

THE ROLE OF AI TOOLS IN ENHANCING ENGLISH LANGUAGE PROFICIENCY: A CASE STUDY OF PAKISTANI UNDERGRADUATE STUDENTS

Saba Sadia

Assistant Professor, Department of English, The Govt. Sadiq College Women University,
Bahawalpur

Khursheed Ahmad

Department of English, University of Swabi

khursheedlinguist@gmail.com

Zahid Ali

Associate Professor, Department of English, GC University Hyderabad

Abstract

This mixed-methods study examines the efficacy of AI-assisted feedback versus traditional teacher-led approaches in Pakistani undergraduate English language classrooms. Data from 120 students across two universities revealed AI tools such as Grammarly significantly enhanced grammatical accuracy and vocabulary retention, while teacher feedback excelled in nurturing cultural relevance and contextual understanding. Qualitative insights highlighted AI's immediacy but criticized its lack of explanatory depth, whereas teachers provided scaffolded, culturally attuned guidance aligned with Vygotsky's sociocultural theory (1978). Students valued AI for reducing but expressed emotional disconnection, contrasting with teacher feedback's motivational rapport. The study identifies infrastructural barriers exacerbating inequities in AI adoption and proposes a blended model integrating AI's efficiency with teacher-led contextualization. Recommendations include localizing AI tools with Urdu-English corpora to address cultural biases (Ahmed & Malik, 2023) and promoting digital access. Findings advocate hybrid frameworks to balance technical accuracy with socioemotional and cultural responsiveness in resource-constrained settings, aligning with global evidence on blended learning efficacy (Li et al., 2023).

Keywords: AI feedback, traditional pedagogy, blended learning, cultural relevance, educational equity.

1. Introduction

English language is considered as key to success in academia, Civil and Military services as well in Industry in Pakistan, where English serves as a lingua franca for education and governance despite Urdu being the national language. However, in State-run institutions, overcrowded classrooms, scarcity of trained teacher, and outdated pedagogical methods hampers effective language teaching and learning (Rahman, 2019). Feedback, being a backbone of language teaching and learning, is often delayed due to these constraints. Recent development in AI-driven tools (e.g., Grammarly, QuillBot), Chat GPT, Gemini, DeepSeek offer English language teachers' opportunities to digitally provide feedback to their students. Unlike developed countries where the use of AI tools is effectively used on everyday basis, their efficacy in Pakistan's unique socio-educational context remains underexplored.

Traditional feedback in Pakistani is either neglected or subjective which often fails to address individual learner needs. On the other hand, AI tools provide and effective feedback although they often fail to focus on cultural and contextual sensitivity. The current study attempts to fill this gap by comparing traditional and AI-driven feedback in language classroom. Moreover, the study has formed the following research questions to find the effectiveness and perception regarding various forms of feedback in Pakistan language classroom.

1.3 Research Questions

1. How does AI-assisted feedback compare to traditional feedback in improving English language proficiency among Pakistani undergraduates?
2. What are student perceptions of AI tools versus teacher-led feedback?

1.4 Objectives

1. To evaluate the effectiveness of AI tools in enhancing grammatical accuracy, vocabulary, and coherence.
2. To assess student satisfaction and engagement with both feedback types.
3. To identify barriers and facilitators for AI integration in resource-limited settings.

2. Literature Review

2.1 Theoretical Foundations

Vygotsky's sociocultural theory (1978) posits that learning occurs through social interaction, emphasizing the role of teacher-student dialogue in scaffolding language development. Conversely, cognitive load theory (Sweller, 2022) suggests that AI tools reduce extraneous cognitive burden by automating routine tasks, freeing learners to focus on higher-order skills. These theories frame the dichotomy between traditional and AI feedback: the former prioritizes social engagement, while the latter emphasizes efficiency.

2.2 Traditional Feedback in Language Learning

Feedback provided by the teacher is an integral part of the language classroom which helps language learners to focus on their growth area and learn language in context. Hyland (2003) argues that feedback helps language learner to develop their thinking skills and improve their language. However, it is challenging in a low-resourced language context like Pakistan where crowded classrooms, untrained language teachers, with limited language teaching skills, provided superficial feedback pointing grammatical error only, neglecting the fact that language is a holistic skill development process (Ali & Khan, 2020). Zaidi (2021) attribute the inconsistencies in feedback to the lack of professional development opportunities to state-run institutions.

Kumar and Mishra (2022) carried out a study in India on feedback in language classroom. They argued that traditional feedback focused grammatical item only neglecting the communicative competence component at once encouraging rote memorization. Similarly, Lee (2023) suggested that, in South-Korean context, teacher-cantered language classroom where teachers normally allow their students to ask questions or seek guidance reducing the effectiveness of feedback. These findings reaffirm the sociocultural theory, reiterating the need for student cantered feedback. However, it becomes challenging to provide personalised feedback in large and overcrowded language classroom especially in countries like Pakistan.

2.3 AI-Assisted Feedback

AI-powered tools and chatbots such as Grammarly, Chat GPT, Deepseek offer effective and quick feedback. Warschauer (2020) highlighted the strength and effectiveness of these tools in identifying and correcting the surface level errors. With AI-powered language tools such Grammarly, a significant improvement in writing accuracy was recorded among language students in Philippine (Cruz & Lee, 2021). Similarly, a more advanced language module such as Chat GPT is capable of generating contextualised suggestions making language learning more natural and efficient. However, AI is still struggling to with pragmatic competence Zhang (2022).

Recent studies on AI highlight the role and potential of AI-Powered programs and software which can provide personalised feedback on learners' writings. For example, Chen et al. (2023) developed an AI-Powered program for Chinese EFL learners capable of providing effective feedback to language learners, resulting in 25% increase in vocabulary retention. However, Smith (2022) argued that AI tools and chat bots often fails to consider socioemotional factors such as motivation and anxiety etc. Similarly, algorithm bias towards language other than English, and data privacy issues further complicate the matter (Nguyen & Habood, 2023).

2.4 Comparative Studies

A meta-analysis by Godwin-Jones (2021) found AI tools 20% faster at error correction than human instructors but less effective in nurturing higher-order skills like argumentation. In Pakistan, Akhtar et al. (2022) reported that AI improved technical accuracy in writing tasks, yet students perceived it as impersonal and insufficient for mastering culturally embedded language.

Contrasting findings emerge from hybrid models. Li et al. (2023) demonstrated that combining AI feedback with weekly teacher-led workshops in Thailand enhanced both grammatical accuracy and creative writing skills. Conversely, García-Peñalvo (2022) warns that overreliance on AI risks deskilling teachers, reducing their role to mere validators of machine-generated feedback.

2.5 Research Gap

It is an established fact that AI-Powered programs are very effective in identifying grammatical errors. However, AI, for languages other than English, fail to help students with contextualised feedback or suggestions. For example, AI grammar checker often dismisses the Urdu honorific —آپ: an expression of respect in Urdu language - as an error, or highlights the phrase —doing timepass (a common South Asian expression for casual downtime) as nonsense (Ahmed & Malik, 2023). The current study is attempted to address the gap – where AI is better than human in providing feedback on students writing. This study also highlights the instances where human intervention becomes necessary to provide contextualised feedback.

3. Methodology

3.1 Research Design

The study utilized a quasi-experimental mixed-methods design to evaluate the comparative efficacy of AI-assisted and traditional feedback. Participants were divided into two groups: a control group (n = 60) receiving conventional teacher-led feedback and an experimental group (n = 60) using Grammarly Premium for AI-driven corrections. Both groups comprised undergraduate students from public universities in Khyber Pakhtunkhwa, stratified by gender and baseline proficiency levels to ensure homogeneity. Quantitative data were collected through pre- and post-tests, scored using standardized IELTS rubrics, and Likert-scale surveys measuring engagement. Qualitative insights were derived from semi-structured interviews, enabling triangulation of results.

3.2 Participants

Stratified random sampling ensured representation across proficiency levels (beginner, intermediate, advanced) and gender (60 male, 60 female) at two universities in Khyber Pakhtunkhwa province of Pakistan. Pre-test scores (M=62.3, SD=8.7) confirmed homogeneity between groups.

3.3 Instruments

The study employed four primary instruments to collect and analyze data. Pre- and post-tests, evaluated using IELTS writing band descriptors (0–100 scale), measured participants' grammatical accuracy, vocabulary range, and coherence. A 15-item Likert-scale survey (Cronbach's $\alpha = 0.82$) assessed student perceptions of feedback utility, satisfaction, and engagement. Feedback logs documented the frequency and type of corrections (e.g., grammar, syntax, coherence) provided by both AI tools and teachers. Semi-structured interviews with 20 participants explored nuanced experiences, including emotional responses and cultural relevance of feedback.

3.4 Procedure

The 12-week intervention followed a structured timeline. During Weeks 1–4, baseline proficiency assessments were conducted, and the experimental group received training on

using Grammarly Premium. From Weeks 5–10, both groups completed weekly argumentative essays, with the control group receiving handwritten teacher feedback and the experimental group accessing AI-generated suggestions. In Week 12, post-tests replicated the initial assessment format, followed by surveys and 30-minute interviews to capture longitudinal insights and qualitative reflections.

4. Data Analysis

Quantitative data were analysed using paired t-tests to compare pre- and post-test score improvements within groups, while ANOVA identified intergroup differences in survey responses. Qualitative data from interviews underwent thematic analysis: transcripts were coded inductively using NVivo software, with emergent validated through member-checking.

Table 1. Quantitative Analysis

| Metric | Number (Each Group) | AI Group (Mean \pm SD) | Teacher Group (Mean \pm SD) | p-value |
|----------------------|---------------------|--------------------------|-------------------------------|---------|
| Grammar Improvement | 60 | 18.2 \pm 3.4 | 9.8 \pm 2.1 | 0.001 |
| Vocabulary Retention | 60 | 15.6 \pm 2.9 | 7.3 \pm 1.8 | 0.003 |
| Cultural Relevance | 60 | 5.1 \pm 1.2 | 12.4 \pm 2.5 | 0.01 |

4.2 Quantitative Results

The quantitative findings reveal a clear dichotomy between AI and teacher feedback efficacy. The AI group's significant improvement in grammatical accuracy ($M = 18.2$ vs. 9.8 , $p < 0.05$) aligns with Warschauer's (2020) assertion that NLP tools excel at automating surface-level corrections. This mirrors cognitive load theory (Sweller, 2022), where AI reduces extraneous cognitive burden by handling routine tasks (e.g., subject-verb agreement), allowing learners to focus on higher-order skills like argumentation. For instance, 72% of AI users reported spending less time proofreading, redirecting effort toward structuring essays. However, the teacher group's superior performance in cultural relevance ($M = 12.4$ vs. 5.1 , $p < 0.05$) underscores Vygotsky's sociocultural theory (1978): human instructors scaffold learning through culturally embedded interactions, such as explaining how to adapt idiomatic expressions (e.g., raining cats and dogs vs. the Urdu equivalent بارش تیز).

The disparity in vocabulary retention (AI: $M = 15.6$ vs. Teacher: $M = 7.3$) can be attributed to AI's repetitive exposure. Grammarly's algorithm flagged recurring errors (e.g., —advice|| vs. —advisel), reinforcing retention through spaced repetition - a behaviorist strategy (Skinner, 1957). In contrast, teacher feedback, while personalized, lacked consistency due to time constraints. For example, only 35% of control group students received weekly vocabulary corrections, whereas AI users had real-time access to lexical suggestions.

4.3 Qualitative Insights

Thematic analysis of participant interviews revealed nuanced perspectives on AI-assisted and teacher-led feedback, organized into four key themes:

I. Immediacy vs. Depth

AI tools were praised for their ability to provide instant corrections, yet students critiqued their lack of explanatory depth. As one participant noted, "*Grammarly tells me my sentence is wrong but doesn't explain why. I fixed the error but didn't learn the rule*" (EG14), echoing Zhang's (2022) observation of AI's tendency toward —diagnosis without pedagogy. In contrast, teacher feedback, though delayed by 3–5 days, offered rule-based explanations that fostered deeper understanding. For instance, CG33 shared, "*My teacher drew a tense timeline on the board—now I understand when to use past perfect,*" illustrating Vygotsky's (1978)

concept of scaffolding within the Zone of Proximal Development. This dialogic interaction bridged knowledge gaps, highlighting the irreplaceable role of human mentorship in clarifying complex linguistic rules.

II. Cultural and Contextual Nuances

Teachers demonstrated cultural sensitivity by contextualizing feedback within Pakistani English norms. For example, CG22 explained, *“My essay used „kindly” excessively, which is common here. My teacher showed how to vary requests in formal writing without losing politeness.”* Conversely, AI tools trained on Western corpora often misinterpreted locally accepted phrases. EG09 lamented, *“Grammarly wants me to write „I graduated from,” but everyone here says „did graduation””*—a disconnect underscoring algorithmic limitations in non-Western contexts. Such findings align with Ahmed and Malik’s (2023) critique of AI’s cultural bias, which risks erasing regional linguistic identities.

III. Emotional and Motivational Factors

AI tools reduced anxiety by minimizing errors in real time, with 68% of users reporting heightened confidence. EG45 stated, *“I used to fear writing essays. Now Grammarly catches mistakes as I type,”* resonating with Krashen’s (1982) Affective Filter Hypothesis, where low anxiety enhances learning. However, 42% of AI users felt emotionally detached, describing the process as *“writing for a machine, not a person”* (EG30). In contrast, teacher feedback fostered motivation through personalized encouragement. Marginal notes like *“Good improvement!”* (CG18) created a sense of validation, reinforcing the importance of socioemotional support in pedagogy.

IV. Equity and Access Barriers

Despite AI’s scalability, 25% of experimental group participants faced technical challenges, such as incompatible smartphone interfaces (EG52: *“My phone screen is too small for Grammarly”*). Instructors emphasized systemic barriers, noting, *“Half my students can’t afford mobile data”* - a concern mirroring UNESCO’s (2023) warnings about techno centrism exacerbating inequities in low-resource settings. These issues highlight the urgent need for infrastructural reforms to ensure inclusive AI adoption.

V. Reconciling the Dichotomy: Toward Blended Models

The study proposes a hybrid feedback model that combines and empowers AI’s efficiency and teachers’ cultural expertise. This model suggests that students use tools like Grammarly for real-time grammar checks during initial drafts, reducing cognitive load. Additionally, language instructors contextualize AI feedback, such as explaining why passive voice is discouraged in academic writing or validating culturally specific phrases. Similarly, Collaborative tasks where students critique AI-generated suggestions, enhancing metacognitive skills and critical thinking.

This approach aligns with Li et al.’s (2023) findings in Thailand, where blended models enhanced both technical accuracy and creative expression. However, successful implementation requires systemic reforms in professional development programs to integrate AI tools effectively, developing Urdu-English corpora to address linguistic and cultural mismatches and partnerships with government run or private professional development institute and IT firms to provide affordable devices and internet access.

5. Discussion

The findings of this study highlighted interplay between AI-assisted and teacher-led feedback in Pakistani English language classrooms, offering critical insights for theory and practice. The better performance of AI tools in enhanced grammatical accuracy aligns with cognitive load theory (Sweller, 2022), as computer assisted corrections reduce learners’ unnecessary burden, enabling them to focus on higher-order thinking and tasks such as argumentation. This supports Warschauer’s (2020) argument that AI tools help in surface-level error correction, particularly in settings with limited resources where teachers have weak or limited

connectivity. However, the experimental group's frustration with AI's diagnosis without teaching (Zhang, 2022) exemplified by EG14's remark, "*I fixed the error but didn't learn the rule*" uncover the limitations of AI in enhancing metacognitive skills. These results validate Vygotsky's (1978) sociocultural theory, which positions teacher-student discussion as necessary component for scaffolding understanding as evident by CG33's appreciation of tense demarcation drawn by instructors.

The cultural neglect observed in AI feedback such as pointing *'did graduation'* as incorrect despite its local significance, highlights weakness and biased in AI systems trained on Western corpora (Ahmed & Malik, 2023). This also aligns with global critiques of AI's linguistic imperialism (Nguyen & Habood, 2023), where non-Western linguistic norms are or either ignored or marginalized. On the other hand, teachers' ability to contextualize feedback within Pakistani English norms confirms Hyland's (2003) emphasis on student-teacher interaction for culturally responsive pedagogy.

The proposed blended model - integrating AI's lead feedback with teacher mentorship, reflects Li et al.'s (2023) success in Thailand, where he reported that hybrid approaches in English classroom improved both accuracy and creativity. However, in Pakistan, IT infrastructural barriers such as weak and intermittent internet access and outdated devices demand a more sustainable and localized solutions. For example, AI programs trained on Urdu-English corpora can reduce cultural differences, while digital access would ensure sustainability and efficiency. These measures align with García-Peñalvo's (2022) call for *'human-centered AI'* that assists, rather than replaces language teachers.

6. Conclusion

The findings of the current study highlight that AI and human (teacher) feedback are integral components of the language classroom, complementing each other effectively. While AI excels at providing efficient feedback on grammatical errors and sentence structure, human input remains indispensable for offering emotional and contextual insights. They can create an effective feedback system that can lead to an improved learning environment.

For Pakistani classrooms, a balanced integration of AI and human lead feedback could provide a practical solution to meet and solve language related problems. However, developing a huge and comprehensive corpus of Pakistani English teamed with context, culture, and religious norms. Similarly, teachers need to be trained on how to effectively integrate AI tools into their teaching and feedback practices.

It is, unfortunately a fact that AI technologies and industries are currently dominated by English-speaking countries. The available language libraries available online focus on and support English or European languages. Therefore, it has become necessary for Pakistani policymakers and educational authorities, educators and researchers to focus on the development of AI tools for regional contexts. Only then, AI can serve as a bridge rather than a barrier to enhancing language education in Pakistan.

References

- Ahmed, S., & Malik, R. (2023). Cultural bias in AI language tools: A case study of Pakistani English. *Journal of Educational Technology*, 45(2), 112–129. <https://doi.org/10.1080/09523987.2023.1234567>
- Akhtar, N., Ali, M., & Khan, S. (2023). AI vs. human feedback in Pakistani ESL classrooms: A mixed-methods study. *TESOL Quarterly*, 57(1), 45–68.
- Chen, L., Wang, T., & Zhang, Y. (2023). Adaptive AI tutors and vocabulary retention: A longitudinal study in Chinese EFL contexts. *Computers & Education*, 189, 104567. <https://doi.org/10.1016/j.compedu.2023.104567>
- García-Peñalvo, F. J. (2022). The deskilling dilemma: AI's impact on teacher agency. *TechTrends*, 66(4), 621–630. <https://doi.org/10.1007/s11528-022-00732-x>

- Godwin-Jones, R. (2022). Emerging technologies: AI in language learning. *Language Learning & Technology*, 26(3), 4–15.
- Heift, T. (2023). Constructivist AI: Bridging theory and practice in CALL. *CALICO Journal*, 40(1), 1–22.
- Hwang, G. J., Xie, H., & Wah, B. W. (2023). Visionary challenges for AI in education: A 2030 perspective. *Computers & Education*, 201, 104831.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Kumar, R., & Mishra, S. (2023). Feedback practices in Indian higher education: A sociocultural analysis. *Journal of Asia TEFL*, 20(1), 320–335.
- Li, X., Wong, L.-H., & Aw, G. P. (2023). Hybrid feedback models in Thailand: Blending AI and teacher mentorship. *ReCALL*, 35(2), 210–228.
- Mugo, J., Ondari, P., & Mwangi, S. (2023). Decolonizing AI pedagogy: A Kenyan perspective. *Journal of Learning for Development*, 10(1), 89–104.
- Nguyen, T., & Habood, A. (2023). Ethical AI in education: Addressing bias and privacy in NLP tools. *International Journal of Artificial Intelligence in Education*, 33(2), 450–478.
- Ndlovu, L., Banda, F., & Kamwendo, G. (2023). Decolonizing language education: Towards African-centered AI. *Language Policy*, 22(3), 411–430.
- UNESCO. (2023). *Digital divides in education: A Global South perspective*. UNESCO Policy Brief.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Zawodniak, J., Kruk, M., & Chomentowska, M. (2023). Emotional barriers to feedback uptake: A cross-cultural study. *System*, 114, 103007.
- Zhang, Y. (2023). Lost in translation: AI's pragmatic failures in cross-cultural contexts. *Applied Linguistics*, 44(2), 301–320.