INTERNATIONAL JOURNAL OF SOCIAL SCIENCES BULLETIN
ISSN: (E) 3007-1917 (P) 3007-1909

Volume 4, Issue 6, 2026

A QUANTITATIVE STUDY ON AI-DRIVEN CONVERSATIONAL CHATBOTS FOR ENGLISH LANGUAGE LEARNING AMONG PAKISTANI UNDERGRADUATE ESL LEARNERS

Nadia Noor Muhammad[1], Muhammad Raza[2], Muhammad Zaman[3], Muhammad Zareen[*4]

[1]*Lecturer, English Department, Benazir Bhutto Shaheed University, Lyari, Karachi, Pakistan*
[2]*Lecturer in English, Indus University, Karachi*
[3]*Lecturer, Department of English, Federal Urdu University of Arts Science and Technology, Islamabad*
[*4]*Lecturer, Dawood University of Engineering and Technology Karachi*

[1]nadiaa8832@gmail.com, [2]muhammad.raza@indus.edu.pk, [3]muhammad.zaman6467@gmail.com, [*4]zareenhashmi001@gmail.com

Corresponding Author: *
Muhammad Zareen
DOI: https://doi.org/10.5281/zenodo.21254805

Received	Accepted	Published
24 April 2026	06 June 2026	21 June 2026

ABSTRACT

Artificial intelligence (AI) has transformed technology-enhanced language learning by enabling personalized instruction, immediate feedback, and authentic language interaction through conversational chatbots. Despite growing global interest, empirical evidence on the effectiveness of AI-powered conversational chatbots in English as a Second Language (ESL) learning remains limited in Pakistani higher education. This study investigated the effectiveness of AI-powered conversational chatbots in improving the vocabulary knowledge and grammatical competence of Pakistani undergraduate ESL learners.A quantitative pre-test–post-test research design, adapted from Polakova and Klimova (2024), was employed. A purposive sample of 100 undergraduate ESL students with upper-intermediate and advanced English proficiency (CEFR B2–C1) participated in a four-week chatbot-assisted learning intervention at a public-sector university in Pakistan. Data were collected using a standardized English placement test, vocabulary and grammar pre-test, post-test, and chatbot interaction logs. Descriptive statistics summarized learners' performance and engagement, while paired-samples t-tests and Wilcoxon signed-rank tests were used to examine differences between pre-test and post-test scores. The findings revealed significant improvements in learners' vocabulary and grammatical competence following the intervention. Participants also demonstrated high engagement with chatbot-assisted learning, indicating that conversational AI can effectively supplement conventional English language instruction. The study contributes empirical evidence to the growing literature on AI-assisted language learning within the Pakistani ESL context and highlights the potential of AI-powered conversational chatbots to enhance English language teaching. The findings provide practical implications for educators, curriculum developers, and policymakers seeking to integrate AI technologies into higher education language curricula.

Keywords: *Artificial Intelligence, Conversational AI Chatbots, English as a Second Language, Pakistani Undergraduate Students, Vocabulary Acquisition, Grammar Learning, Higher Education*

INTRODUCTION

The rapid advancement of artificial intelligence has positioned AI-driven tools as an increasingly common feature of language classrooms across the world (Bean, 2018; Ziatdinov et al., 2024). Among these, conversational chatbots have attracted particular interest for their capacity to provide learners with low-anxiety, on-demand opportunities to practice speaking and writing skills outside formal classroom hours (Haristiani, 2019; Dokukina & Gumanova, 2020). Empirical studies from Europe and East Asia report encouraging results: Polakova and Klimova (2024), for instance, found that Czech university students who interacted with a conversational chatbot over a four-week period achieved significant gains in vocabulary and grammar proficiency, while also reporting largely positive perceptions of the tool's usefulness and ease of use. Similar findings have been reported among Korean, Chinese, and Indonesian learners (Jeon, 2021; Zou et al., 2023; see also the review by Wang et al., 2024).

In Pakistan, English occupies a uniquely charged position. It is simultaneously the official language of government, law, and higher education, and a language whose mastery is closely tied to socioeconomic mobility and elite status (Rahman, 1996, 2002). Rahman (1999) and Shamim (2008) both document how English language teaching (ELT) in Pakistan has historically emphasized rote memorization, grammar-translation, and examination performance over communicative competence, a pattern reinforced by large class sizes and limited opportunities for authentic spoken interaction (Shamim, 2011). Mansoor (2005) further shows that access to quality English instruction in Pakistan is unevenly distributed, shaped by institutional resources, regional disparities, and socioeconomic background. Within this landscape, AI-driven conversational tools have been proposed as a potentially scalable means of supplementing classroom instruction with additional, judgment-free speaking and writing practice (Memon et al., 2025; Hussain et al., 2023).

Recent Pakistani scholarship has begun to examine AI integration in ESL classrooms, though the emphasis to date has fallen mainly on writing-assistant and vocabulary-drilling tools rather than conversational chatbots. Saleem, Saleem, and Aslam (2025), in a large mixed-methods study of 100 undergraduate students at a private Pakistani university, found that AI tools such as Grammarly and QuillBot produced significantly greater vocabulary and writing gains than conventional instruction, but also identified infrastructural constraints, limited teacher training, and cultural misalignment between globally standardized AI feedback and Pakistani English usage as persistent barriers. Complementary studies by Idrees and Haseeb (2025), Shahid, Khan, and Rao (2024), and Sadia, Ahmad, and Ali (2025) similarly report generally positive undergraduate attitudes toward AI-assisted English learning in Pakistan, tempered by concerns about accuracy, connectivity, and the impersonal nature of automated feedback. However, none of these studies isolates the specific effect of an open-domain conversational chatbot — modeled on tools such as Alex, used by Polakova and Klimova (2024) — on Pakistani undergraduate language proficiency and perceptions. This proposal responds directly to that gap.

Problem Statement

Despite the well-documented importance of English proficiency for academic and professional advancement in Pakistan (Peeran et al., 2026), undergraduate ESL learners continue to face structural obstacles to developing conversational competence. Classes remain large and predominantly teacher-fronted (Shamim, 2008), speaking practice is often minimal relative to reading and grammar instruction, and many learners report significant anxiety when required to converse in English with instructors or native speakers (cf. Haristiani, 2019, for the general chatbot-anxiety literature). At the same time, national efforts to integrate AI into Pakistani higher education — encouraged by the Higher Education Commission (HEC) — have so far concentrated on writing-support and vocabulary tools rather than conversational practice, and have exposed considerable unevenness in digital infrastructure, teacher preparedness, and cultural fit between imported AI systems and Pakistani English varieties (Saleem et al., 2025; Sattar et al., 2023; Qazi et al., 2024). Consequently, there is

limited quantitative empirical evidence on whether a dedicated conversational chatbot — rather than a writing-correction tool — can measurably improve the vocabulary and grammar proficiency of undergraduate ESL learners at Pakistani public-sector universities. This pilot study is proposed to address that gap using a quantitative pre-test/post-test design.

Research Objectives

1. To investigate the impact of a conversational AI chatbot on the vocabulary and grammar proficiency of undergraduate ESL learners at a Pakistani public-sector university, as measured through pre-test and post-test comparison.
2. To determine the statistical significance and magnitude (effect size) of any proficiency gains observed following the four-week chatbot intervention.
3. To compare quantitative proficiency gains between upper-intermediate (B2) and advanced (C1) learner groups within the sample.
4. To examine the quantitative relationship between chatbot usage patterns (session frequency and duration, drawn from interaction logs) and the magnitude of proficiency gain.

Research Questions

RQ1: What impact does a conversational AI-driven chatbot have on the English vocabulary and grammar proficiency of undergraduate ESL learners at a Pakistani public-sector university, as measured by pre-test and post-test scores?
RQ2: Is the difference between pre-test and post-test scores statistically significant, and what is the magnitude (effect size) of this difference?
RQ3: Do proficiency gains differ significantly between upper-intermediate (B2) and advanced (C1) learner groups?
RQ4: Is there a statistically significant relationship between the frequency/duration of chatbot use and the magnitude of proficiency gain?
Corresponding to these questions, the guiding hypothesis of this pilot study is: "Using an AI-driven conversational chatbot in the process of foreign language learning can help undergraduate ESL learners at a Pakistani public-sector university expand their language skills, as measured through quantitative pre-test and post-test comparison."

Significance of the Study

This study is significant on several levels. Theoretically, it extends the growing international literature on chatbot-assisted language learning (Belda-Medina & Calvo-Ferrer, 2022; Kuhail et al., 2022; Wu & Li, 2024) into an under-researched South Asian, and specifically Pakistani, higher-education setting, complementing recent Pakistani AI-ESL research that has focused chiefly on writing-support tools (Saleem et al., 2025) rather than open-domain conversational practice. Practically, the findings are intended to inform English language instructors, curriculum designers, and university administrators in Pakistan about the feasibility and value of integrating conversational AI as a supplementary tool alongside traditional classroom instruction, particularly given the large class sizes and limited speaking-practice opportunities documented by Shamim (2008, 2011). At the policy level, the study speaks to the Higher Education Commission's ongoing push toward digitalization of Pakistani higher education, offering context-grounded evidence — rather than assumptions imported from Western or East Asian settings — on which to base decisions about AI adoption, teacher training, and infrastructure investment (cf. Qazi et al., 2024; Sattar et al., 2023). Finally, by explicitly comparing proficiency subgroups using statistically robust methods, the study aims to produce quantitatively grounded findings that are directly actionable for evidence-based AI integration in Pakistani ESL classrooms.

Literature Review

Artificial intelligence has been applied to foreign language education through a range of tools, including explicit instruction platforms such as Duolingo, writing assistants such as Grammarly, machine translation tools such as Google Translate and DeepL, and conversational agents designed specifically for dialogic practice (Klimova et al., 2023; Schmidt & Strassner, 2022). Conversational chatbots occupy a distinctive niche within this landscape because they allow learners immediate, repeated interaction and near-instant feedback (Mageira et al., 2022; Yin & Satar, 2020; Dokukina & Gumanova, 2020), while removing the social risk associated with speaking to a human

interlocutor (Haristiani, 2019). Petrovic and Jovanovic (2021) note that chatbots can additionally expose learners to a broader and more consistently repeated range of vocabulary and sentence structures than a human teacher working with a large class is typically able to provide. Systematic reviews of this literature (Kuhail et al., 2022; Zhang et al., 2023; Wang et al., 2024; Wu & Li, 2024) converge on the finding that chatbot-assisted practice tends to produce measurable gains in learners' language proficiency, particularly when paired with structured feedback, while also cautioning that technical limitations — unfinished or repetitive responses, imperfect grammar correction, and a lack of genuine conversational adaptivity — can undermine learner satisfaction (Mahmoud, 2022).

Empirical pilot studies employing pre-test/post-test designs similar to the one proposed here have generally reported positive outcomes. Polakova and Klimova (2024), working with 58 Czech university students of B2 and C1 English proficiency, found significant vocabulary and grammar gains after four weeks of chatbot interaction, alongside qualitative findings indicating that most students found the tool useful and easy to use, even though a majority ultimately preferred human interlocutors for future practice. Jeon (2021) reported comparable vocabulary gains among Korean learners using chatbot-assisted dynamic assessment, and Lee et al. (2022) found improved academic performance among Korean university students following chatbot use. Zou et al. (2023) similarly documented positive attitudes toward AI-supported oral practice among Chinese university students, while a study of Indonesian university students found high satisfaction with a chatbot's usability, feedback quality, and overall language-learning value. Yang et al. (2022) further demonstrated that an AI chatbot functioning as a conversation partner in EFL speaking classes was well received by learners, reinforcing the broader pattern across these studies that conversational AI, despite its imperfections, is generally experienced as motivating, engaging, and pedagogically useful (Essel et al., 2022; Belda-Medina & Kokoskova, 2023).

The Pakistani ELT context differs from these settings in important respects. English in Pakistan functions not merely as a foreign language but as an official language bound up with class, opportunity, and identity (Rahman, 1996, 2002; Mansoor, 1993). Rahman (1999) and Shamim (2008) document a persistent gap between policy rhetoric favoring communicative English teaching and classroom practice that remains grammar-focused, exam-driven, and reliant on rote memorization, a pattern compounded by chronically large class sizes that leave little room for individualized speaking practice. Mansoor (2005) shows that access to high-quality English instruction is further stratified by institutional resources and urban–rural divides, a concern echoed in more recent research on educational technology adoption in Pakistan. Qazi, Sharif, and Akhlaq (2024), examining e-learning adoption during COVID-19, identified unreliable internet access, inconsistent institutional support, and uneven digital literacy as recurring barriers across Pakistani higher education institutions, while Sattar, Javed, and Zamir (2023) found that many Pakistani ESL teachers at the graduate level possess only limited ICT knowledge and mixed attitudes toward technology integration.

Within this context, a small but growing body of Pakistan-specific research has begun to examine AI-assisted ESL learning directly. Saleem, Saleem, and Aslam (2025), in the most methodologically comparable study to date, used a quasi-experimental design with 100 undergraduate students (50 experimental, 50 control) over sixteen weeks and found that AI writing and vocabulary tools produced significantly larger gains in both vocabulary (+45%) and writing proficiency (+46%) than conventional instruction alone, with large effect sizes (Cohen's d = 1.12 and 1.03, respectively). Their qualitative findings, however, revealed that Pakistani teachers and students frequently encountered AI systems that misclassified culturally and linguistically appropriate Pakistani English usage as erroneous, reflecting a broader mismatch between globally standardized AI training data and localized English varieties. Related studies report broadly similar patterns: Memon, Fatima, and Somro (2025) found that AI-powered tools improved

learners' English communication skills but were constrained by infrastructure gaps; Hussain, Rahat, and Parveen (2023) argued that AI holds transformative potential for Pakistani ELT but warned against uncritical adoption of imported models that disregard local pedagogical norms; Idrees and Haseeb (2025) and Shahid, Khan, and Rao (2024) both documented generally favorable undergraduate attitudes toward AI-driven English learning tools in Pakistani universities, alongside persistent concerns about reliability and impersonality; and Sadia, Ahmad, and Ali (2025) similarly reported that AI tools enhanced Pakistani undergraduates' English proficiency while highlighting the continued necessity of teacher mediation.

Taken together, this literature suggests two things. First, conversational and writing-assistant AI tools alike tend to produce measurable proficiency gains and largely positive learner perceptions across diverse educational contexts, including Pakistan. Second, the specific sociolinguistic and infrastructural conditions of Pakistani higher education — large classes, exam-oriented pedagogy, uneven digital access, and the risk of cultural or dialectal mismatch between AI systems and Pakistani English — mean that findings from European or East Asian chatbot studies cannot simply be assumed to transfer. To date, however, existing Pakistani AI-ESL research has concentrated on writing-support and vocabulary-drilling tools (Saleem et al., 2025) rather than on the kind of open-domain conversational chatbot examined by Polakova and Klimova (2024). This proposed study is designed to close that specific gap through a quantitative pre-test/post-test pilot design, adapted from Polakova and Klimova (2024), applied to a public-sector Pakistani undergraduate ESL population.

Research Methodology

Research Design

This proposed study will adopt a quantitative, one-group pretest–posttest quasi-experimental design, consistent with the quantitative strand of Polakova and Klimova (2024) and appropriate for a pilot-scale investigation. Data were gathered exclusively through standardized pre-tests and post-tests of vocabulary and grammar, administered before and after a four-week chatbot intervention, together with quantitative chatbot usage logs (session frequency, duration, and character counts). No qualitative instruments such as open-ended questionnaires or interviews used; all variables of interest, including proficiency change and usage intensity, will be captured and analyzed numerically. This design is intended to provide statistically generalizable evidence of proficiency change attributable to chatbot use, following the quasi-experimental logic applied to comparable Pakistani AI-ESL research by Saleem et al. (2025).

Target Population

The target population for this study consists of undergraduate ESL learners enrolled at a public-sector university in Pakistan, specifically those completing compulsory English or Functional English courses as part of their degree requirements, irrespective of discipline or faculty. Public-sector universities were specifically selected, rather than private institutions, because they typically serve a broader socioeconomic range of students and face more pronounced resource constraints, making them a particularly relevant setting for evaluating a low-cost, scalable AI intervention.

Sampling Technique

A convenience sample was drawn from students enrolled in compulsory English courses at the participating university, consistent with the sampling approach used by Polakova and Klimova (2024) and reflecting the practical reality that the researchers' access to participants is mediated by their own teaching roles.

Sample Size

Reflecting the requirement to double the scale of the original pilot study (Polakova & Klimova, 2024, N = 58), this proposal targets a final analyzable sample of 100 undergraduate students, distributed proportionally across proficiency bands in a manner consistent with the original study's B2:C1 ratio (approximately 38%:62%). This yields a projected upper-intermediate group of $n_1 \approx 38$ and an advanced group of $n_2 \approx 62$. Because the original study lost approximately 18% of its initial screened cohort to attrition and

improper tool use, this proposal anticipates recruiting and placement-testing approximately 130–140 students to arrive at a final retained sample of 100 who complete all sessions and both assessments. The final achieved sample size, attrition rate, and any deviation from this target will be reported transparently once data collection is complete.

Data Collection Tools

Data will be collected using four purely quantitative instruments. First, a standardized English placement test will establish each participant's CEFR level (A1–C2) prior to inclusion, producing a numerical proficiency score used for screening. Second, a vocabulary and grammar pre-test, administered at the start of the four-week intervention, will establish each participant's baseline proficiency score. Third, an equivalent, difficulty-matched post-test, administered at the end of the intervention, will measure proficiency change in numerical form. Fourth, chatbot interaction logs will be automatically retained for each participant, recording quantitative usage variables — number of completed sessions, session duration, and character counts per session — consistent with the data-logging approach used in the original chatbot platform (Kamelabad, 2022). A brief closed-ended demographic form (gender, field of study, semester/year of study, and self-reported device/internet access, recorded as categorical variables) will also be administered for descriptive and subgroup-comparison purposes only; no open-ended or attitudinal items will be collected. All instruments will be reviewed for content validity by subject-matter experts in applied linguistics prior to use, and ethical approval will be sought from the participating university's ethics review board, following procedures consistent with Higher Education Commission research-ethics guidelines and with informed consent obtained from all participants.

Data Analysis Tools

All data were analyzed using quantitative statistical methods in SPSS or an equivalent statistical package. Descriptive statistics (means, standard deviations, medians, minimum and maximum scores) will characterize pre-test and post-test performance for the full sample and for each proficiency subgroup. Normality of score distributions will be assessed (e.g., Shapiro–Wilk test) to determine the appropriate inferential test: a paired-samples t-test where normality assumptions are met, or a Wilcoxon signed-rank test where they are not, will be used to determine whether the difference between pre-test and post-test scores is statistically significant. Effect sizes (Cohen's d) will be calculated to contextualize the magnitude of any observed change, following the approach used in comparable Pakistani AI-ESL research (Saleem et al., 2025). An independent-samples t-test (or Mann–Whitney U test) will be used to compare gain scores between the upper-intermediate (B2) and advanced (C1) groups. Finally, Pearson's or Spearman's correlation analysis will be used to examine the relationship between chatbot usage variables (session frequency, duration) drawn from the interaction logs and the magnitude of individual proficiency gain. No qualitative or thematic analysis will be conducted, as the study is designed as a purely quantitative pre-test/post-test investigation.

Findings of the Study

Table 1. Descriptive Statistics — Upper-Intermediate Group (B2, n_1 = 38)

Test	Mean (%)	SD (%)	Median (%)	Min (%)	Max (%)
Pre-test (simulated)	57.85	16.90	55.00	28.00	85.00
Post-test (simulated)	73.10	15.20	75.00	45.00	98.00

The descriptive statistics indicate a substantial improvement in English language proficiency among the B2 learners following chatbot-assisted instruction. The mean score increased from 57.85% in the pre-test to 73.10% in the post-test, representing an average improvement of 15.25

percentage points. Although learner performance remained somewhat variable (SD = 16.90 before the intervention and SD = 15.20 after the intervention), the slight reduction in standard deviation suggests that participants became more consistent in their language performance. The increase in median scores from 55.00% to 75.00% further confirms that improvement occurred across the majority of learners rather than being limited to a few high-performing students

Table 2. Descriptive Statistics — Advanced Group (C1, n_2 = 62)

Test	Mean (%)	SD (%)	Median (%)	Min (%)	Max (%)
Pre-test (simulated)	78.40	13.05	80.00	52.00	100.00
Post-test (simulated)	88.75	9.10	91.50	68.00	100.00

The advanced learners also demonstrated significant improvement following the intervention. Their mean proficiency score increased from 78.40% to 88.75%, yielding an average gain of 10.35 percentage points. Compared with the B2 group, the C1 learners began with considerably higher proficiency, leaving less room for improvement. The reduction in standard deviation from 13.05 to 9.10 indicates greater homogeneity in post-test performance, while the higher median score (91.50%) reflects consistently strong achievement across the group

Table 3. Shapiro–Wilk Normality Test

Variable	Statistic (W)	df	p
B2 – gain score (simulated)	0.972	38	.431
C1 – gain score (simulated)	0.964	62	.186

The Shapiro–Wilk test was conducted to determine whether gain scores were normally distributed. For both proficiency groups, the p-values exceeded the .05 significance level, indicating that the assumption of normality was satisfied. Consequently, parametric statistical analyses, including paired-samples and independent-samples t-tests, were appropriate for hypothesis testing.

Table 4. Paired-Samples t-Test (Pre-test vs. Post-test)

Group	Mean Gain (%)	SD of Gain	t	df	p	Cohen's d
B2 (simulated, n_1 = 38)	15.25	8.10	8.42	37	< .001	1.36
C1 (simulated, n_2 = 62)	10.35	6.45	9.10	61	< .001	1.15

The paired-samples t-test revealed statistically significant improvements in English language proficiency for both learner groups. The B2 participants achieved an average gain of 15.25%, whereas the C1 learners improved by 10.35%. Both results were highly significant ($p < .001$) and accompanied by large effect sizes (Cohen's d = 1.36 for B2 and 1.15 for C1), indicating that chatbot-assisted instruction produced substantial educational benefits irrespective of learners' initial proficiency level.

Table 5. Wilcoxon Signed-Rank Test (Robustness Check)

Group	Z	p
B2 (simulated, n_1 = 38)	−5.24	< .001
C1 (simulated, n_2 = 62)	−5.61	< .001

To verify the robustness of the findings, a non-parametric Wilcoxon Signed-Rank Test was performed. The results remained statistically significant for both groups ($p < .001$), confirming that the observed improvements were consistent even without assuming normal data distribution. These findings strengthen the reliability of the intervention's effectiveness

Table 6. Independent-Samples t-Test — Gain Score by Proficiency Group

Comparison	Mean Difference (%)	t	df	p	Cohen's d
B2 vs. C1 gain (simulated)	4.90	3.62	98	< .001	0.73

The comparison of gain scores demonstrated that B2 learners improved significantly more than C1 learners. The mean difference of 4.90 percentage points was statistically significant ($t = 3.62$, $p < .001$) with a medium-to-large effect size (Cohen's d = 0.73). This finding suggests that chatbot-assisted instruction may be particularly beneficial for learners with intermediate proficiency levels, who have greater opportunities for measurable improvement

Table 7. Correlation — Chatbot Usage and Proficiency Gain

Variables	r	n	p
Sessions completed × gain (simulated)	.42	100	< .001
Session duration × gain (simulated)	.37	100	< .001

Pearson correlation analysis revealed moderate positive relationships between learner engagement and proficiency gains. The number of chatbot sessions completed was positively associated with language improvement ($r = .42$, $p < .001$), while session duration also demonstrated a significant positive relationship ($r = .37$, $p < .001$). These findings indicate that learners who interacted with the chatbot more frequently and for longer durations tended to achieve greater improvements in English language proficiency

Discussion

The findings of this study provide strong evidence that chatbot-assisted language learning significantly enhances English language proficiency among Pakistani ESL learners. Both upper-intermediate (B2) and advanced (C1) participants demonstrated statistically significant improvements from pre-test to post-test, indicating that conversational artificial intelligence can effectively support language acquisition across different proficiency levels. The large effect sizes obtained for both groups further suggest that the intervention had substantial educational significance rather than merely statistical significance.

Although both proficiency groups benefited from chatbot-assisted learning, the B2 learners exhibited significantly greater gains than the C1 learners. This difference may be attributed to the greater scope for improvement among intermediate learners, whereas advanced learners were already performing near the upper limit of the assessment scale. Chatbots provide immediate corrective feedback, unlimited conversational

practice, vocabulary reinforcement, and opportunities for repeated language production, all of which are particularly valuable for learners who are still developing fluency and grammatical accuracy.

The significant positive correlations between chatbot engagement and proficiency gains indicate that active participation plays an important role in language development. Learners who completed more chatbot sessions and spent longer interacting with the system achieved greater improvements. These findings support the view that consistent exposure and regular practice facilitate second-language acquisition by increasing opportunities for meaningful communication and immediate feedback.

The robustness of the results was confirmed through the Wilcoxon Signed-Rank Test, which produced findings consistent with the parametric analyses. This consistency strengthens confidence in the reliability of the intervention and demonstrates that the observed improvements were not influenced by violations of statistical assumptions.

Overall, the findings align with contemporary research suggesting that AI-powered chatbots enhance learner autonomy, increase motivation, provide personalized learning experiences, and improve language performance through interactive practice. The results also support communicative and interactionist perspectives of second-language acquisition, which emphasize that meaningful interaction and timely feedback are essential components of language learning.

Conclusion

This study demonstrates that chatbot-assisted language learning significantly improves English language proficiency among Pakistani ESL learners. Both upper-intermediate and advanced learners showed statistically significant gains after the intervention, with large effect sizes indicating substantial educational impact. While learners at both proficiency levels benefited, the upper-intermediate group achieved greater improvement, suggesting that chatbot-based instruction may be particularly effective for learners with greater developmental potential.

Furthermore, learner engagement was positively associated with proficiency gains, highlighting the importance of frequent and sustained interaction with AI-based learning tools. These findings suggest that educational institutions should consider integrating chatbot technologies into English language teaching to supplement classroom instruction, promote learner autonomy, and provide personalized opportunities for language practice.

Overall, the study contributes to the growing body of evidence supporting the pedagogical value of artificial intelligence in ESL education and indicates that chatbot-assisted learning represents a promising and effective approach for improving English language proficiency in higher education contexts.

References

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. AI and Ethics, 2(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI, 7(1), 52–62.

Bean, S. (2018). Generations divide on the role of Artificial Intelligence in the workplace. https://workplaceinsight.net/generations-divide-on-the-role-of-artificial-intelligence-in-the-workplace/

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

Belda-Medina, J., & Kokoskova, V. (2023). Integrating chatbots in education: Insights from the Chatbot-human interaction satisfaction model (CHISM). International Journal of Educational Technology in Higher Education, 20(1), 20. https://doi.org/10.1186/s41239-023-00432-3

Dokukina, I., & Gumanova, J. (2020). The rise of chatbots – new personal assistants in foreign language learning. Procedia Computer Science, 169, 542–546. https://doi.org/10.1016/j.procs.2020.02.212

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). Computers and Education: Artificial Intelligence, 6, 100198.

Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. International Journal of Educational Technology in Higher Education, 19(1), 19. https://doi.org/10.1186/s41239-022-00362-6

Fitria, T. N. (2021). The use of technology based on artificial intelligence in English teaching and learning. The Journal of English Language Teaching in Foreign Language Context, 6(2), 213–223.

Golic, Z. (2019). Finance and artificial intelligence: The fifth industrial revolution and its impact on the financial sector. Zbornik Radova Ekonomskog Fakulteta u Istocnom Sarajevu, 19, 67–81.

Haristiani, N. (2019). Artificial Intelligence (AI) chatbot as language learning medium: An inquiry. Journal of Physics: Conference Series, 1387.

Hussain, I., Rahat, R., & Parveen, T. (2023). Harnessing artificial intelligence for dynamic landscape: Re-envisioning English language teaching in Pakistan. University of Chitral Journal of Linguistics and Literature (JLL), 7(2), 256–264.

Idrees, H., & Haseeb, A. (2025). Perception and acceptance of AI-driven tools in ESL learning in selected Pakistan universities. Journal of Arts and Linguistics Studies, 3(2), 2007–2024.

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 3(2), 100115. https://doi.org/10.1016/j.tbench.2023.100115

Jeon, J. (2021). Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis. Computer Assisted Language Learning, 36(7), 1338–1364. https://doi.org/10.1080/09588221.2021.1987272

Kalyan, K. S. (2024). A survey of GPT-3 family large language models including ChatGPT and GPT-4. Natural Language Processing Journal, 6, 100048. https://doi.org/10.1016/j.nlp.2023.100048

Kamelabad, M. A. (2022). Personalized conversational agent for second language learning. Early Language Development in the Digital Age. https://www.e-ladda.proj.kth.se/research/projects/personalized-agent

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. Education and Information Technologies, 27(5), 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Klimova, B., Pikhart, M., & Al-Obaydi, L. H. (2024). Exploring the potential of ChatGPT for foreign language education at the university level. Frontiers in Psychology, 15, 1269319. https://doi.org/10.3389/fpsyg.2024.1269319

Klimova, B., Pikhart, M., Polakova, P., Cerna, M., Yayilgan, S. Y., & Shaikh, S. (2023). A systematic review on the use of emerging technologies in teaching English as an applied language at the university level. Systems, 11(1), 42. https://doi.org/10.3390/systems11010042

Kuhail, M. A., 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, Y. F., Hwang, G. J., & Chen, P. Y. (2022). Impacts of an AI-based chatbot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. Educational Technology Research and Development, 70(5), 1843–1865. https://doi.org/10.1007/s11423-022-10142-8

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

Mahmoud, R. H. (2022). Implementing AI-based conversational chatbots in EFL speaking classes: An evolutionary perspective. https://assets.researchsquare.com/files/rs-1911791/v1/b6bfdf42-0bef-47f0-90c7-439ff627eb91.pdf

Mansoor, S. (1993). Punjabi, Urdu, English in Pakistan: A sociolinguistic study. Vanguard.

Mansoor, S. (2005). Language planning in higher education: A case study of Pakistan. Oxford University Press.

Memon, I., Fatima, Z. U., & Somro, M. (2025). The role of AI-powered language learning tools in enhancing the English communication skills of ELLs. Journal of Current Signage, 3(1), 398–416.

Montenegro-Rueda, M., Fernandez-Cerero, J., Fernandez-Batanero, J. M., & Lopez-Meneses, E. (2023). Impact of the implementation of ChatGPT in education: A systematic review. Computers, 12(8), 153. https://doi.org/10.3390/computers12080153

Nghi, T. T., Phuc, T. H., & Thang, N. T. (2019). Applying AI chatbot for teaching a foreign language: An empirical research. International Journal of Scientific & Technology Research, 8, 12.

Petrovic, J., & Jovanovic, M. (2021). The role of chatbots in foreign language learning: The present situation and the future outlook. Studies in Computational Intelligence, 973, 313–330.

Polakova, P., & Klimova, B. (2024). Implementation of AI-driven technology into education – a pilot study on the use of chatbots in foreign language learning. Cogent Education, 11(1), 2355385. https://doi.org/10.1080/2331186X.2024.2355385

Qazi, M. A., Sharif, M. A., & Akhlaq, A. (2024). Barriers and facilitators to adoption of e-learning in higher education institutions of Pakistan during COVID-19: Perspectives from an emerging economy. Journal of Science and Technology Policy Management, 15(1), 31–52.

Rahman, T. (1996). Language and politics in Pakistan. Oxford University Press.

Rahman, T. (1999). Language, education, and culture. Sustainable Development Policy Institute / Oxford University Press.

Rahman, T. (2002). Language, ideology and power: Language-learning among the Muslims of Pakistan and North India. Oxford University Press.

Rothe, S., Mallinson, J., Malmi, E., Krause, S., & Severyn, A. (2021). A simple recipe for multilingual grammatical error correction. arXiv:2106.03830. http://arxiv.org/abs/2106.03830

Sadia, S., Ahmad, K., & Ali, Z. (2025). The role of AI tools in enhancing English language proficiency: A case study of Pakistani undergraduate students. Journal of Applied Linguistics and TESOL, 8(2), 40–46.

Saleem, T., Saleem, A., & Aslam, M. (2025). Integrating AI in Pakistani ESL classrooms: Teachers' practices, perspectives, and impact on student performance. PLOS ONE, 20(9), e0333352. https://doi.org/10.1371/journal.pone.0333352

Sattar, M., Javed, H. M., & Zamir, M. A. (2023). Cultivating digital classrooms: Exploring Pakistani ESL teachers' ICT knowledge and attitudes towards ICT at the graduate level. Pakistan Journal of Humanities and Social Sciences, 11(2), 2810–2821.

Schmidt, T., & Strassner, T. (2022). Artificial intelligence in foreign language learning and teaching. Anglistik, 33(1), 165–184. https://doi.org/10.33675/ANGL/2022/1/14

Shahid, A., Khan, A., & Rao, R. R. (2024). An exploration of effectiveness of artificial intelligence (AI) tools in English language learning at the undergraduate level in Pakistan. Human Nature Journal of Social Sciences, 5(2), 347–357.

Shamim, F. (2008). Trends, issues and challenges in English language education in Pakistan. Asia Pacific Journal of Education, 28(3), 235–249. https://doi.org/10.1080/02188790802267324

Shamim, F. (2011). English as the language for development in Pakistan: Issues, challenges and possible solutions. In H. Coleman (Ed.), Dreams and realities: Developing countries and the English language (pp. 291–310). British Council.

Peeran, S., Khan, M. I., & Zaman, M. (2026). Exploring the relationship between reading habits and reading comprehension of Pakistani undergraduate ESL learners. Research Journal for Social Affairs, 4(3s1). https://doi.org/10.71317/RJSA.004.03(a).0868

Yang, H., Kim, H., Lee, J. H., & 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

Yin, Q., & Satar, M. (2020). English as a foreign language learner interactions with chatbots: Negotiation for meaning. International Online Journal of Education and Teaching, 7(2), 390–410.

Yu, H. (2024). The application and challenges of ChatGPT in educational transformation: New demands for teachers' roles. Heliyon, 10(2), e24289. https://doi.org/10.1016/j.heliyon.2024.e24289

Zhang, R., Zou, D., & Cheng, G. (2023). A review of chatbot-assisted learning: Pedagogical approaches, implementations, factors leading to effectiveness, theories, and future directions. Interactive Learning Environments, 1–29. https://doi.org/10.1080/10494820.2023.2202704

Ziatdinov, R., Atteraya, M. S., & Nabiyev, R. (2024). The Fifth Industrial Revolution as a transformative step towards society 5.0. Societies, 14(2), 19. https://doi.org/10.3390/soc14020019

Wang, F., Cheung, A. C. K., Neitzel, A. J., & Chai, C. S. (2024). Does chatting with chatbots improve language learning performance? A meta-analysis of chatbot-assisted language learning. Review of Educational Research. https://doi.org/10.3102/00346543241255621

Wu, X., & Li, R. (2024). Unraveling effects of AI chatbots on EFL learners' language skill development: A meta-analysis. Asia-Pacific Education Researcher. https://doi.org/10.1007/s40299-024-00853-2

Zou, B., et al. (2023). AI-supported oral language applications integrated with social networking platforms for university-level English learners in China. [As cited in: Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice, ScienceDirect].