AI-Powered Chatbots for English Conversation Practice: Impact on Speaking Proficiency

Mohammad Aliakbari[1], Pooria Barzan*[2], Seyyed Pedram Allahveysi[3], Morteza Bakhtiarvand[4]

[1]Department of English Language and Literature, Ilam University, Ilam, Iran. m.aliakbari@ilam.ac.ir

[2]Corresponding author, Department of English Language and Literature, Ilam University, Ilam, Iran. p.barzan@ilam.ac.ir

[3]Department of Language and Linguistics, University of Essex, Colchester, UK. s.p.allahveysi@essex.ac.uk

[4]Instructional Technology at Europa-Universität Flensburg, Flensburg, Germany. m.bakhtiarvand@atu.ac.ir

Article Info	Abstract
	In the era of advancing artificial intelligence in education, this study investigates whether regular practice with AI-powered chatbots improves English speaking proficiency more effectively than traditional peer-to-peer conversation practice. Grounded in the interaction hypothesis (Long, 1996) and output hypothesis (Swain, 1985), which emphasize the role of interactive practice and feedback in language acquisition, the research employed an experimental design. Two groups of English learners participated over 12 weeks: an experimental group practiced speaking with Elsa AI, an AI chatbot equipped with natural language processing for real-time feedback on pronunciation, vocabulary, grammar, and fluency, while a control group engaged in peer-to-peer conversation practice. Speaking proficiency was assessed pre- and post-intervention using standardized tests (IELTS speaking), and statistical methods (e.g., t-tests or ANOVA) were used to compare proficiency gains. Statistical analyses revealed that both groups improved significantly over time, but the chatbot group achieved larger mean gains than the peer group. These findings suggest that AI chatbots offer scalable, on-demand practice with immediate corrective feedback, addressing limitations of human partners such as availability and consistency. The findings could inform the integration of chatbots in diverse English Language Teaching (ELT) settings, enhancing speaking practice opportunities globally.

Introduction

English-speaking proficiency is a cornerstone of communicative competence in second-language acquisition, essential for effective participation in academic, professional, and social contexts in an increasingly globalized world (Huang et al., 2021; Yu & Nazir, 2021). The ability to communicate fluently in English, recognized as the lingua franca of international exchange, directly correlates with enhanced opportunities and outcomes (Zhang, 2025). However, learners frequently encounter significant barriers to developing this skill, including limited access to sustained, meaningful oral practice in traditional classroom settings and with human interlocutors (Huang et al., 2021; Warschauer, 1997). Traditional peer-to-peer conversation practice, while a staple of communicative language teaching, is often hampered by practical constraints such as variability in peer proficiency, inconsistent feedback quality, and restricted availability of conversation partners (Kuhail et al., 2022; Belda-Medina & Calvo-Ferrer, 2022; Shadiev & Yang, 2020). Additionally, affective factors like speaking anxiety and lack of confidence further impede learners' willingness to engage, exacerbating the gap between classroom instruction and real-world communicative demands (Kohnke, 2022). These challenges highlight the urgent need for innovative, scalable solutions to enhance speaking practice opportunities.

Immediate corrective feedback, defined as real-time responses to learner errors that facilitate noticing and self-correction (Singh & Halim, 2023), plays a pivotal role in language acquisition by promoting linguistic accuracy and retention. Recent advancements in artificial intelligence (AI), particularly in natural language processing (NLP) and automatic speech recognition (ASR), have catalyzed the development of AI-powered chatbots capable of simulating human-like conversations (Patra & Kumar, 2020). Tools such as Elsa AI can process spoken input in real time, detect errors across pronunciation, vocabulary, grammar, and fluency, and deliver immediate, adaptive feedback (Ye et al., 2022; Nguyen & Pham, 2021). Operating 24/7 and tailoring interactions to individual learner needs, these chatbots offer a consistent, on-demand alternative to peer practice, mitigating accessibility issues and variability in feedback (Belda-Medina & Calvo-Ferrer, 2022; Yang et al., 2022). By providing a non-judgmental environment, they also reduce anxiety, fostering greater learner engagement and confidence (Ballıdağ & Aydın, 2025; Shazly, 2021). This technological evolution prompts a re-evaluation of pedagogical strategies in English Language Teaching (ELT), positioning AI chatbots as potential game-changers in language education.

Empirical research on chatbot-mediated speaking practice has yielded promising yet preliminary insights. Studies such as Ye et al. (2022) report significant gains in pronunciation accuracy and fluency among learners using smartphone-based chatbots, while Xu et al. (2024) note enhanced engagement with ChatGPT-based agents for IELTS preparation. A meta-analysis by Lyu et al. (2024) found moderate-to-large proficiency improvements from chatbot interventions, corroborating their potential (Lyu et al., 2024). However, much of this research relies on single-group designs or descriptive evaluations, with few controlled experimental

comparisons to traditional peer practice (Kohnke et al., 2023; Lyu et al., 2024). Moreover, evidence on chatbot effectiveness relative to human interlocutors, using standardized measures like the IELTS speaking test, remains scarce (Xu et al., 2024; Buckingham & Alpaslan, 2017). This gap underscores the need for rigorous investigation into whether AI-driven practice can outperform conventional methods in fostering speaking proficiency, which the present study addresses through a direct comparison against the de facto classroom alternative (peer practice), using a gold-standard assessment like the IELTS speaking test—thereby overcoming key methodological limitations in existing research.

To address this, the present study examines whether regular practice with an AI-powered chatbot, specifically Elsa AI, yields greater improvements in English-speaking proficiency than traditional peer-to-peer conversation practice during a 12-week intervention. This research aims to provide empirical evidence on the comparative efficacy of these modalities, testing the hypothesis that the chatbot's scalable, feedback-rich interactions enhance speaking skills more effectively than peer-based exchanges. By grounding the investigation in the Interaction and Output Hypotheses, it bridges theoretical SLA principles with practical application.

The findings hold significant implications for ELT pedagogy and technology integration. Should chatbot-mediated practice prove superior, it could encourage broader adoption of AI tools to supplement classroom instruction, expanding access to high-quality, individualized speaking practice globally, particularly in resource-scarce contexts (Hidayat et al., 2022; Zhang, 2025). Furthermore, insights from learner interactions with Elsa AI may inform the design of future educational chatbots, enhancing their pedagogical utility (Nguyen & Pham, 2021). Ultimately, this study contributes to evidence-based decision-making in language education by offering a replicable framework for evaluating AI-driven interventions and expanding opportunities for English learners worldwide. The following research questions guide the study:

- To what extent does practice with an AI-powered chatbot improve English speaking proficiency as measured by the IELTS speaking test?

- How do proficiency gains from AI chatbot practice compare to those from traditional peer-to-peer conversation practice?

- What implications do these findings have for integrating AI tools in ELT settings?

Literature review

The integration of AI-powered chatbots into English as a Foreign Language (EFL) and English Language Teaching (ELT) contexts has emerged as a transformative approach to enhancing speaking proficiency. By leveraging advanced NLP and machine learning, these systems provide scalable, on-demand conversational practice with immediate, personalized feedback on pronunciation, grammar, vocabulary, and fluency (Pérez et al., 2020; Ye et al., 2022; Zhang, 2024). This literature review synthesizes theoretical foundations, technological advancements,

empirical evidence, affective impacts, comparisons with traditional methods, and challenges of employing AI chatbots in English conversation practice.

Theoretical Foundations

The efficacy of AI chatbots in language learning is theoretically supported by foundational Second Language Acquisition (SLA) frameworks, notably Long's Interaction Hypothesis (1996) and Swain's Output Hypothesis (1985). The Interaction Hypothesis asserts that negotiated interaction—where learners resolve communication breakdowns—enhances language acquisition by making input comprehensible and by promoting opportunities for output (Atta, 2022; Kohnke, 2022). Likewise, it posits that negotiation of meaning and corrective feedback during interaction facilitate interlanguage development (Yang et al., 2022; Zou et al., 2025). AI chatbots operationalize these principles by engaging learners in interactive dialogues and delivering real-time feedback, such as recasts and clarification requests, in response to learner errors (Daniels, 2015; Hoang et al., 2023; Vančová, 2023; Zhang, 2024). The Output Hypothesis emphasizes that producing language, especially when coupled with corrective feedback, drives learners to notice linguistic gaps and restructure their interlanguage (Atta, 2022; Kohnke, 2022; Swain, 1985). Chatbots foster this process by encouraging repeated output in low-stakes environments, enabling learners to test hypotheses, refine their skills, and benefit from immediate feedback that promotes linguistic accuracy and retention (Carroll et al., 1992; Hoang et al., 2023; Saito, 2021; Vančová, 2023).

Technological Advancements

AI chatbots represent a significant evolution from early Computer-Assisted Language Learning (CALL) systems, which relied on static, drill-based exercises (Bodnar et al., 2011). Modern chatbots employ sophisticated NLP algorithms, including tokenization, syntactic parsing, and transformer-based models, to interpret and generate contextually appropriate responses (Lee et al., 2010; Yang et al., 2022). These systems provide multidimensional feedback across pronunciation, grammar, vocabulary, and fluency, adapting to individual learner profiles through machine learning (Pérez et al., 2020; Zhang, 2024). Notably, pronunciation-focused tools like Elsa AI differ from open-ended generative AI such as ChatGPT by emphasizing targeted, ASR-driven feedback on specific speaking metrics (e.g., intonation, phonemes), rather than broad conversational generation, making them particularly suited for proficiency assessments like IELTS that score pronunciation and fluency discretely (Kohnke et al., 2023; Xu et al., 2024). Multimodal interfaces, integrating text, voice, and visual prompts, further enhance engagement by supporting both written and spoken dialogue (Kemelbekova et al., 2024). Generative AI models, such as ChatGPT, enable dynamic, human-like conversational simulations, though pedagogical alignment remains critical to mitigate risks like hallucination (Wang et al., 2025; Xu et al., 2024).

Empirical Evidence on Speaking Proficiency

Empirical studies consistently demonstrate that AI chatbot interventions yield significant gains in speaking proficiency across diverse learner populations. Ye et al. (2022) reported improvements in grammar accuracy, pronunciation, and oral performance among undergraduates who used a smartphone-based chatbot, compared with a non-practice control group. Similarly, Duong and Suppasetseree (2024) found that Vietnamese EFL learners improved fluency and complexity after biweekly chatbot sessions over eight weeks. Recent research further supports these findings; for instance, Grab (2025) explored integrated AI chatbot practice in multicultural ESL settings, showing enhanced speaking skills through personalized interactions, while Du and Daniel (2024) conducted a systematic review highlighting improved confidence and engagement in English speaking via AI tools (Du & Daniel, 2024; Grab, 2025). Additionally, Ding and Yusof (2025) investigated AI conversation bots' role in enhancing speaking proficiency, reporting positive outcomes in naturalistic learning environments (Ding & Yusof, 2025). Meta-analyses, such as Wu and Yu (2023), confirm moderate to large effect sizes on speaking outcomes, though heterogeneity arises from variations in chatbot design and learner proficiency. Pronunciation training, a key focus, benefits from speech recognition technologies, with studies showing reductions in segmental and suprasegmental errors through immediate, multimodal feedback (Hoang et al., 2023; Vančová, 2023). Additionally, chatbots integrated into task-based and Content and Language Integrated Learning (CLIL) contexts enhance fluency and content knowledge by facilitating authentic, discipline-specific dialogues (Mageira et al., 2022; Kemelbekova et al., 2024).

Affective and Motivational Impacts

AI chatbots positively influence affective factors critical to language learning, such as reduced speaking anxiety and increased willingness to communicate (WTC). Their nonjudgmental nature creates a low-anxiety environment, encouraging risk-taking and extended speaking turns (Aliakbari et al., 2025; Ballıdağ & Aydın, 2025; Zou et al., 2025). Shazly (2021) documented concurrent reductions in anxiety and improvements in task performance, while Zou et al. (2025) noted increased WTC over time, linked to greater fluency and complexity. The motivational advantage of chatbots stems from their accessibility and personalized feedback, fostering learner autonomy and self-regulated learning (Zhang, 2024; Zhou & Privitera, 2024). Qualitative data indicate high learner satisfaction with the immediacy and objectivity of chatbot feedback, which reduces affective filters and enhances engagement (Saito, 2021; Wanderley et al., 2021).

Comparisons with Traditional Methods

Comparative studies highlight the advantages of chatbot-based practice over traditional peer-to-peer and teacher-mediated interactions. Ballıdağ and Aydın (2025) found that chatbot users exhibited lower anxiety and equivalent accuracy gains compared to peer interaction groups, attributed to consistent feedback and availability. Kim et al. (2021) reported superior improvements in pronunciation and lexical variety with chatbots, due to their fidelity in

feedback delivery. While peer interactions offer socio-cultural negotiation, their effectiveness is often limited by partner proficiency and scheduling constraints (Wang, 2023; Zou et al., 2023). In contrast, chatbots provide uniform, immediate corrections, mitigating variability in human feedback (Zhang, 2024). However, human-mediated feedback retains advantages in addressing complex pragmatic nuances, suggesting a complementary role for AI in hybrid learning models (Lee, 2002; Singh & Halim, 2023).

Challenges and Limitations

Despite their promise, AI chatbots face several challenges. Developing high-quality, pedagogically aligned content is resource-intensive and requires expertise in both NLP and instructional design (Koç & Savaş, 2024; Nhạc, 2021). Technical limitations in NLP can hinder the handling of nuanced or context-dependent interactions, potentially leading to inaccurate feedback (Huang et al., 2021; Wang et al., 2025). The digital divide, encompassing disparities in access and digital literacy, poses barriers to equitable implementation (Zhang et al., 2024). Additionally, novelty effects and cognitive load may diminish long-term engagement, necessitating careful design to sustain learner motivation (Fryer et al., 2017). Human oversight remains essential to address complex linguistic and cultural subtleties that automated systems may overlook (Ulfa, 2023).

While the reviewed literature establishes the potential of AI chatbots for enhancing speaking proficiency and highlights their affective benefits over traditional methods, a critical methodological gap persists. Rigorous experimental studies directly comparing the efficacy of AI chatbot-mediated practice with established peer-to-peer conversation practice, particularly using standardized proficiency measures such as the IELTS speaking test, remain scarce. Most existing comparative research either lacks controlled designs or relies on non-standardized assessments, limiting definitive conclusions about the relative effectiveness of AI-driven interaction for measurable speaking gains. This gap underscores the need for precisely the controlled, longitudinal comparison undertaken in the present study.

Method

Design

This study adopted a quasi-experimental pretest-posttest control-group design to compare the effectiveness of AI-powered chatbot practice versus traditional peer-to-peer conversation practice in enhancing English-speaking proficiency. The experimental group engaged in speaking practice with Elsa AI, an AI chatbot equipped with natural language processing and automatic speech recognition capabilities, while the control group participated in structured peer-to-peer conversation sessions. The intervention lasted 12 weeks, during which speaking proficiency was measured both before and after using standardized assessments.

Participants

The study involved 60 intermediate-level non-native English learners (B1-B2 on the CEFR scale) from Ilam University, all of whom were university students in an EFL context. Participants were aged 18–25 (M = 20.3, SD = 1.8) and had similar baseline English proficiency, confirmed via the Oxford Placement Test. Using stratified random sampling, participants were assigned to either the experimental group (n = 30) or the control group (n = 30), ensuring balance in gender, age, and initial proficiency levels.

Instruments

Speaking Proficiency Assessment

Speaking proficiency was assessed using the International English Language Testing System (IELTS) Speaking test (Academic version), a standardized measure widely recognized for its reliability and validity in evaluating English-speaking skills. The IELTS speaking test assesses four key areas: pronunciation, vocabulary, grammar, and fluency. Both pre- and post-intervention assessments were conducted by trained IELTS examiners who were blind to the group assignments to prevent bias. The pre-test was administered one week before the intervention began, and the post-test was conducted one week after the intervention concluded. Participants' scores were recorded on a scale from 0 to 9, with higher scores indicating greater proficiency. Inter-rater reliability was established using Cronbach's α = 0.92; although Intraclass Correlation Coefficient (ICC) is more commonly used for inter-rater agreement, Cronbach's α was selected here as it provides a measure of internal consistency among multiple raters' scores.

AI Chatbot (Elsa AI)

The experimental group used Elsa AI, a mobile-based chatbot designed for English-speaking practice. Elsa AI provides real-time feedback on pronunciation, vocabulary, grammar, and fluency, tailoring responses to individual learner needs. Its natural language processing and automatic speech recognition capabilities enable interactive, human-like conversations, available 24/7.

Peer-to-Peer Conversation Protocol

The control group followed a structured peer-to-peer conversation protocol, involving paired discussions on pre-assigned topics aligned with IELTS speaking themes (e.g., education, technology, travel). Sessions were facilitated by trained instructors who provided peers with a brief training session and a rubric for delivering feedback, focusing on pronunciation, vocabulary, grammar, and fluency, and ensuring structured yet natural variation in peer input.

Procedure

Pre-Intervention

Participants completed the IELTS speaking pretest one week before the intervention began, administered individually in a controlled setting. Baseline demographic and proficiency data were collected via a questionnaire and Oxford Placement Test, respectively.

Intervention

The intervention spanned 12 weeks, during which both groups engaged in speaking practice sessions three times a week. Each session lasted 60 minutes (total: 36 hours), ensuring that both groups received equal practice time. However, the nature of the practice differed between the groups.

Experimental Group (AI-Powered Chatbot Practice): Participants practiced speaking with Elsa AI, an AI-powered chatbot equipped with natural language processing (NLP) and automatic speech recognition (ASR). Elsa AI provided real-time feedback on pronunciation, vocabulary, grammar, and fluency. During each 60-minute session, participants completed a variety of speaking tasks, including describing images, responding to questions, and engaging in simulated conversations. The tasks and topics within the Elsa AI platform were explicitly aligned with the IELTS-themed topics assigned to the control group to ensure equivalence. The chatbot's feedback was immediate and adaptive, tailored to each learner's performance, allowing for personalized correction and reinforcement.

Control Group (Peer-to-Peer Conversation Practice): Participants in the control group engaged in traditional peer-to-peer conversation practice. They were paired with a partner and completed similar speaking tasks to those of the experimental group, such as describing images, answering questions, and engaging in conversations. However, feedback was provided by their peers, guided by instructors rather than an AI system. Pairs rotated weekly to ensure diverse interactions, and instructors provided general feedback on common errors after each session. To ensure consistency, participants were instructed to focus on pronunciation, vocabulary, grammar, and fluency when giving feedback, but the quality and accuracy of peer feedback varied naturally, reflecting real-world conditions of peer practice.

Attendance and engagement were monitored via logs (experimental group) and instructor records (control group).

Post-Intervention

The post-test was conducted one week after the intervention concluded, under identical conditions to the pretest. Post-intervention surveys were administered to gather qualitative insights on participant experiences, though these data are reported separately.

Data Analysis

To compare proficiency gains between the two groups, a two-way repeated-measures Analysis of Variance (ANOVA) was conducted. The within-subjects factor was time (pre-test vs. post-test), and the between-subjects factor was group (experimental vs. control). The dependent variable was the IELTS speaking test score. This statistical method allowed for the examination of both the main effects of time and group, as well as the interaction effect between time and group, which was critical for determining whether the chatbot intervention led to greater improvements than peer practice. Additionally, post hoc independent-samples t-tests were conducted to compare the groups' proficiency levels at each time point (pre-test and post-test) and to analyze differences in proficiency gains. All statistical analyses were performed using SPSS version 27, with a significance level set at $p < 0.05$.

Ethical Considerations

The study adhered strictly to ethical guidelines governing research with human participants, prioritizing participant welfare and data integrity. Participation was entirely voluntary, and informed consent was secured from all individuals prior to their involvement. During the consent process, participants were thoroughly briefed on the study's purpose, procedures, potential benefits, and their unconditional right to withdraw at any point without repercussions. To safeguard privacy, all data were anonymized by replacing personal identifiers with unique codes, and confidentiality was maintained throughout the research process, from data collection to reporting.

Findings

To address the research questions, we compared pre- and post-intervention IELTS speaking scores for the experimental (Elsa AI) and control (peer-to-peer) groups. The primary aim was to determine whether the AI-chatbot practice led to significantly greater gains than structured peer interaction. The main findings are summarized below and presented in Table 1 and Figures 1–2.

Table 1 presents the mean IELTS speaking scores (and standard deviations) for both the experimental (AI-chatbot) and control (peer-to-peer) groups at pre- and post-intervention. It also includes the mean gain scores (post-test minus pre-test) for each group. Both groups began with comparable baseline proficiency (pre-test), but the experimental group exhibited a larger increase by the post-test.

Table 1

Descriptive statistics of IELTS speaking scores by group and time

Group	Time	n	Mean (SD)	Gain Mean (SD)
Experimental (Elsa AI)	Pre-Test	30	5.21 (0.48)	
	Post-Test	30	6.85 (0.61)	+1.64 (0.34)
Control (Peer)	Pre-Test	30	5.18 (0.50)	
	Post-Test	30	6.10 (0.58)	+0.92 (0.29)

Figure 1 illustrates each group's average IELTS speaking scores before and after the intervention, with error bars showing one standard deviation. It highlights that while both groups improved, the experimental (AI-chatbot) group's post-test mean (6.85 ± 0.61) increased more than the control group's (6.10 ± 0.58).
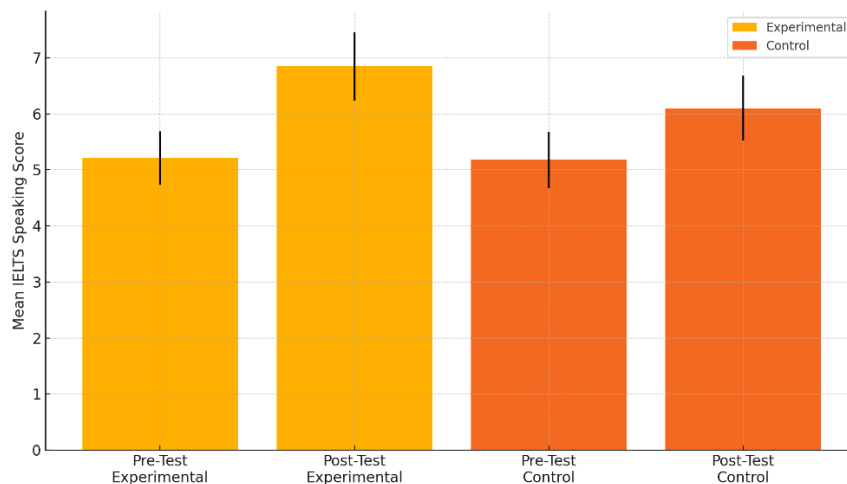
Figure 1 *Pre-Test vs. Post-Test Means*

A two-way repeated measures ANOVA was conducted with Time (pre-test vs. post-test) as the within-subjects factor and Group (experimental vs. control) as the between-subjects factor.

Table 2

ANOVA results for main and interaction effects

Effect	F(1,58)	p-value	η_p^2
Time	450.23	< 0.001	0.89
Group	12.15	0.001	0.17
Time × Group Interaction	25.47	< 0.001	0.31

The analysis revealed a significant main effect of Time, indicating that IELTS speaking scores increased substantially from pre-test to post-test for both groups combined, $F(1, 58) = 450.23$, $p < 0.001$, $\eta_p^2 = 0.89$. There was also a significant main effect of Group, showing that, when averaged across both time points, the experimental group scored higher overall than the control group, $F(1, 58) = 12.15$, $p = 0.001$, $\eta_p^2 = 0.17$. Crucially, a significant Time × Group interaction was found, $F(1, 58) = 25.47$, $p < 0.001$, $\eta_p^2 = 0.31$, demonstrating that the improvement in scores over time was significantly greater in the experimental group than in the control group.

To further explore the interaction, independent-samples t-tests were used to compare the groups at each time point, and gain scores (post-test minus pre-test) were analyzed.

Table 3

Post-hoc t-tests for group differences

Comparison	t(58)	p-value	Experimental M – Control M
Pre-test	0.36	0.72	0.03
Post-test	4.97	< 0.001	0.75
Gain Scores	8.11	< 0.001	0.72

Initial pre-test scores revealed no significant difference between the experimental and control groups (t(58) = 0.36, p = .72), confirming comparable baseline proficiency at the start of the study. However, on the post-test, the experimental group demonstrated significantly higher performance than the control group (t(58) = 4.97, p < .001), with a mean difference of 0.75. Analysis of gain scores further confirmed the effectiveness of the intervention, showing that the mean improvement for the experimental group (M = 1.64, SD = 0.34) was significantly

greater than the mean improvement for the control group (M = 0.92, SD = 0.29), t(58) = 8.11, p < .001.
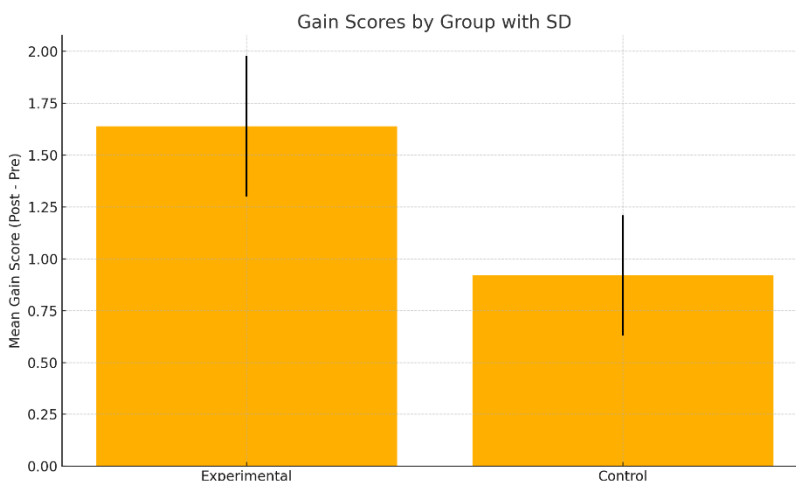
Figure 2 *Gain Scores by Group*

The statistical analyses demonstrate that both AI-powered chatbot practice and traditional peer-to-peer conversation practice led to significant improvements in English-speaking proficiency during the 12-week intervention. However, the AI chatbot group achieved larger mean proficiency gains than the peer-to-peer group. The significant interaction effect in the ANOVA and the post-hoc t-tests confirms that the AI chatbot intervention was more effective in enhancing speaking skills. These findings support the hypothesis that the AI chatbot provides immediate, personalized feedback and offers a scalable and effective alternative to traditional practice methods in English language teaching.

Discussion

The present study examined the comparative efficacy of AI-powered chatbot practice (using Elsa AI) versus traditional peer-to-peer conversation practice in improving English speaking proficiency over a 12-week intervention. Grounded in the Interaction Hypothesis (Long, 1996) and the Output Hypothesis (Swain, 1985), it was hypothesized that learners engaging with the chatbot would experience greater gains, owing to its capacity for immediate, tailored feedback. The quantitative results—significant main effects of time and group, as well as a robust time×group interaction—unequivocally support this hypothesis: while both groups improved, the chatbot cohort exhibited significantly larger proficiency gains.

The superior performance of the AI chatbot group can be interpreted through the lens of foundational SLA theories, particularly the Interaction Hypothesis (Long, 1996) and the Output Hypothesis (Swain, 1985). The Interaction Hypothesis posits that language acquisition is facilitated by negotiated interaction and corrective feedback, which make input comprehensible and promote opportunities for output. Similarly, the Output Hypothesis emphasizes the importance of producing language and receiving feedback to notice and correct linguistic gaps.

In this study, the AI chatbot's ability to provide immediate, personalized feedback on pronunciation, vocabulary, grammar, and fluency likely enhanced the quality of interaction and output, leading to more substantial proficiency gains. Unlike peer-to-peer practice, where feedback quality can vary with partner proficiency and consistency, the AI chatbot provided uniform, real-time corrections, thereby optimizing the learning process. Notably, the control group also demonstrated significant improvements, underscoring the enduring value of structured peer-to-peer practice in fostering speaking skills through social interaction and mutual support (Wang, 2023). This suggests that while AI chatbots may outperform traditional methods in certain aspects, peer practice remains a beneficial component of ELT, particularly for developing pragmatic and sociocultural competencies that AI may not fully replicate.

These findings are consistent with prior empirical research on AI chatbots in language learning. Studies such as Ye et al. (2022) and Duong and Suppasetseree (2024) reported significant improvements in speaking skills among learners using chatbot-based practice, while meta-analyses like Wu and Yu (2023) confirmed moderate to large effect sizes for chatbot interventions on speaking outcomes. More recent investigations, such as those by Ding and Yusof (2025) and Du and Daniel (2024), further corroborate these results by demonstrating AI's role in boosting engagement and proficiency across diverse contexts, reinforcing the scalability of chatbot interventions. Our results align with Ye et al. (2022), who found gains in pronunciation and fluency with smartphone chatbots, but extend this by demonstrating superiority over peer practice using standardized IELTS measures. In contrast, Duong and Suppasetseree (2024) focused on fluency in Vietnamese learners, showing similar improvements but without direct comparison to human methods, highlighting our study's contribution to comparative efficacy. More recent investigations, such as Chen et al. (2025) and Li et al. (2025), further corroborate these results by showing AI's role in boosting engagement and proficiency in diverse contexts, reinforcing the scalability of chatbot interventions; however, unlike Li et al.'s (2025) systematic review which aggregates broad AI tools, our controlled design isolates Elsa AI's targeted feedback, revealing larger gains than generalized approaches. Furthermore, the larger gains in the experimental group may be partly attributable to affective factors highlighted in the literature. The non-judgmental, low-stakes environment provided by the AI chatbot likely reduced speaking anxiety, encouraging more frequent risk-taking, extended output, and overall engagement—factors known to amplify proficiency development (Ballıdağ & Aydın, 2025; Zou et al., 2025).

The results suggest that AI-powered chatbots like Elsa AI can serve as a highly effective supplement to traditional language-teaching methods, particularly for enhancing speaking proficiency. Their ability to provide scalable, on-demand practice with consistent, adaptive feedback makes them especially valuable in contexts where access to human conversation partners—such as native speakers or qualified instructors—is limited. This is particularly relevant in EFL settings, where learners often face barriers to sustained oral practice. For educators and curriculum designers, these findings support integrating AI chatbots into ELT programs to expand speaking practice opportunities. Hybrid models combining AI-driven

practice with human-mediated interactions may offer a balanced approach, leveraging the consistency of chatbots while retaining the socio-cultural depth of human conversations. In addition, the demonstrated efficacy of AI tools could inform policy decisions in educational institutions, promoting investments in technology to bridge resource gaps and enhance global access to quality language education (Zhang, 2025). Teachers might incorporate chatbots for homework or supplementary practice, allowing classroom time to focus on advanced interpersonal skills.

In summary, this study provides empirical evidence that AI-powered chatbots, with their capacity for immediate, personalized feedback, can enhance English speaking proficiency more effectively than traditional peer-to-peer conversation practice. These findings contribute to the growing body of research on AI in language education and offer practical insights for integrating technology into ELT settings to meet the global demand for accessible, high-quality speaking practice.

Conclusion

This study provides robust evidence that regular practice with an AI-powered chatbot, Elsa AI, yields significantly greater gains in English-speaking proficiency than traditional peer-to-peer conversation practice during a 12-week intervention. Both groups demonstrated meaningful improvement, confirming that sustained, structured speaking activities promote language development. Crucially, however, the AI-chatbot group's larger mean gains, significant Time × Group interaction ($\eta_p^2 = 0.31$), and higher post-test scores underscore the pedagogical value of immediate, consistent, and adaptive corrective feedback in facilitating more efficient noticing and repair of interlanguage errors.

By operationalizing core SLA principles—the negotiated interaction of Long's Interaction Hypothesis and the learner output and noticing emphasized in Swain's Output Hypothesis—Elsa AI effectively scaffolds learner production in a low-anxiety environment, promoting risk-taking and sustained engagement. These findings suggest that AI chatbots can serve as scalable complements to conventional ELT methods, particularly in contexts where access to proficient interlocutors or trained instructors is limited. Educators and curriculum designers are encouraged to integrate chatbot-mediated practice into hybrid learning models, leveraging both automated feedback and human interaction to optimize speaking development.

Despite the promising results, several limitations must be acknowledged. First, the sample size was relatively small (n = 60), and participants were all university students in a specific EFL context (Ilam University), which may limit the generalizability of the findings to other populations or learning environments. Second, the intervention lasted only 12 weeks, and the long-term retention of proficiency gains was not assessed. Third, while the study controlled for practice time and task types, it did not account for potential differences in individual learner characteristics, such as motivation or prior technology experience, which could have influenced outcomes.

Future research should aim to replicate these findings with larger, more diverse samples, including learners from different proficiency levels, age groups, and linguistic backgrounds, to enhance generalizability. Longitudinal studies are needed to investigate the sustained impact of AI chatbot practice on speaking proficiency over extended periods and to determine whether initial advantages persist. Additionally, exploring the effectiveness of AI chatbots for other language skills, such as writing or listening, would provide a more comprehensive understanding of their pedagogical utility. Finally, research into optimal chatbot design features—such as the balance between automated feedback and human oversight—could inform the development of more effective AI tools for language learning.

Authors' Contributions

All authors participated equally in conducting the research and preparing the manuscript.

Declaration

The language model ChatGPT assisted in editing and elevating the scholarly writing in this work.

Transparency Statement

The corresponding author will provide datasets for research purposes upon a valid request.

Declaration of Interest

The authors state there are no competing interests to disclose.

References

Aliakbari, M., Barzan, P. and Sayyadi, M. (2025). Exploring the Impact of AI Chatbots on EFL Learners' Conversational Proficiency. *Interdisciplinary Research in English Language Communication*, *1*(2), 66-80. https://doi.org/10.30470/irelc.2025.2058800.1022

Atta, A. (2022). Theories and complexity of second language acquisition. S*candic Journal of Advanced Research and Reviews*, *2*(2), 92-99. https://doi.org/10.55966/sjarr.2022.2.2.0037

Ballıdağ, M. and Aydın, S. (2025). A comparison of the effects of AI-based chatbots and peer interactions on speaking anxiety among EFL learners. *Future in Educational Research*, *3*(2), 224-238. https://doi.org/10.1002/fer3.70005

Belda-Medina, J. and Calvo-Ferrer, J. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, *12*(17), 8427. https://doi.org/10.3390/app12178427

Bodnar, S., Vries, B., Cucchiarini, C., Strik, H., & Hout, R. (2011). Feedback in an ASR-based call system for l2 syntax: a feasibility study. Proc. *Speech and Language Technology in Education (SLaTE 2011)*, 113-116. https://doi.org/10.21437/slate.2011-19

Buckingham, L. and Alpaslan, R. (2017). Promoting speaking proficiency and willingness to communicate in Turkish young learners of English through asynchronous computer-mediated practice. *System*, *65*, 25-37. https://doi.org/10.1016/j.system.2016.12.016

Carroll, S., Swain, M., & Roberge, Y. (1992). The role of feedback in adult second language acquisition: error correction and morphological generalizations. *Applied Psycholinguistics*, *13*(2), 173-198. https://doi.org/10.1017/s0142716400005555

Daniels, P. (2015). Using web speech technology with language learning applications. *The Jalt Call Journal*, *11*(2), 177-187. https://doi.org/10.29140/jaltcall.v11n2.192

Ding, D., & Yusof, A. M. B. (2025). Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: a mixed methods study. *Humanities and Social Sciences Communications*, *12*(1), 1-16. https://doi.org/10.1057/s41599-025-05550-z

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education Artificial Intelligence*, *6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Duong, T. and Suppasetseree, S. (2024). The effects of an artificial intelligence voice chatbot on improving Vietnamese undergraduate students' English speaking skills. *International Journal of Learning Teaching and Educational Research*, *23*(3), 293-321. https://doi.org/10.26803/ijlter.23.3.15

Fryer, L., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: an experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, *75*, 461-468. https://doi.org/10.1016/j.chb.2017.05.045

Grab, M. Ö. (2025). Integrated AI chatbot practice: A pathway to improved ESL speaking skills. *Social Sciences & Humanities Open*, *12*, 101933. https://doi.org/10.1016/j.ssaho.2025.101933

Hidayat, S., Djatmika, D., & Ngadiso, N. (2022). Anxiety no more: the EFL students voice on the use FTF vs SCMC toward their speaking anxiety. *Veles Voice of English Language Education Society*, *6*(2), 392-406. https://doi.org/10.29408/veles.v6i2.6290

Hoang, N., Duong, N., & Le, D. (2023). Exploring chatbot AI in improving vocational students' English pronunciation. *Asiacall Online Journal*, *14*(2), 140-155. https://doi.org/10.54855/acoj.231429

Huang, W., Hew, K., & Fryer, L. (2021). 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/jcal.12610

Kemelbekova, Z., Degtyareva, X., Yessenaman, S., Ismailova, D., & Seidaliyeva, G. (2024). Ai in teaching English as a foreign language: effectiveness and prospects in Kazakh higher education. *Xlinguae*, *17*(1), 69-83. https://doi.org/10.18355/xl.2024.17.01.05

Kim, H., Kim, N., & Cha, Y. (2021). Is it beneficial to use AI chatbots to improve learners' speaking performance?. *The Journal of Asiatefl*, *18*(1), 161-178. https://doi.org/10.18823/asiatefl.2021.18.1.10.161

Koç, F. and Savaş, P. (2024). The use of artificially intelligent chatbots in English language learning: a systematic meta-synthesis study of articles published between 2010 and 2024. *Recall*, *37*(1), 4-21. https://doi.org/10.1017/s0958344024000168

Kohnke, L. (2022). A pedagogical chatbot: a supplemental language learning tool. *Relc Journal*, *54*(3), 828-838. https://doi.org/10.1177/00336882211067054

Kohnke, L., Moorhouse, B., & Zou, D. (2023). ChatGPT for language teaching and learning. *Relc Journal*, *54*(2), 537-550. https://doi.org/10.1177/00336882231162868

Kuhail, M., Alturki, N., Alramlawi, S., & Alhejori, K. (2022). Interacting with educational chatbots: a systematic review. *Education and Information Technologies*, *28*(1), 973-1018. https://doi.org/10.1007/s10639-022-11177-3

Lee, L. (2002). Enhancing learners' communication skills through synchronous electronic interaction and task-based instruction. *Foreign Language Annals*, *35*(1), 16-24. https://doi.org/10.1111/j.1944-9720.2002.tb01829.x

Lee, S., Lee, C., Lee, J., Noh, H., & Lee, G. (2010). Intention-based corrective feedback generation using context-aware model. In *Proceedings of the 2nd International Conference on Computer Supported Education* - Volume 2: CSEDU, SciTePress, pages 11-18. https://doi.org/10.5220/0002773200110018

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic Press.

Lyu, B., Lai, C., & Guo, J. (2024). Effectiveness of chatbots in improving language learning: a meta-analysis of comparative studies. *International Journal of Applied Linguistics*, *35*(2), 834-851. https://doi.org/10.1111/ijal.12668

Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences*, *12*(7), 3239. https://doi.org/10.3390/app12073239

Nguyen, L. and Pham, A. (2021). Using synchronous online discussion to develop EFL learners' productive skills: a case study. *The Journal of Asiatefl*, *18*(1), 179-207. https://doi.org/10.18823/asiatefl.2021.18.1.11.179

Nhạc, H. (2021). Corrective feedback and uptake patterns in English lessons at a tertiary institution. *English Education and Applied Linguistics Journal*, *4*(2), 92-105. https://doi.org/10.31980/eealjournal.v4i2.1510

Patra, B. and Kumar, M. (2020). Natural language processing in chatbots: a review. *Turkish Journal of Computer and Mathematics Education*, *11*(3), 2890-2894. https://doi.org/10.61841/turcomat.v11i3.14655

Pérez, J., Daradoumis, T., & Puig, J. (2020). Rediscovering the use of chatbots in education: a systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549-1565. https://doi.org/10.1002/cae.22326

Saito, K. (2021). Effects of corrective feedback on second language pronunciation development. In H. Nassaji & E. Kartchava (Eds.), *The Cambridge Handbook of Corrective Feedback in Second Language Learning and Teaching* (pp. 407–428). chapter, Cambridge: Cambridge University Press.

Shadiev, R. and Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. *Sustainability*, *12*(2), 524. https://doi.org/10.3390/su12020524

Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: a case study. *Expert Systems*, *38*(3), 1-15. https://doi.org/10.1111/exsy.12667

Singh, A. and Halim, H. (2023). Addressing challenges in language teaching in India: exploring the role of corrective feedback in enhancing learning. *Advanced Education*, *10*(22), 152-184. https://doi.org/10.20535/2410-8286.278042

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.

Ulfa, K. (2023). The transformative power of artificial intelligence (AI) to elevate English language learning. *Majalah Ilmiah Methoda*, *13*(3), 307-313. https://doi.org/10.46880/methoda.vol13no3.pp307-313

Vančová, H. (2023). Ai and AI-powered tools for pronunciation training. *Journal of Language and Cultural Education*, *11*(3), 12-24. https://doi.org/10.2478/jolace-2023-0022

Wanderley, L., Zhao, Z., & Epp, C. (2021). Negative language transfer in learner English: a new dataset. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3129-3142. https://doi.org/10.18653/v1/2021.naacl-main.251

Wang, C., Li, X., & Zou, B. (2025). Revisiting integrated model of technology acceptance among the generative AI-powered foreign language speaking practice: through the lens of positive psychology and intrinsic motivation. *European Journal of Education*, *60*(1). https://doi.org/10.1111/ejed.70054

Wang, Y. (2023). Peer corrective feedback: a review of its unique contributions to enhancing second language development. *International Journal of New Developments in Education*, *5*(8), 50-54. https://doi.org/10.25236/ijnde.2023.050809

Warschauer, M. (1997). Computer-mediated collaborative learning: theory and practice. *Modern Language Journal*, *81*(4), 470-481. https://doi.org/10.2307/328890

Wu, R. and Yu, Z. (2023). Do AI chatbots improve students learning outcomes? evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10-33. https://doi.org/10.1111/bjet.13334

Xu, X., Su, Y., Zhang, H., Zhang, Y., & Hao, S. (2024). *Beyond theory: a mixed-methods investigation of postgraduate engagement with ChatGPT for IELTS speaking*. PREPRINT. https://doi.org/10.21203/rs.3.rs-4503199/v1

Yang, H., Kim, H., Lee, J., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *Recall*, *34*(3), 327-343. https://doi.org/10.1017/s0958344022000039

Ye, Y., Deng, J., Liang, Q., & Liu, X. (2022). Using a smartphone-based chatbot in EFL learners' oral tasks. *International Journal of Mobile and Blended Learning*, *14*(1), 1-17. https://doi.org/10.4018/ijmbl.299405

Yu, H. and Nazir, S. (2021). Role of 5G and artificial intelligence for research and transformation of English situational teaching in higher studies. *Mobile Information Systems*, *2021*, 1-16. https://doi.org/10.1155/2021/3773414

Zhang, J. (2025). Integrating chatbot technology into English language learning to enhance student engagement and interactive communication skills. *Journal of Computational Methods in Sciences and Engineering*, *25*(3), 2288-2299. https://doi.org/10.1177/14727978241312992

Zhang, Z. (2024). Advancements and challenges in ai-driven language technologies: from natural language processing to language acquisition. *Applied and Computational Engineering*, *57*(1), 146-152. https://doi.org/10.54254/2755-2721/57/20241325

Zhou, Y. and Privitera, A. (2024). Subjective versus objective language proficiency measures in the investigation of bilingual effects on cognitive control. *International Journal of Bilingualism*, *29*(3), 549-565. https://doi.org/10.1177/13670069241229393

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability*, *15*(4), 1-19, 2872. https://doi.org/10.3390/su15042872

Zou, B., Xie, S., & Wang, C. (2025). Students' willingness to communicate (WTC) in using artificial intelligence (AI) technology in English-speaking practice. *International Journal of Information and Communication Technology Education*, *21*(1), 1-18. https://doi.org/10.4018/ijicte.375387