ChatGPT essentially shows you the impact it has on fluency, accuracy, pronunciation, and confidence of your speaking in English.

### **Research Design**

The study is conducted following the single-group pretest and posttest approach. What this means is that the speaking skills of the participants are assessed before and after four weeks of daily ChatGPT usage. With this setup, it is also straightforward to compare their skills over time and observe ChatGPT's impact on improvement. The study uses structured pre and post assessments as data gathering means for quantitative (number-based) data, and collects participant

feedback to get personal findings on ChatGPT's experience in terms of speaking practice for humans.

### Participants

The study involves 50 undergraduate EFL learners from a university in Bangladesh, aged 21-25. To maintain consistency, participants had no prior experience with AI-powered chatbots for learning English speaking. All participants provided informed consent, and their anonymity was protected throughout the study.

### Procedure

There were three phases in this research: pretest, intervention, and posttest. First, participants were given an initial speaking test. The four skills measured in this test were fluency, accuracy, pronunciation, and confidence. To introduce themselves, describe what they do daily, their opinions on current events, as well as what they wish to accomplish in the future. Participants, however, rated their performer's performance on a 5-point scale (1 = very poor, 5 = excellent). It became a starting point for comparing these scores.

Participants trained regularly for the next four weeks with ChatGPT, speaking with it as much as possible. Participants were provided with real-time feedback on the grammar, the vocabulary, and the pronunciation, and were asked to correct mistakes from ChatGPT. The topics varied from those of the pretest for consistency, but drew from the most common topics used in previous performance measuring studies.

The tasks and scoring of the post test were exactly the same as they were in the pre test. As such, it was easily to compare scores before and after the four weeks. They also gave feedback on how easy ChatGPT is to use, and provided feedback on whether they found it effective and if there were challenges they faced.

### Data Analysis

Data were analyzed using R 3.6.0+. The analysis involved the following steps:

#### Quantitative Analysis:

**Paired samples t-test:** This statistical test was applied to compare the pretest and posttest scores across the four key dimensions (fluency, accuracy, pronunciation, and confidence). The test determined whether the improvements observed in

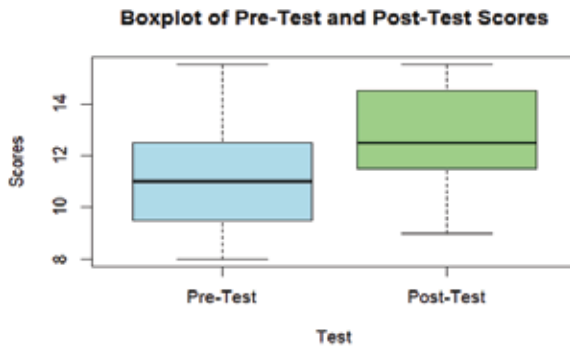
participants' speaking proficiency were statistically significant.

**Qualitative analysis:**

**Thematic Analysis:** Responses from the qualitative feedback survey were analyzed using thematic analysis. Common themes were identified to highlight learners' perceptions, including any challenges they faced while using ChatGPT and the perceived benefits of the intervention.

**Findings and Discussions**

Analysis of Pre-Test and Post-Test Scores using R 3.6.0+



This section summarizes the findings of the statistical analyses used to assess the intervention's impact on students' speaking efficiency, as evaluated by pre- and post-test scores. The Shapiro-Wilk test was performed to determine if variations between pre-test and post-test scores were normally distributed. It was found that the differences are not regularly distributed, prompting us to conduct a t-test.

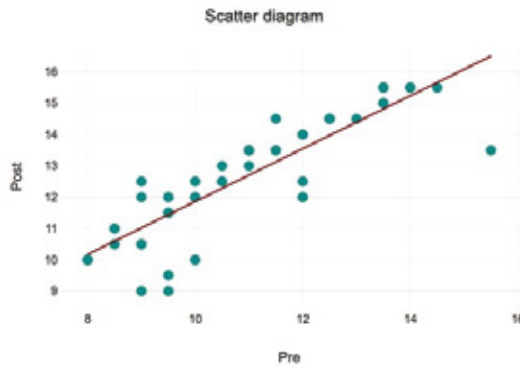
The box plots reveal significant differences in pre- and post-test results. The spread (interquartile range) is slightly greater in the post-test, indicating a broader range of results. This indicates prospective improvement following the intervention.

A paired-sample t-test was performed manually to confirm the experimental group's pre-test and post-test scores.

Experiment	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>
Pre-test	10.97	1.867	4.563	98	0.05	1.20
Post-test	12.68	1.841				

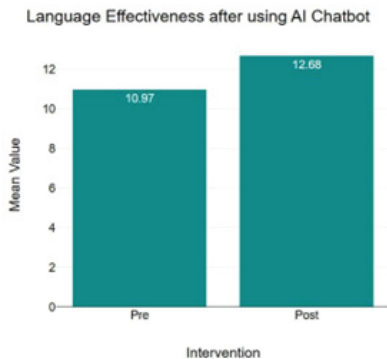
The mean pre-test score for the experimental group was 10.97, based on a sample size of 50 participants. In comparison, the post-test results had an average of 12.68.

The pre-test score has a standard deviation of 1.867, whereas the post-test score has a standard deviation of 1.58. The effect size was calculated using Cohen's  $d$ , which came out to be 1.20, suggesting a large influence. Finally, the  $t$ -value obtained from R is 4.563. After comparing the results to the critical  $t$  value, we can conclude that the result is significant enough to reject the null hypothesis and that the intervention was advantageous.



The scatter plot shows the correlation between pre-test scores (x axis) and post-test scores (y axis). Each point represents an individual score comparing pre- and post-test performance.

The line of best fit shows an increasing trend, demonstrating a positive correlation: as pre-test scores climb, so does post-test performance. This shows that students who scored higher on the pre-test also scored higher on the post-test, indicating consistency or improvement in performance after an intervention or over time. The points are spread out around the trend line, demonstrating some variation but an overall strong, positive relationship between pre-test and post-test outcomes.



This bar chart presents the mean values of "Language Effectiveness" before (Pre) and after (Post) an intervention using an AI chatbot. The y-axis shows the mean value, and the x-axis denotes the intervention stages, with "Pre" indicating the value before the chatbot was used and "Post" indicating the value after.

The typical value during the "Pre" stage is 10.97. After incorporating the AI chatbot, the "Post" stage mean value rises to 12.68. This implies that the AI chatbot intervention had a good impact, as the mean value increased from pre- to post-intervention.

### **Qualitative analysis**

This study used a pre-test and post-test design with 50 Bangladeshi undergraduate EFL learners to evaluate their speaking skills in four main areas: fluency, accuracy, pronunciation, and confidence. The speaking tasks included topics such as introducing themselves, describing daily routines, discussing current events, and expressing future goals.

### **Pre-Test Phase**

The participants demonstrated significant challenges in all four speaking skill domains, including fluency and accuracy, together with pronunciation and confidence during the pre-test phase. The participants experienced difficulty with fluent speech as they paused between sentences while translating Bengali into English. During the pre-test phase, Monsur demonstrated multiple pauses when he said "My name... is Monsur... and I... live in..." His limited ability to speak English spontaneously became evident when he said "My name... is Monsur... and I... live in... Dhaka." Students struggled to maintain proper sentence flow while making numerous grammatical errors during their speech. The most frequent errors among students involved tense inconsistencies together with incorrect pronoun selection and improper preposition usage. During the interview Amanullah used the pronoun "he" to describe his mother before correcting himself to "she." The participant's use of inconsistent verb tense created confusion when he said "I go to market yesterday" instead of "I went to the market yesterday." Several participants chose inappropriate words in their speech which reduced both clarity and meaning understanding such as when Sadiya used "good" instead of "important" in her statement "This is a good issue for me." The participants' pronunciation problems stemmed mainly from their local speech patterns. The participants from Comilla and Chittagong regions pronounced words differently because they said "tink" instead of "think" and "dat" instead of "that." The incorrect pronunciation of words created two problems: it made it harder for listeners to understand and it seemed to

make students less willing to speak. Confidence levels were generally low. The participants displayed excessive worry about making errors that led to response delays and feelings of nervousness. Sumaiya demonstrated her fear through frequent interruptions during speech as she stated "I... I am... sorry, I don't know... maybe I should... start again." The participant expressed self-doubt through the statement "I am... sorry, I don't know... maybe I should... start again" because of limited English practice and fear of mistakes.

### **Post Test Phase**

Instructions through the use of ChatGPT led to major enhancements in the proficient understanding of all necessary skills among participants. The participants achieved significant improvements in their speech fluency because they started delivering their words more effortlessly. The participants experienced fewer delays due to their routine practice of familiar content. During the post-test interview, Monsur spoke fluently without pauses to deliver his statement, which began with "My name is Monsur, I reside in Savar, Dhaka since I lost my parents." Accuracy also improved. The participants' attention to grammatical structure and vocabulary use increased through immediate feedback provided by ChatGPT. Amanullah correctly used pronouns after overcoming his previous difficulties because his mother supported him both academically and emotionally through his education. Students moved from saying "I go to market yesterday" toward correct verb tense usage by stating "I went to the market yesterday" while maintaining the past tense. Student writing showed better coherence along with enhanced clarity because learners started using the word "important" instead of "good" to denote meaning importance as in the statement "This is an important issue for me."

Studying pronunciation produced substantial progress by concentrating on proper articulation without focusing on accent, so regional dialects would not affect the results. The post-tests revealed that Comilla and Chittagong participants who previously said "think" as "tink" and "that" as "dat" achieved proper pronunciation of these words. The participants achieved a better understanding because their enhanced pronunciation made their speech easier to understand. Confidence levels improved markedly. ChatGPT offered participants judgment-free feedback that allowed them to talk without hesitation and free of worries about errors. The individual who often delayed her speech shifted to more confident statements about English language progression through "I believe that learning English is a journey, and my skills are improving steadily." There was no hesitancy during her speech. The increased confidence among participants allowed them to communicate more naturally because of their regular practice of using ChatGPT with positive feedback.

### **Limitations**

The study establishes how ChatGPT assists Bangladeshi EFL learners in enhancing their speaking ability, yet it presents specific constraints. Because this research study did not include a control group, the investigators cannot demonstrate that results were obtained solely through ChatGPT. Some study participants encountered technical issues because they lacked access to suitable devices, and their internet connection experienced interruptions. The unstable digital resources throughout Bangladesh make the process of solving these problems more difficult. The spoken accents that ChatGPT failed to comprehend negatively affected both learner motivation levels and the quality of provided feedback. The brief four-week duration of the research proved insufficient for demonstrating how ChatGPT could effectively enhance speech practice because adequate improvement measurement was restricted.

### **Conclusion**

The research demonstrates that Bangladeshi EFL students can improve their speaking abilities with the use of ChatGPT technology. The participants reported they enhanced their speaking level while their comfort with conversational practice using ChatGPT improved. Student outcomes from both pre- and post-assessments appeared to validate growth rates because participants displayed rising confidence and falling anxiety regarding speech activities after leveraging the tool. The research findings indicate that ChatGPT effectively supports language learning because native speaker communication may become limited. The system needs better voice recognition technology, together with standardized technology access for its user population to provide full effectiveness. The resolution of current service challenges will enhance the efficiency of AI chatbots for educating all student populations. Studies reveal how AI chatbots present useful opportunities to support EFL students with their development of spoken fluency skills.

### **Recommendations**

Educational organizations, together with language teaching programs, need to integrate AI chatbots, specifically including ChatGPT, for their spoken language instruction. The tools enable students to conduct independent training between classes, which increases their language abilities. The next development of chatbots needs to strengthen their recognition of various accents so they become more effective tools for speakers. Combining chatbot activities with teacher-led lessons could also enhance learning. Including a control group in future studies would also strengthen the findings.

## References

- Akanda, F., Suchona, I. J., & Shorn, S. A. (2024). Evaluating ChatGPT's influence on English language teaching and learning in Bangladesh: *Insights from teachers and students*. SSRN. <https://doi.org/10.2139/ssrn.5006965>
- Aysu, S. (2024). The use of AI-powered tools in language learning. *Dil Eğitimi ve Öğretiminde*, 1.
- Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. In *Proceedings of the CALL Conference*.
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning – Are they really useful? A systematic review of chatbot supported language learning. *Journal of Computer Assisted Learning*, 38(1), 237–257. <https://doi.org/10.1111/j-cal.12614>
- Jinming, D. U., & Daniel, B. K. (2024). A systematic review of AI-powered chatbots in EFL speaking practice: Transforming language education. *Computers and Education: Artificial Intelligence*, 5, 100230. <https://doi.org/10.1016/j.caeai.2024.100230>
- Lin, C. C., Huang, A. Y., & Yang, S. J. (2023). A review of AI-driven conversational chatbots implementation methodologies and challenges (1999–2022). *Sustainability*, 15(5), 4012. <https://doi.org/10.3390/su15054012>
- Mohamed, A. M. (2024). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies*, 29, 3195–3217. <https://doi.org/10.1007/s10639-023-11917-z>
- Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, 37(4), 814–840. <https://doi.org/10.1080/09588221.2022.2056203>
- Xiao, F., Zhao, P., Sha, H., Yang, D., & Warschauer, M. (2023). Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning*. <https://doi.org/10.1515/jccall-2022-0032>
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 2021(1), 8812542. <https://doi.org/10.1155/2021/8812542>