Comparative Effectiveness of AI-Assisted Feedback vs. Traditional Teacher Feedback in Enhancing English Language Proficiency amongst Pakistani Undergraduate Students

Shah Nawaz Barich
PhD Scholar Institute of English Language & Literature, University of Sindh Jamshoro
Email: ali.shahnawaz1235@gmail.com

Irum Pervez Memon
Universiti Sains Malaysia
Email: irumpervez89@student.usm.my

Zeeshan Ali Rahu
Lecturer (English), Shaheed Benazir Bhutto University, Sanghar Campus
Email: rahuzeeshan@gmail.com

ABSTRACT
Using a mixed-methods approach, this study investigates the comparative effectiveness of AI-enabled feedback in comparison to traditional teacher-centered practices in English language global classrooms. The study draws upon data obtained from 120 undergraduate students at two higher educational institutions in Pakistan, which suggests that AI tools such as Grammarly can dramatically improve grammatical accuracy and vocabulary retention. On the other hand, whereas the teacher feedback mechanism is limited in immediateness, it fosters cultural relevance and learners' comprehension assisted by Vygotsky's socio-cultural theory. As corroborated by the provided qualitative findings, teachers offer necessary guidance that benefit from scaffolding and a more profound understanding of the students' cultural backgrounds. Nonetheless, while students generally report satisfaction with AI-informed collaboration due to time efficiency, they indicate an aspect of disconnection attributed to emotional rapport with a teacher. In addition, this study reveals what infrastructure-related obstacles impede the efficient integration of AI and exacerbate disparities amongst the less-privileged, suggesting a blend of AI and traditional practices. Further recommendations include the need for establishing context and localizing AI tools using Urdu-English corpora to eliminate cultural bias, as well as making sophisticated AI products available to educators. Ultimately, the findings imply the need for hybrid AI-centered applied linguistics frameworks that balance precision and cultural sensitivity in marginalized settings, in alignment with comparative research evidence showing the effectiveness of a blended approach.

Keywords: AI, Corpora, English, Grammarly, Teacher,

INTRODUCTION
The English language is the medium of academic success, Civil and Military services as well as Industry in Pakistan, where it is used as a lingua franca for education and governance, however, the national language is Urdu. In State-run institutions, overcrowded classrooms, a scarcity of trained teachers, and outdated pedagogy impede effective education and feedback, which is the backbone of language teaching and

leering. In Pakistan, feedback on writing is still given in the same way it was when it was created, and it is dependent on the teacher, who might be dismissive or prejudicial. With the development of AI tools such as Grammarly, QuillBot, the latest advancement in AI-driven tools I have used is Chat GPT, Gemini, DeepSeek. English language teachers now have opportunities to provide feedback to their students digitally. Moreover, in the developed countries where these tools are widely being used in day-to-day life, it is an underexplored issue in the Pakistani socio-educational context. Traditional feedback is either disregarded, as described above, or biased, as no information is focused on group or individual learner needs. In the AI-tools features, all students are served with their needs met, however, cultural and contextual sensitivity is often missing. The current study tries to fill this gap by comparing traditional written and AI-driven written feedback in the language classroom in Pakistan. Additionally, to find the response about various forms of feedback in context with the effectiveness and efficiency, the study has developed the following research questions:

RESEARCH QUESTIONS
Research Questions:
To what extent is AI-assisted feedback more effective in improving English language proficiency in Pakistani undergraduates than the traditional approach?
How do students perceive AI based tools compared to teacher-led feedback?

RESEARCH OBJECTIVES
Offer insights on alleged efficacy of items on grammatical accuracy, vocabulary, and coherence.
Assess student satisfaction and participation in both focus groups.
Explore enablers and barriers in AI integration in low resource contexts.

REVIEW OF LITERATURE
In traditional language classrooms, feedback is one of the critical components. They provide the learners with the opportunity to identify and work on the weaknesses. Within language, context feedback is a process that enhances leaner thinking when producing an expatiation. Hyland, 2003, as cited in Ali and Khan (2020), pointed texture of traditional feedback in sub-context of feedback giving in language. They stated that the teacher only focuses only students' grammatical errors, which is why learners of such context become poor readers and writers. Generally, the traditional context of feedback is challenging and inadequate in such a setting of classroom. This usually intensifies when the setting is low-resourced. Kumar and Mishra (2022) conducted a study on the feedback process in language laboratories in India context. They discovered that feedback only focused on grammatical items and one point language acquisition that encouraged road memory. In South-Korea context, Lee postulated that due to the teacher centredness and the willingness of the teacher to allow his/her students to ask question some did carry the language, making feedback processing complex. Studies provide empirical evidence to sociocultural theory. Lee (2023) stated that teacher interaction is essential in language classrooms. The teachers should provide student-centred feedback; however, it could be somehow difficult to provide the same in informal and overcrowded classrooms as in Pakistan.

AI-Assisted Feed back

Experts found that AI-powered tools and chatbots such as Grammarly, Chat GPT, Deepseek offer effective and quick feedback. According to Warschauer (2020), the strength and effectiveness of AI are identifying and correcting the surface level errors. In the study of improvement in writing accuracy among language students in Philippine (Cruz & Lee, 2021), a significant improvement was observed with the aid of AI-powered language tools such Grammarly, in a similar vein, with Chat GPT a more advanced language module generating contextualized suggestions making language more natural and efficient approach, AI are still struggling with pragmatic competence Zhang (2022). Recent research on AI focused on the possible and current use of AI-Powered programs and software's which can provide learners' writings with personalized feedback. For instance, AI-Powered program for Chinese EFL learners develop by Chen et al., (2023) which has shown to provide feedback to language learners and increase vocabulary retention rate by 25%. On the other hand, Smith (2022) argued that AI tools and chat bots often fails to consider socio-emotional factors such as motivation and anxiety etc. Similarly * algorithm bias for other than English language and data privacy issues complicated the factors affecting the AI tools (Nguyen & Habood, 2023) arise.

Comparative Studies

According to Godwin-Jones (2021), also found that AI tools error correction was 20% quicker than the human instructors but less effective in cultivating higher-order skills such as argumentation. For Akhtar et al. (2022), in Pakistan, AI increased technical knowledge in writing duties but is derogatory to students as _impersonal and less proficient for mastering culturally embedded language. Contrasting evidence comes from hybrid models. According to Li et al. (2023), for instance, investigated that _Responsive of AI feedback with teacher-led workshops on a weekly basis in Thailand increase both the grammatical accuracy and creative writing skill. The study by García-Peñalvo (2022), argues that AI over-reliability risks teachers engenderment, reducing them to mere validates of the machine-generated feedback.

Research Gap

While it is indisputable that AI-Powered programs are highly instrumental in identifying grammatical errors, it can, however, be countered that AI programs have been developed for any language other than English do not provide feedback or recommendations that are contextualised to the sentence or paragraphs level for students. For instance, an Urdu honorific " آپ " – a respectful expression in the Urdu language –might be dismissed as inappropriate by an AI grammar checker or the colloquial phrase " doing time pass" that means spending time, often referred to as lazy time, in South Asia is designated as unintelligible by an AI system as (Ahmed & Malik, 2023) expressed. This research aimed at filling the gap – where AI tools outperform humans in providing feedback on students writing. Additionally, while this paper addressed those limitations and highlighted meaningful human interference for feedback, this paper would address that part of it.

METHODOLOGY
Research Design

A quasi-experimental mixed-methods design was used to compare the effectiveness of the type of feedback delivered by AI-assisted feedback and conventional feedback in this study. Participants were divided into two categories: a control group = 60, who received

conventional feedback from a teacher, and an experimental group n = 60, which received AI-corrected feedback through Grammarly Premium. All participants were university-level students from Sindh public universities who were equalized by gender and initial course levels to make sure homogeneity. Quantitative data was collected using pre-and post-tests and IELTS rubrics to evaluate and compare mean and standard deviations, and the participants expressed their sentiments on a Likert scale to describe their experience. Qualitative data were also gathered by semi-structured perceptions interviews if necessary.

Participants
Stratified random sampling was utilized to choose participants from an undergraduate level in universities selected based on language competency level: beginners, intermediate and expert students and gender level in Sindh. From two universities in Sindh, 60 males and 60 females participants were chosen. A pre-test mean of m = 62.3, and SD = 8.7. The two groups were homogenized.

Instruments
The four primary instruments used in the study were utilized to gather, analyze and determine data. Pre-and post-tests, aligned to the IELTS writing band descriptors measured on a scale of 0–100, surveyed participants' grammatical accuracy, vocabulary range, and coherence. A 15-item Likert-scale survey, with a reliability of Cronbach's $\alpha = 0.82$ was employed to capture students' perceptions and qualities of the feedback including utility, satisfaction, and engagement. It included Feedback logs capturing the AI tool and teacher corrections' frequency and type, including grammar, syntax, and coherence. Semi-structured interviews conducted with 20 participants captured the nuanced experiences like emotional reactions and the cultural-oriented relevance perception of feedback.

Procedure
The 12-week randomized controlled trial followed a structured protocol. During Weeks 1 to 4, baseline proficiency tests were conducted to assess the experiment and the control group. During the same period, the experiment group was trained on how to use Grammarly Premium. From Weeks 5 10, both groups participated in writing the argumentative essays, while the control group received handwritten teacher feedback, the experimental group received AI tool-generated suggestions. In Week 12, students undertook post-tests targeting toning the initial assessment quality and format. The post-tests were followed by 15-question survey checklists and 30-minute interviews aimed at capturing longitudinal approaches and qualitative perceptions.

DATA ANALYSIS
Quantitative data were analysed with paired t-tests to examine pre-and post-test score improvements within groups, and ANOVA to note inter group differences in survey responses. Qualitative data from interviews were thematically analyzed. The interviews were coded inductively from transcripts using NVivo software and the emergent themes were validated by member-checking.

Vol. 3 No. 11 (November) (2025)

Quantitative

Metric No.	Group	AI (Group) Mean (SD)	Teacher (Group Mean (SD)	P Value
Improvement (Grammar)	60	18.2+3.4 -	9.9+2.0 -	0.001
Retention of Vocabulary	60	16.6+3.0 -	7.2+1.9 -	0.003
Relevance of Culture	60	5.0+1.1 -	11.2+2.3 -	0.01

Quantitative results show a clear dichotomy between the efficacy of AI and teacher feedback. The AI group's grammatical accuracy was significantly better than the teacher group; the improvement supported Warschauer's (2020) assertion that NLP tools are particularly adept at automating surface-level corrections, resulting in twenty-first-century learning outcomes. This can be described through cognitive load theory; using AI reduces extraneous cognitive load by undertaking routine tasks such as subject-verb agreement, allowing students to develop higher-order skills such as argumentation. For example, 72% of AI learners reported spending less time on proofreading, which helped them redirect efforts towards essay structure. However, the teacher group outperformed the AI group in the cultural relevance dimension. Vygotsky's sociocultural theory explains it suggestive that human teachers provide scaffolding for learning through culturally embedded interactions, e.g. by telling me how to adapt idioms. The teacher group was also significantly worse at remembering vocabulary. This can be linked to spaced repetition, a behaviorist technique Grammarly's algorithm on personalized repetition flagged recurring vocabulary mistakes (i.e. advice vs. advise), whereas the teacher feedback was less consistent due to time constraints. For example, only 35% of the control group received weekly vocabulary corrections, whereas all AI users received lexical relevance suggestions in real-time.

Qualitative
Thematic analysis of participant interviews elicited nuanced perspectives on AI-assisted and teacher-led feedback, which were organized into four key themes.

Immediacy vs. Depth
AI tools were lauded for their immediacy, allowing students to know the correct answer right away. However, they were criticized for their lack of depth. As Student EG14 phrased it, "Grammarly tells me my sentence is wrong but doesn't explain why. I fixed the error but didn't learn the rule!" This experience reflects Zhang's (2022) observation of AI as more inclined to "diagnose without pedagogy". Teacher feedback, by contrast, took 3–5 days to obtain but was underpinned by explanations of rules. For example, as Student CG33 noted, "My teacher drew a tense timeline on the board. Now I understand when to use past perfect." This exchange illustrates Vygotsky's concept of the zone of proximal development.

Cultural and Contextual Nuances
Additionally, all the teachers were also particularly aware of the Pakistani English usage, allowing them to contextualize feedback as per local norms. "My essay used 'kindly' excessively, which is common here. My teacher showed how to vary requests in formal

writing without losing politeness" (CG22). In contrast, the AI tools trained on Western corpora would often be unable to interpret "everyone here says 'did graduation,'" as described by EG09. The same trend was echoed in Ahmed and Malik (2023) argued that the cultural bias ingrained in every AI risks perpetuating the dominance of Western cultural standards.

Emotional and Motivational Factors
As a result, 85% of AI users reported higher confidence in writing and success level due to the ease of rectifying mistakes. "I used to fear writing essays. Now Grammarly catches mistakes as I type," shared EG45, directly tying into Krashen (1982) that low anxiety equals high motivation in learning. However, 42% of individuals viewed the entire process as "writing for a machine, not a person" and were thus emotionally detached. In comparison, receiving feedback from the teacher made students more motivated as they felt validated. Regardless of the content, marginal notes like "Good improvement!" put a smile on the face of CG18. This underlines Azorin Reyes and Mckenzie that socio-emotionality plays a crucial role in learning.

Equity and Access Barriers
Age, gender, class, and even experiment group – singular innovations should erode such barriers. 25% of EG participants faced tech challenges, with the smartphone interface being incompatible with the tool . Additionally, the teacher contribution brings other systemic issues, "Half my students can't afford mobile data". This concern is in line with UNESCO (2023) warnings that the techno-centric approach to education deteriorates already existing inequalities. It thus highlights the future science-driven reforms this country needs and deserves.

Reconciling the Dichotomy: Toward Blended Models
AI is superior, but it will never compensate for the teacher's role in language education. This model should, however, be smart – students should use Grammarly to avoid trivial grammar errors and typos that heavily load the cognitive. The teacher should work after that, reminding the more profound language principles that AI is not yet adept to teach and guide the feedback instead of designating it. As for cooperative tasks, students can also review the suggestions and insights on language from the AI and praise or criticize it together, promoting meta-cognitive minds. According to Li et al. (2023), such models proved to be smarter in Thailand as they not only increase accuracy but also broaden identity. However, everything will pop without systemic reform: professional we need to adapt the new teaching schemes and get relevant urdu-English corpora; the government should partner with IT to distribute devices and internet to the most remote areas.

DISCUSSION
This study has detailed the interplay between AI-assisted and teacher-led feedback in Pakistani English language classrooms, with wide-ranging implications for theory and practice. The enhanced grammatical accuracy of AI tools can be justified by cognitive load theory (Sweller, 2022), as per the cognitive load theory, computer-assisted correction provides less work for learners unnecessarily, so they can allot more time to higher-order thinking and work, such as argumentation. Therefore, the AI tools can correct surface-level errors: Warschauer (2020) highlights this reasoning AI tools on their surface, especially in pandemic-ridden settings such as Pakistan, where most teachers have little to no access to internet connectivity (Zhang, 2022). At the same time,

Vol. 3 No. 11 (November) (2025)

the frustration of the experimental group of AI's diagnosis without treatment as can be seen from EG14's comment, "I fixed the error but didn't learn the rule" emphasized the limitation of AI feedback in terms of improving metacognitive skills. On the one hand, this finding provides backing for Vygotsky's sociocultural theory: the teacher-student dialogue is necessary, as clearly shown by the appreciating comment of CG33 about the tense demarcation given by teachers. The neglect of Pakistani norms in AI feedback is also worrying: the AI pointed 'did graduation' as an error, despite it being perfectly normal in Pakistani English, shows the systemic bias of AI trained on a Western corpus (Ahmed & Malik, 2023). The sociocultural theory promotes teacher-learner interaction, making the culture error prevalent in AI irrelevant. Similarly, the cultural preparedness of teachers supports Hyland's (2003) argument for a teacher component in language teaching and learning. The proposed blended model perfectly satisfies Li et al.'s (2023) rationale for hybrid approaches in English classrooms, with the challenge in Pakistan being the low connectivity and outdated devices such as PC as limiting factors for sustainable success It also followers García-Peñalvo's (2022) call for human-centered AI that doesn't replace language instructors but rather assists them.

CONCLUSION

The results of the ongoing study indicate that AI and human can effectively coordinate with one another in a language classroom. Pakistani classrooms present an actual fighting chance of integrating AI into the existing feedback system. AI could fill the gap to correctly help AI in the learning process. AI can multi-learn various languages and recommend the best way, but ensuring time requires enough learning data. However, other sides are encouraging formulating a balanced integration of AI, and human can work in Pakistani theories and solve problems related to their learners. One main challenge for Pakistani education would be the corpus. A vast corpus database should be developed for Pakistani English. Teachers must be trained on how to utilize AI systems effectively. In summary, I must say that it is a fact that the majority of AI technologies and industries are from English-speaking countries. It is almost all libraries available online support the European or eastern languages. Consequently, I feel that it has become necessary for policymakers and educational authorities, educators, and researchers in Pakistan to focus on developing AI tools. Only then will AI act as real language learning support, rather than a stringent challenge to tackle.

REFERENCES

Ahmed, S., & Malik, R. (2023). Cultural bias inAI language tools:Acase study of Pakistani English.JournalofEducationalTechnology,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 ESLclassrooms: Amixed-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).Thedeskillingdilemma:AI'simpactonteacheragency. 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). ConstructivistAI: Bridging theory and practice in CALL. CALICO Journal, 40(1), 1–22.

Vol. 3 No. 11 (November) (2025)

Hwang,G.J.,Xie,H.,&Wah,B.W.(2023).Visionary challenges for AIineducation: A 2030 perspective. Computers & Education, 201, 104831.

Krashen,S.D.(1982).Principlesandpracticeinsecondlanguageacquisition.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: BlendingAI 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. UNESCOPolicyBrief.

Vygotsky,L.S.(1978).Mindinsociety:Thedevelopmentofhigherpsychological 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.(2022).Lostintranslation:AI'spragmaticfailuresincross-cultural contexts. Applied Linguistics, 44(2), 301–320.