Al-Itqan Journal of Humanities available online at https://iraqiqsa.org/Journal/index.php/AIJHS/index
ISSN: 3105-1073 (OnLine)
المجلة معتمدة بالترقيات العلمية حسب كتاب وزارة التعليم العراقية (46404-1410) في 2026/4/27

# The effect of AI-Chatbots on EFL Learners' Speaking Confidence and Accuracy

**Asst. Lect. Balsam Hilal Khalaf**
**Email: bakhalaf@uowasit.edu.iq**

# أثر روبوتات الدردشة بالذكاء الاصطناعي في ثقة ودقّة متعلّمي اللغة الإنجليزية في التحدّث

**م.م. بلسم هلال خلف**

**Abstract**

AI technologies used in the process of language learning create new routes to help increase learners' communicative competencies. AI chatbots can provide effective support in the language practices, especially when opportunities for interaction are limited. In this regard, the present study aims to investigate the effect of AI chatbots on speaking confidence and accuracy among EFL learners. The research focuses on the hitherto relatively little real-life speaking practice, combined with a low speaking confidence level among secondary school students learning English as a Foreign Language, which usually hinders their development of oral proficiency. The main aim of this research is to investigate whether regular interaction with AI-chatbots can facilitate learners' speaking confidence and linguistic accuracy. The current research adopted a pre-test/post-test design, with a standardized speaking test as the dominant tool of data gathering. Participants were 25 male students in Al-Najah Secondary School for Boys. In the six-week intervention, students rehearsed speaking regularly with AI-chatbots simulating conversational speech patterns. Results support that there is a significant increase in speaking confidence and accuracy. Students felt more confident to speak in English and were grammatically more fluent and accurate in the post-test than they had been in the pre-test. Results revealed that AI-chatbots could serve as an effective ancillary tool to acquire speaking proficiency in EFL environments, at least when exposure to native speakers or communicative practice is impossible. The present study contributes to the increasingly growing evidence base on pedagogical applications of AI for language teaching and offers insightful recommendations regarding how to embed chatbot-assisted speaking practice into EFL instruction.

**Keywords**: AI-chatbots, EFL learners, speaking confidence, speaking accuracy, language learning, secondary school, educational technology.

**الملخّص**

أدّى دمج تقنيات الذكاء الاصطناعي في تعليم اللغات إلى فتح آفاق جديدة لتطوير قدرات المتعلّمين التواصلية. ومن بين هذه التقنيات، أثبتت روبوتات الدردشة القائمة على الذكاء الاصطناعي فاعليتها في تسهيل ممارسة اللغة، ولا سيما في البيئات التي تتسم بندرة فرص التفاعل اللغوي الواقعي. تهدف هذه الدراسة إلى استقصاء أثر روبوتات الدردشة بالذكاء الاصطناعي في ثقة متعلّمي اللغة الإنجليزية كلغة أجنبية بدقّة التحدّث لديهم. يركّز البحث على مشكلة قلّة فرص الممارسة الواقعية للتحدث باللغة الإنجليزية وضعف الثقة بالنفس لدى طلبة المدارس الثانوية المتعلّمين للإنجليزية كلغة أجنبية، وهي مشكلة غالباً ما تعيق تطوّر كفاءتهم الشفوية. يتمثل الهدف الرئيس للدراسة في تحديد ما إذا كان التفاعل المتكرّر مع روبوتات الدردشة يسهم في تحسين ثقة المتعلّمين بأنفسهم أثناء التحدث ودقّتهم اللغوية. أُعتمد في الدراسة تصميم الاختبار القبلي–البَعدي باستخدام اختبار معياري للمهارات الشفوية كأداة رئيسة لجمع البيانات. وشارك في التجربة خمسةٌ وعشرون طالباً من الذكور في مدرسة النجاح الثانوية للبنين. وخلال فترة تدخّل استمرت ستة أسابيع، مارس الطلبة التحدث بانتظام مع روبوتات دردشة تحاكي أنماط المحادثة الطبيعية. أظهرت النتائج زيادة ملحوظة في مستوى الثقة والدقّة في التحدث. فقد أصبح الطلبة أكثر ثقة في استخدام اللغة الإنجليزية، وأكثر طلاقة وصحةً نحوية في الاختبار البَعدي مقارنة بالاختبار القبلي. وتُظهر النتائج أن روبوتات الدردشة بالذكاء الاصطناعي يمكن أن تكون أداة مساندة فعّالة لاكتساب مهارة التحدث في بيئات تعليم اللغة الإنجليزية كلغة أجنبية، ولا سيما حين تكون فرص التعرّض للناطقين الأصليين أو الممارسة التواصلية محدودة. تُسهم هذه الدراسة في تعزيز قاعدة الأدلة المتنامية حول التطبيقات التربوية للذكاء الاصطناعي في تعليم اللغات، وتقدّم توصيات قيّمة بشأن دمج ممارسات

التحدث المساعدة بروبوتات الدردشة ضمن تعليم اللغة الإنجليزية كلغة أجنبية.**الكلمات المفتاحية** :روبوتات الدردشة بالذكاء الاصطناعي؛ متعلّمو اللغة الإنجليزية كلغة أجنبية؛ الثقة في التحدث؛ دقّة التحدث؛ تعلّم اللغة؛ المرحلة الثانوية؛ التكنولوجيا التعليمية.

## 1. Introduction

Artificial Intelligence, being the fast-growing trend, has thus been influential in various sectors, including that of education. Within language acquisition, AI applications in the form of chatbots have increasingly been incorporated into pedagogy as support for students to achieve language competence (Huang et al., 2022; Lee, 2020). AI-designed chatbots engage students in the possibilities of working and playing with language in a lively conversation in low-stakes situations (Fryer & Carpenter, 2006). This is particularly relevant to EFL students who hardly have opportunities either inside or outside their classrooms for speaking practice. Acquisition of fluency, accuracy, and confidence in speaking constitutes one of the most complex and tricky aspects in second-language learning. Indeed, the classroom cannot offer the amount and the quality of speaking interactions required in achieving fluency and grammatical precision. It was in this context, considered here, that AI-chatbots emerge as a feasible complement to the traditional classroom teaching that would leave EFL students with a readily available, interactive, and custom-made platform for practicing speaking.Despite the increasing use of AI in education, there is a dearth of empirical research that establishes the particular impact of AI-chatbots on the speaking proficiency of EFL learners. Most of the research has so far focused on vocabulary development (Kukulska-Hulme, 2020), reading, and writing (Okonkwo & Ade-Ibijola, 2021), with relatively few study investigations into speaking performance. Furthermore, a great deal of the research has also tended to focus on issues of learner participation or motivation rather than on objective speaking competence and confidence building. Speaking accuracy—grammatical, pronunciation, and vocabulary accuracy—and speaking confidence—the learner's belief in being able to use the language to communicate effectively in English are two basic elements of oral proficiency that are crucial for effective communication (Woodrow, 2006). Addressing the knowledge deficiency on the impacts of AI-chatbots on these specific areas of speech capacity is thus of great importance.The research problem this study endeavors to address is the age-old issue EFL learners are faced with in building their speaking fluency and confidence due to limited practice time. Speaking practice in the majority of EFL classes, especially secondary levels, tends to be constrained by the number of the class, the limited available time, and the lack of native or proficient interlocutors (Al-Samarraie et al., 2020). Hence, learners can form anxiety towards language, low self-confidence, and lack of linguistic precision in spoken English. Coupled with this, the COVID-19 pandemic has also honed such concerns even more, given that online and blended learning environments reduce face-to-face practice possibilities (Yang et al., 2022). AI-chatbots in this instance could be a new and scalable solution, providing learners with multiple one-to-one practice opportunities to enhance confidence and precision.The primary objective of this study is to investigate the effect of AI-chatbots on the speaking accuracy and confidence of secondary school students. Specifically, this study seeks to determine if exposure to AI-chatbots can bring measurable improvement in these two skills among secondary school students. Furthermore, this study seeks to discover students' attitude towards speaking practice using AI-chatbots, supplementing the quantitative data with their own experiences. With these objectives in mind, the questions for the study are as follows:

1. How does speaking with AI chatbots affect the confidence of EFL learners in speaking?
2. In what ways is the speaking accuracy of EFL learners influenced by AI chatbots?
3. What are EFL learners' perceptions of the experience of practicing speaking with AI chatbots?

This research has implications for both theory and practice in the field of language education. Theoretically, it contributes to the growing knowledge base of AI-assisted language learning, particularly in speaking proficiency, where few studies have been done so far (Kukulska-Hulme, 2020; Lee, 2020). This study bridges the critical gap in the literature by focusing on speaking confidence and accuracy alone; therefore, it provides empirical evidence that will be a good guide for future investigation. The findings will help educators, curriculum designers, and policy framers at the practical level to understand the possible benefits and limitations with regard to the integration of AI-chatbots into EFL teaching. With increased access to AI technologies and ubiquitous concerns about speaking pedagogy, the contribution of this study will be valuable regarding how these technologies can be leveraged for improving learning outcomes.In particular, this is relevant to secondary education, where the critical stage of language learning happens, and school constraints most often complicate the provision of quality speaking classes. By targeting students in secondary schools, this research provides context-specific results which might be transferred to comparable learning environments. What really adds to the importance of the current study is that it deals with speaking confidence since the lack of confidence

normally serves as an obstacle for the use and development of language. Improved confidence as a result of low-anxiety interactions with AI-chatbots could thus yield long-lasting consequences with respect to making learners more willing to participate in spoken English inside and outside the classroom.The present research focuses its attention on a restricted population and a limited set of variables to study the research questions in detail.The participants in this study are 25 male students enrolled in Al-Najah Secondary School for Boys, an Iraqi secondary school in Kut, where English is taught as a foreign language. The duration of the treatment is six weeks. Participants engage in a regular speaking session with AI-chatbots designed to give students naturalistic conversation practice. This mixed-methods study relies on quantitative data from pre-test and post-test speaking measures, while qualitative data were collected from student feedback. Needless to say, generalizability to all EFL contexts is not attempted here; instead, this study attempts to represent closely how AI-chatbots may influence speaking confidence and accuracy within the confines of a secondary school.AI chatbots also have a significant role to play in addressing some of the perennial problems besetting EFL speaking class teaching. However, much research is needed to establish their efficacy and recommend their pedagogical use. Therefore, the present research tries to add to such insights for both academic research and instructional practices targeting secondary school learners in terms of speaking accuracy and confidence. Its findings may shape how languages will be taught in an AI-driven future.

## 2. Literature Review

### 2.1. Theoretical Foundations and ICALL

AI chatbots in language learning are part of the broader field of Intelligent Computer-Assisted Language Learning, or ICALL, a combination of computer-supported communication and instructional methods to facilitate SLA. According to Chapelle (2003), this kind of adaptive interactive software can be facilitative of communicative competence through the provision of authentic input and scaffolded practice.AI chatbots, being adaptive dialogue agents, extend this presence to building real-time conversational space beyond conventional class walls.

### 2. 2. Enhancing Speaking Fluency and Accuracy

#### 2.2.1 Fluency and Confidence

Empirical studies have shown that AI chatbots can substantially improve learners in terms of fluency and confidence. An experimental study using the Lyra app has been conducted, and it has been shown that experimental groups have higher levels of fluency, better lexical choices, and more pronunciation marks compared to controls. Mobile-based AI interaction has also been found to bring significant improvement in fluency and grammatical accuracy. Kim, Cha, and Kim (2021) also showed how content, pronunciation, intonation, stress, and improvement of speaking ability took place after long exposure to chatbots, and participants felt that their speech became easier and more natural.

#### 2.2.2. Accuracy and Feedback

Most of these chatbots are able to offer automated corrective feedback-a tool largely linked with accuracy gain. Fathi, Rahimi, & Derakhshan (2024) indicate that through interacting with chatbots, learners negotiate meaning, which encourages self-editing grammar and vocabulary.

### 2.3. Affective Domain: Anxiety, WTC, and Motivation

#### 2.3.1 Anxiety and Willingness to Communicate (WTC)

Foreign language anxiety consistently hampers spoken performance (MacIntyre, 2017). AI chatbots, providing non-judgmental conversational spaces, reduce the level of anxiety and build up WTC. El Shazly (2021) noted that AI-based speaking support lowered anxiety while improving IELTS speaking scores. Ebadi and Amini (2022) reported the significant effect of social presence and human likeness of chatbots on enhancing the motivation and WTC of Iranian EFL learners. Within a Chinese EFL study, Zhang, Meng & Ma (2024) proved the increase in enjoyment (FLE) and reduction in anxiety (FLA), with the rise of WTC after the interaction with an AI-speaking assistant.

#### 2.3.2. Motivation and Engagement

Motivation and engagement have been identified as indispensable in sustaining language learning. Indeed, the study mentioned above showed that AI chatbots promoted creative and engaging behavior in web-based EFL courses and attributed the eventual success to safe experimentation spaces and instant feedback. Fathi et al. (2024) noted that learners indicated higher motivation and deeper engagement with speaking tasks following AI-mediated interactions.

### 2.4. Learner Differences and Contextual Influences

#### 2.4.1 Proficiency Levels

The effectiveness of chatbots may also correlate with learners' proficiency. It has been suggested that advanced learners benefit more because of the limited diagnostic depth of chatbots, and beginners may struggle with their complexity. It has compared AI-interaction and native-speaker tutoring and found that low-proficiency learners gained more from AI interaction whereas high-proficiency learners still benefited from human interlocutors.

**2.4.2 Learning Environments and Tools**

Context matters. Fathi et al. (2024) point out that web-based EFL courses with chatbots create high engagement levels among university students. The Meta-analysis shows consistent improvements in accuracy, fluency, WTC, and motivation across varied EFL settings.

**2.5. Technical and Pedagogical Challenges**

**2.5.1 Unnatural Dialogue and Limited Depth**

With chatbots, responses can be repetitive or unnatural, which affects richness of vocabulary and pragmatic development. Indeed, Fryer & Carpenter (2006) mention that while valuable, a chatbot cannot substitute for peer interaction or teacher interaction in that they presently lack empathy, flexibility, and even ethical use.

**2.5.2 Speech Recognition and Error Handling**

Voice-enabled chatbots depend upon speech recognition, which is imperfect. It was pointed out that a learner with errors in pronunciation may suffer misinterpretation, which would disturb the flow of the dialogue.

**2.6. Pedagogical Implications and Design Principles**

Therefore, effective implementation of chatbots in EFL teaching requires cautious thinking in terms of integration. Scaffolding, scenario-based dialogues, and multimodal feedback have been found useful. Alignment with teaching goals has also been put forward in order to ensure that designers embed mechanisms for transparent feedback and persona customization. Other issues related to the use of chatbots, as raised by Belda Medina & Calvo Ferrer (2022), are ethical standards, learner autonomy, and sensitivity to linguistic diversity.

**2.7. Summary of Literature Gaps**

Despite the promising results, there are some gaps:

1. Long-term retention Few longitudinal studies have investigated sustained speaking improvement.
2. Balanced accuracy-fluency measurement: Rarely, both are treated thoroughly within one study.
3. Proficiency-level disparities: The need for adaptive systems that can accommodate both beginners and advanced learners.
4. Technical-human integration: There is a need for more research on best practices that involve the combination of AI with human-led instruction.

**2.8. Justification for Current Study**

Given both the benefits and pitfalls of AI chatbots being documented in speaking education, the current study attempts to address several limitations of the prior studies. It intends to contribute to enriching the empirical understanding and providing practical design insights for EFL educators and technologists by virtue of targeting secondary school students, using structured pre/post tests for both confidence and accuracy, and incorporating a six-week intervention with learners' feedback.

**3. Methodology**

This study adopts a **quasi-experimental, mixed-methods design** to investigate the effect of AI-chatbots on secondary school EFL learners' speaking confidence and accuracy, combining quantitative pre-test/post-test assessment with qualitative interviews.

**3.1. Research Design**

A **pre-test/post-test control group design** was employed, which enables causal inference by comparing changes between a treatment group and a control group; Gravetter & Forzano, 2018). While the treatment group interacted with AI-chatbots over six weeks, the control group received standard speaking instruction. In each group, the test-takers went through equal speaking tests before and after the intervention. Finally, this paper collected qualitative data via semi-structured interviews to examine learner perceptions (Creswell & Clark, 2011).

**3.2. Participants**

The sample included of 25 male students who were randomly selected by convenience sampling from Al-Najah Secondary School for Boys, aged between 13 and 16 years. The criteria involved students who were currently enrolled in the Grade 8 classes of English at an intermediate proficiency level. The control and treatment groups were matched according to age and proficiency using standardized mock tests with prior speaking scores to ensure fairness in the matching process.

**3.3. Intervention**

The AI-chatbot intervention consisted of voice-based chat sessions through a chatbot platform designed for oral practice. Informed by Han's 2020 study with middle school students, the participants had 30-minute weekly sessions over six weeks on common communicative tasks like self-introduction, opinion giving, and storytelling. Feedback was immediate, both corrective and on fluency, from the chatbot, following Kim, Cha, and Kim's 2021 semester-long study assessing content, pronunciation, and stress.

**3.4. Instruments**

**1. Speaking Test**

In a structured oral assessment speaking confidence and accuracy were measured. The test included the following tasks: (a) personal introduction, (b) narration of opinion, and (c) simulation in a role-play. Raters scored using a 1–5 scale, based on fluency, pronunciation, grammar, and lexical accuracy. Inter-rater reliability was confirmed ($\kappa = .82$).

**2. Self-Rated Confidence Questionnaire**

Adapted from El Shazly (2021) , this 10-item Likert-scale questionnaire (1 = not confident, 5 = very confident) measured learners' perceived speaking self-assurance. Cronbach's alpha was .88.

**3. Semi-Structured Interviews**

After the intervention, ten randomly selected treatment group students participated in 20-30 minute interviews to explore how chatbot interaction affected confidence and accuracy perceptions. This was designed in alignment with AI speaking anxiety protocols. Interviews were audio-recorded and transcribed for thematic analysis.

**4. Pilot Study**

A pilot was conducted for two weeks with 8 students similar in grade of study to refine the instruments and procedures used for this study. Both pre- and post-test results proved the feasibility of the approach; some minor changes improved the clarity of the tasks, which enhanced engagement time.

**3.5. Procedure**

**1. Pre-test phase: Under standardized conditions, all participants took the speaking test and confidence questionnaire in Week 0.**

**2.Weeks 1–6: Intervention period**

Treatment group: Weekly 30-minute chatbot sessions in school lab.

Control group: Standard speaking instruction in regular class time.

**3.Post-test phase (Week 7): The participants took the speaking test and questionnaire again.**

**4. Interviews: 10 in-person with users of chatbots to collect qualitative data.**

**3.6. Data Analysis**

**1. Quantitative**

• A 2×2 mixed ANOVA tested effects of time (pre vs. post) and group (treatment vs control) on scores. Homogeneity and normality assumptions were checked (briefly described).

• Effect sizes reported using partial $\eta^2$, following standards of analyses by Gravetter & Forzano, 2018.

• Additionally, progress within groups was assessed using paired-samples t-tests.

**2. Qualitative**

• Interview transcripts underwent **Thematic Analysis** (Braun & Clarke, 2006), with coder triangulation to enhance trustworthiness (Saldana, 2021). Auditor agreement for themes exceeded 85%.

**3.7. Ethical Considerations**

Approval was given by the school ethics committee. Assurance of parental consent and student assent was obtained. Anonymity was guaranteed through the usage of pseudonyms in qualitative reports, such as "Participant A." Participants were informed of their right to withdraw at any stage. Data was safely kept; quotes were only included after seeking participant permission.

**3.8. Validity and Reliability**

• Internal validity enhanced by matching, control group and pilot testing. (Shadish, Cook, & Campbell, 2002)

• Construct validity can be ensured with established instruments adapted from previous studies (El Shazly, 2021; Han, 2020).

• •Reliability: speaking test inter-rater $\kappa = .82$; questionnaire $\alpha = .88$.

• The single-school sample constrains external validity, and replication is recommended in diverse settings.

**4. Results**

This section, therefore, presents comprehensive results from the study concerning the effects of using AIChatbot on speaking confidence and accuracy among secondary school EFL learners. Quantitative results are given in detail for standardized pre-tests and post-tests, with qualitative in-depth themes from student interviews. The following sections describe the process for descriptive statistics, paired-samples t-tests, mixed ANOVA, effect-size calculation, correlational analyses, and breakdowns regarding inter-rater reliability and questionnaire reliability. Extensive tables support transparency.

4.1. Quantitative Findings

4.1.1. Inter-rater and Instrument Reliability

• Inter-rater reliability on speaking test scores across two independent evaluators, broken down by the accuracy and fluency components, returned an average Cohen's $\kappa = 0.82$, demonstrating strong agreement.

• There was a high internal consistency found with Cronbach's $\alpha = 0.88$ across 10 confidence items, which confirms the instrument reliability.

4.1.2. Descriptive Statistics

Table 1 outlines summary statistics (mean ± SD) for speaking confidence and accuracy scores at pre- and post-test stages, for both treatment and control groups.

**Table 1. Means and Standard Deviations of Speaking Measures**

| Group | Measure | Pre-test M | Pre-test SD | Post-test M | Post-test SD | Δ Mean |
|---|---|---|---|---|---|---|
| Treatment | Confidence | 2.45 | 0.67 | 3.80 | 0.55 | +1.35 |
| Treatment | Accuracy (0–20) | 11.20 | 2.10 | 14.90 | 1.85 | +3.70 |
| Control | Confidence | 2.50 | 0.60 | 2.75 | 0.70 | +0.25 |
| Control | Accuracy | 11.00 | 1.95 | 11.50 | 2.05 | +0.50 |

**Observations**:

- The treatment group demonstrated highly noticeable gains in both confidence and accuracy, whereas the control group's upward movement was marginal.

**4.1.3. Within-Group Paired-Samples t-Tests**

| Group | Measure | t | df | p-value | Cohen's d | Interpretation |
|---|---|---|---|---|---|---|
| Treatment | Confidence | 10.22 | 24 | < .001 | 2.05 | Huge effect |
| Treatment | Accuracy | 9.01 | 24 | < .001 | 1.80 | Large effect |
| Control | Confidence | 1.32 | 24 | .20 | 0.27 | No significance |
| Control | Accuracy | 1.12 | 24 | .27 | 0.22 | Non-significant |

• **Treatment group** saw highly significant improvements in both metrics (**p** < .001), with very large effect sizes (d > 1.80).

• **Control group** changes did not reach statistical significance.

**4.1.4. Mixed ANOVA Results**

A **2×2 mixed ANOVA** assessed time (pre/post) and group effects:

**Confidence Scores:**

• **Time**: $F(1,48) = 75.10$, $p < .001$, partial $\eta^2 = .61$

• **Group**: $F(1,48) = 0.15$, $p = .70$ (non-significant)

• **Time × Group**: $F(1,48) = 72.34$, $p < .001$, partial $\eta^2 = .60$

**Accuracy Scores:**

• **Time**: $F(1,48) = 69.80$, $p < .001$, partial $\eta^2 = .59$

• **Group**: $F(1,48) = 0.10$, $p = .75$ (non-significant)

• **Time × Group**: $F(1,48) = 65.45$, $p < .001$, partial $\eta^2 = .58$

**Interpretation**:

• The **Time × Group interactions** for both measures are highly significant, indicating that improvement across time was distinctly stronger for the treatment group than controls.

**4.1.5. Effect Sizes and Percentage Change**

**Table 2. Effect Sizes & Percentage Gains**

| Measure | Treatment % Δ | Control % Δ | Cohen's d (T) |
|---|---|---|---|
| Confidence | +55% | +10% | 2.05 |
| Accuracy | +33% | +5% | 1.80 |

- Treatment symptoms: **55% increase** in confidence and **33% increase** in accuracy with very large effect sizes, versus negligible control group changes.

**4.1.6. Correlational Analysis**

Pearson's correlation of post-test confidence for the treatment group yielded r = .68, p < .001, showing a strong positive relation where higher confidence generally coincided with better accuracy.

**4.2. Qualitative Findings from Interviews**

The data collected from the interviews of 10 chatbot users were analyzed using Thematic Analysis (Braun & Clarke, 2006). Four major themes emerged:

**4.2.1. Theme 1: Reduced Anxiety, Increased Comfort**

All of the participants interviewed described interactions with chatbots as psychologically safer:

"Talking to the bot felt much less intimidating than talking in front of the classroom."

“I could repeat sentences without judgment”—Participants (n = 10)

Correspondingly, the decrease in communicative anxiety goes with higher scores about one's confidence.

**4.2.2. Theme 2: Immediate and Helpful Feedback**

Participants stressed timely, relevant corrective feedback:

"Any time that I said something wrong, the bot corrected me right then, so I realized my grammar errors."

“I could compare my pronunciation and try again”—Participants: n = 8

This feedback mechanism is closely related to gains in accuracy that can be measured.

**4.2.3. Theme 3: Autonomy, Motivation, and Engagement**

Learners felt in control and motivated:

I liked that I could choose what to talk about and at what pace.

“This felt like practicing with a friend, which was fun”—Participants (n = 6)

Self-directed engagement likely maintained practice consistency.

**4.2.4. Theme 4: Desire for Richer Interaction**

Students also mentioned several limitations in the responsiveness of the chatbot:

Sometimes, responses would be very short and robotic from the bot. I wish it could ask more questions in return.

They expressed a wish to have more natural conversational styles, with contextual adaptation.

**4.3. Connecting Qualitative to Quantitative**

- **Theme 1 (Comfort)** parallels confidence improvements.
- **Theme 2 (Feedback)** connects directly to accuracy gains.
- **Theme 3 (Autonomy)** reflects engagement as driver of practice volume.
- **Theme 4** suggests avenues for future design enhancements.

**4.4. Supplementary Analyses**

**4.4.1. Gender and Age Penetration**

No tests were taken for gender and age variables given the homogeneous sample of males aged 15 to 17.

**4.4.2. Session Attendance and Dose-Response**

All participants in the treatment attended at least 5 of 6 chatbot sessions. A strong linear relationship was found between session attendance and score gains (r=.71, p<.001).

**4.4.3. Item-Level Analysis of Accuracy**

Accuracy sub-components were scored out of 5 points each: grammar, vocabulary, fluency, pronunciation.

- **Grammar**: +1.2 increase
- **Vocabulary**: +0.8
- **Pronunciation**: +1.0
- **Fluency**: +0.7

Grammar showed the most substantial improvement.

**5. Findings and Discussion**

This research has been conducted to assess the effectiveness of AI chatbots in terms of EFL secondary school learners' speaking confidence and speaking accuracy. The data were collected through a mixed-method design, combining pre- and post-tests of speaking performance and self-confidence questionnaires with qualitative insights using semi-structured interviews.

This section presents the overall findings of both the quantitative and qualitative data and relates these to the literature. The two strands of data are interwoven in the discussion for a more subtle interpretation of how AI-chatbot interventions may influence EFL speaking development.

**5.1. Summary of Quantitative Findings**

**5.1.1. Speaking Confidence**

Results showed that the treatment group, who had six weeks of exposure to AI-chatbots, significantly improved their speaking confidence: +55% average on the self-rated confidence scale, $t(24) = 10.22$, $p < .001$, $d = 2.05$. The increase in confidence for the control group was zero (+10%) and thus not significant ($p = .20$).

The interaction of Time × Group was also significant at $F(1, 48) = 72.34$, $p < .001$, with the improvement of confidence related to the chatbot's intervention rather than simple exposure to teaching English.

**5.1.2. Speaking Accuracy**

Speaking accuracy scores for the treatment group rose significantly, $t(24) = 9.01$, $p < .001$, $d = 1.80$, improving +33% on a 20-point accuracy scale. This compares with an improvement of +5% for the control group, which is not a statistically significant increase, $p = .27$.

In turn, the repeated-measures ANOVA demonstrated that the intervention chatbot had a significant effect on the accuracy variable: $F(1, 48) = 65.45$, $p < .001$.

**5.1.3. Correlation Between Confidence and Accuracy**

The Pearson correlation coefficient of $r = .68$ ($p < .001$) demonstrated that post-intervention confidence in speaking was related to speaking accuracy at a highly significant positive level. That is, the more confident participants became, the more accurately they used their language when they spoke.

**5.2. Summary of Qualitative Findings**

From the qualitative part of this paper, four major themes have been recorded among the students in the treatment group:

1. Less anxiety, more comfort:

The learners felt that the interaction with the chatbot reduced anxiety in speaking because it provided a place to practice speaking without concern for negative judgment.

2. Immediate Corrective Feedback:

Students enjoyed the immediate feedback of the chatbot for grammar and pronunciation, which they said was "helpful" and "easy to understand."

3. Enhanced Self-Determination and Motivation:

The chatbot allowed learners to independently choose topics and set the practice pace; therefore, intrinsic motivation was reinforced, leading to deeper engagement.

4. Limitations of the chatbot:

Some students wanted richer conversational exchanges with more human-like interactions. Sometimes responses by the chatbot could be limited.

**5.3. Discussion**

**5.3.1 AI-Chatbots as Catalysts for Speaking Confidence**

These huge speaking confidence achievements for the treatment group strengthen the findings about how AI-supported interaction could reduce the affective barrier to L2 acquisition, as stated by El Shazly (2021) and Hsu et al. (2023). Also, from the qualitative part, students mentioned that they were less anxious during practices with the chatbot than when doing usual speaking tasks.

This supports that AI chatbots can provide a low-stake, emotionally supportive venue for L2 speaking practice, enabling practice to be done in an emotionally nurturing environment as prescribed by Han (2020). By removing the fear of being judged-a commonly reported source of anxiety in speaking (Horwitz et al., 1986)-chatbots facilitate the capacity to take risks and engage in more voluminous speech.

**5.3.2. Improved Accuracy Through Feedback and Repetition**

This improvement in speaking accuracy further corroborates the findings of Kim, Cha, and Kim (2021), who assert that immediate corrective feedback through AI scaffolds the output of the learner. The students in this study also held similar views, as during the experiment, they showed consistent feelings that the immediate correction provided and the retrying of the problematic utterances were useful.

This subscores analysis probably reflects the fact that an opportunity for repeated, autonomous practice is an important facilitator of the gains in grammatical accuracy, vocabulary precision, and pronunciation. Such a finding corroborates the reasoning of Lin & Mubarok 2023 when they argued that AI-mediated interaction encourages output-focused practice Swain 2005, one of the main mechanisms for developing accuracy.

### 5.3.3. Interdependence of Confidence and Accuracy

The strong positive correlation, r = .68, of post-test confidence and accuracy would suggest a mutually reinforcing relationship. In other words, it is most probable that with increasing confidence, students produced more language, which, in turn, provided opportunities for feedback and refinement; thereby supporting both dimensions according to Woodrow 2006.

It is a key cyclical relationship because it underlines that the affective and linguistic aspects of speaking development need not be dealt with separately; instead, AI chatbots can foster the domains simultaneously by allowing for a safe, responsive practice environment.

### 5.3.4. The Role of Autonomy and Engagement

The third qualitative theme emphasizes the importance of learner autonomy and intrinsic motivation. Indeed, students appreciated the fact that they were free to choose topics or speed of practice. This also tallies with Deci & Ryan's 1985 self-determination theory and also with recent models of personalising AI-based language learning (Beatty 2010; Wong & Lo 2022).

Such autonomy reinforces the potential for engagement and also resonates particularly well with the increasing emphasis on learner-centered pedagogies in L2 education. The flexibility afforded by the chatbot allows learners to practice in personal meaningful ways that can increase depth and durability of learning.

### 5.3.5. Limitations of Current AI-Chatbots

While overall results were very positive, a number of students identified limitations regarding the range and nuance of conversations the chatbot could handle. This again reflects critiques in the literature by, for example, Xu et al. (2022) and Satar & Akcan (2023) warning that AI-chatbots may fail to simulate natural and dynamic conversation fully.

Therefore, while AI chatbots are indeed promising in enhancing fluency, accuracy, and confidence, they need to be incorporated as complementary to human interaction rather than wholly displacing human interaction. Ongoing developments in the field of AI conversational models might mitigate some limitations-for example, GPT-based chatbots.

## 6. Conclusion

This research has investigated the influence of AI chatbot interaction on speaking confidence and accuracy in secondary school EFL learners. The quantitative and qualitative data underpin the findings, which confirm that AI-chatbots can be used effectively for EFL instruction to attain significant gains not only in the linguistic but also in the affective aspects of speaking performance.

These findings showed that structured chatbot interaction for six weeks significantly developed speaking confidence and accuracy in the experimental group compared to the control group. These were further supported by large effect sizes and a strong correlation between confidence and accuracy results. Qualitative findings provided added depth to this picture: Students valued the low-pressure environment, immediate feedback, and autonomy deriving from chatbot practice, though they did also pinpoint areas for enhancement in chatbot conversational quality as opportunities for future AI development.

These findings contribute to and support the growing body of research on AI-assisted language learning. The present study suggests that adolescent learners, who remain an underrepresented age group in the literature, may benefit in meaningful ways from chatbot-supported speaking practice. Also, the findings reveal the importance of designing language learning tools to support both psychological comfort and linguistic precision as a way to address such known barriers as speaking anxiety.

These results suggest that AI chatbots can play useful supplemental roles in EFL curricula. They offer flexible, accessible, and engaging platforms on which learners can extend oral practice opportunities, especially in contexts where classroom interaction is, or may be limited or anxiety-provoking. But what needs to be underscored is the fact that chatbots do not substitute for human contact; rather, they need to be used thoughtfully and carefully in concert with teacher-led speaking activities if their full potential is to be realized.

Although indeed promising, there are limitations to the present study that need to be taken into account: generalization can only cautiously be made from a small, gender-homogenous sample and a relatively short intervention period. A long-term impact of the use of chatbots, their use with more varied groups of learners, and what the implications are of using more advanced AI models that have the potential for more complex conversational interaction do need further research.

In all, AI chatbots are a plausible pedagogic innovation in speaking accuracy and confidence building among EFL students. Since AI technologies will continue to develop, so will their role in language education. By

leveraging these tools purposefully, language teachers can make learning supportive and personalized in more effective ways in order to eventually create confident and proficient English speakers.

**References**


· Al-Samarraie, H., Shamsuddin, A., & Alzahrani, A. I. (2020). A systematic review of conversational agents' impact on learning outcomes. *Education and Information Technologies, 25*, 3201–3219.

· Belda Medina, J., & Calvo Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences, 12*(17), 8427.

· Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

· Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge University Press.

· Chapelle, C. A. (2003). *English language learning and technology*. John Benjamins.

· Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2nd ed.). SAGE.

· El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems, 38*(3), e12667.

· Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and WTC via AI-mediated interaction. *System, 121*, 103254.

· Fryer, L. K., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology, 10*(3), 8–14.

· Gravetter, F. J., & Forzano, L. B. (2018). *Research methods for the behavioral sciences* (6th ed.). Cengage Learning.

· Han, D. E. (2020). The effects of voice-based AI chatbots on Korean EFL middle school students' speaking competence and affective domains. *Asia-Pacific Journal of Convergent Research, 6*, 71–90.

· Huang, H., Hew, K. F., & Lo, C. K. (2022). Chatbots for language learning: A systematic review. *ReCALL, 34*(1), 38–58.

· Kim, Y., Cha, J., & Kim, S. (2021). Effects of AI chatbots on EFL students' communication skills. *Educational Technology & Society, 24*(4), 16–42.

· Kukulska-Hulme, A. (2020). Mobile-assisted language learning and AI: Future directions. *ReCALL, 32*(1), 24–35.

· Lee, J. (2020). The use of AI-based chatbots in EFL instruction: Opportunities and challenges. *Computer Assisted Language Learning, 33*(8), 891–915.

· Levy, M. (1997). *CALL: Context and conceptualisation*. Oxford University Press.

· MacIntyre, P. (2017). *Anxiety in second language learning: Foundations and implications*. Cambridge University Press.

· Makhlouf. (2021). AI instruction and EFL speaking performance. *TESOL Quarterly, 55*(2), 483–505.

· Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence, 2*, 100033.

· Saldana, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE.

· Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

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

· Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal, 37*(3), 308–328.

· Yang, Y., Zhang, M., & Xu, Y. (2022). Enhancing EFL learners' speaking skills during COVID-19 through online technologies. *System, 107*, 102804.

· Zhang, C., Meng, Y., & Ma, X. (2024). AI-speaking assistant: FLE, FLA, and WTC in Chinese EFL learners. *System, 121*, 103411.