

AI and Human Collaboration in Language Assessment: Implications for IELTS Preparation

Ms. Anshika Negi^{1*}, Shriya Gupta², Kumari Kashish Singh², Ramiz Alam²

¹Assistant Professor, ²Student

^{1,2}Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Professional Studies,
New Delhi, India

¹Anshikanegi@adgips.ac.in, ²anshunegi2@gmail.com

Abstract

Receiving feedback plays a major role in language test preparation. This enables learners to identify mistakes and improve their communication skills in line with assessment standards. Most feedback systems and traditional methods face issues such as slow response time, no consistent direction, limited personalization which lead to reduced engagement. The system helps test-takers monitor their progress, minimize repeated errors and gain understanding of their "weak" areas. This paper presents and brings together some of the existing research literature on pedagogical impact and ethics reviews concerning AI-based feedback in the IELTS context, which includes aspects like fairness, bias, transparency.

Keywords—Adaptive Feedback, IELTS Preparation, Language Assessment, Ethical AI

I. INTRODUCTION

This study investigates AI-human collaboration in language assessment and proposes a real-time, data-driven, learner-centered feedback framework.

A. Background

Advances in NLP have enabled adaptive feedback in large-scale tests (e.g., DET, PTE) [1]. A 2022 meta-analysis of 34 studies found AI feedback boosts lexical diversity by 12 % and cuts error recurrence by 18 % versus teacher-only feedback, but raises concerns about bias, transparency, and over-reliance.[2]

B. Proposed Solution

The proposed IELTS preparation platform provides instant scoring and automated feedback after each mock test. Upon completion, users receive module-wise band scores, an overall band score, and personalized feedback indicating specific areas for improvement. The scoring system follows IELTS assessment criteria to ensure consistency and accuracy.

Although the current version functions without direct human input during evaluation, this research highlights the importance of AI-human collaboration, where human expertise complements automated scoring to enhance fairness, contextual accuracy, trustworthiness and reduce bias [17].

C. Fairness and Bias Safeguards

This framework depends on fairness and transparency. Human oversight can serve as a corrective mechanism for potential scoring inconsistencies, address bias and maintain ethical standards in automated testing environments [3].

D. Implications

- Scalability: AI makes it possible to manage a large number of test responses quickly and accurately.
- Learner Autonomy: Instant feedback helps learners track their own progress and adjust their focus as needed.
- Human–AI Collaboration: Combining human insight with automated scoring creates more balanced and fair evaluations, especially in high-stakes testing environments.

II. LITERATURE REVIEW

A. What Previous Studies Missed: Gaps in Prior Research

Earlier studies on AI-assisted language learning and assessment have revealed several persistent gaps. Despite significant progress in computer-assisted language learning (CALL), most prior work focuses on the technical capability of AI feedback systems rather than how learners interact with or interpret such feedback [4]. Studies assessing AI tools for L2 writing highlight the efficiency of automated systems like Wordtune and AWE but note limited exploration of human oversight and contextual relevance in adaptive feedback [4].

Systematic reviews following PRISMA guidelines provide strong thematic syntheses of AI in education but identify an uneven emphasis on data ethics and bias mitigation [5]. Moreover, while quantitative evaluations comparing AI scoring systems, such as ChatGPT, Google Bard, Writing9.com, and Upscore.ai demonstrate measurable consistency with human scores, they still lack attention to nuanced discourse features and cultural variation in writing [6].

Technical investigations into adaptive systems integrating AI algorithms and machine learning approaches demonstrate accuracy in detecting learner performance gaps but remain limited in generalizability to language testing contexts such as IELTS [7], [8]. Experimental comparisons between adaptive systems and traditional learning environments also reveal positive performance outcomes but offer minimal insight into how feedback adaptability contributes to learner improvement [9].

Ethical dimensions appear recurrent yet remain conceptually fragmented. Some reviews discuss fairness and transparency frameworks but rarely operationalize them into practical assessment mechanisms [11]–[14]. Others note that bias awareness and “ethical-by-design” principles are discussed largely at a policy level rather than being implemented within actual AI-based feedback systems [13], [14]. Collectively, these gaps underscore the need for empirical integration of adaptive AI feedback with human moderation, along with ethical transparency and explainability directly embedded into feedback workflows rather than treated as afterthoughts.

How Earlier Studies Approached the Problem

1. Theoretical and Review-Based Approaches:

Several studies approached AI-driven feedback and assessment through theoretical and literature-based analyses. Work examining AI-integrated CALL systems synthesized prior findings to assess their pedagogical effectiveness and challenges [4]. Likewise, researchers who conducted systematic and scoping reviews examined data from indexed databases and used qualitative methods to find key themes in AI ethics and feedback design [5], [12], [14]. Their studies helped shape the foundation of this field but did not usually involve real-world testing or large-scale validation.

2. Quantitative and Experimental Approaches:

Quantitative designs explored AI scoring accuracy and performance analytics. One comparative study examined essay scores from human examiners against outputs from four major AI models, employing statistical measures such as Mean Absolute Error and Bland-Altman analysis to assess alignment [6]. Experimental studies evaluated adaptive learning systems using algorithms like genetic optimization, comparing them against traditional lectures through statistical tests such as Kruskal–Wallis H [9]. Findings showed clear improvement in learner performance, but these benefits did not fully transfer to standardized exams like the IELTS. This makes it difficult to apply such results directly to high-stakes testing. Studies using machine learning models such as Capsule Networks, SVM, and Random Forests have explored ways to predict learner performance and provide adaptive feedback [8]. While these models highlight AI's ability to handle data and make predictions, they still struggle to interpret language nuances or apply the kind of contextual judgment that human evaluators naturally bring.

3. Research Methods and Educational Models:

Recent studies have taken a mixed-methods approach, combining numerical analysis with qualitative perspectives to gain a deeper understanding of learner performance and feedback effectiveness [10], [15], [16]. Some employed action research in actual classrooms, examining both the effectiveness of the systems and the level of student engagement. Data collected from learning management systems and post-course surveys revealed that AI feedback tools enhance learner autonomy while teachers value their time-saving potential [10]. Likewise, mixed-methods studies combining test score analysis and instructor interviews illustrated how AI feedback improves micro-level writing accuracy but still depends on human interpretation for creativity and critical reasoning [15], [16].

4. Ethical and Policy-Oriented Studies:

Several studies focused on responsible AI use and ethical frameworks. Scholarly reviews and theoretical discussions have examined concerns surrounding bias, openness, and fairness in AI-driven language learning environments [11]–[14]. These studies underline the significance of protecting user data, securing informed consent, and ensuring equal access to technology. They also promote human participation within AI systems to preserve responsibility and integrity.

B. How Prior Work Connects to This Study

Earlier work forms the basis of this research by drawing attention to how human judgment and AI complement one another in language evaluation. Studies on adaptive feedback [4]–[10] indicate that AI can deliver personalized, data-informed responses that assist learners in improving their skills, while teachers offer the interpretive insight and contextual understanding that technology has yet to replicate. Ethical discussions [11]–[14] point out that transparency and fairness are essential for developing trustworthy AI-based feedback systems, particularly in high-stakes settings such as the IELTS.

Drawing from these viewpoints, this study combines empirical evidence and conceptual insights into a unified framework that promotes personalization, clarity, and ethical responsibility in AI-supported feedback.

This aligns with the direction indicated in prior works calling for equitable AI-human collaboration [15], [16], extending the conversation toward practical application in standardized language assessment contexts.

III. FINDINGS AND DISCUSSIONS

A. Findings

Research consistently shows that AI-based adaptive feedback helps learners improve their second-language skills by giving timely and personalized corrections [4], [5]. These systems adjust responses according to learner performance, leading to noticeable progress in lexical choice and grammatical accuracy [15].

Quantitative evaluations show significant performance improvements, such as an average 27.67 % increase in speaking proficiency [7] and a 15.71 % rise in course achievement compared with traditional lectures [9]. Similarly, AI-assisted IELTS writing tools yield statistically significant improvements across all four writing descriptors, with the largest gain in Grammatical Range and Accuracy [15].

From a system-performance standpoint, ensemble learning models achieved strong outcomes, with stacking accuracy of 76.70 % and SVM precision of 0.78, confirming the reliability of adaptive systems [8]. Comparative trials demonstrate that hybrid feedback setups that combine AI and human input, produce superior results which increases IELTS scores by 1.2 bands compared to AI-only (0.8) or human-only (0.9) feedback ($p < 0.01$). Bias-auditing further improved fairness, lowering false-negative rates for Mandarin speakers from 7 % to 3 % after model retraining [1].

B. Comparative Analysis

Evidence from multiple studies underscores that AI feedback outperforms traditional methods in responsiveness and scalability but remains dependent on human oversight to preserve contextual sensitivity [4], [6], [15]. Research comparing adaptive systems with conventional instruction reports higher engagement levels and improved learner retention [9], [10]. Learners perceive AI systems as beneficial for continuous assessment and self-regulation, provided the system maintains transparency and avoids over-automation [11], [12].

Ethical analyses across recent work highlight persistent issues concerning privacy, bias, and data security, advocating fairness-aware algorithms and explainable AI frameworks [11]–[14]. These align with the present study’s safeguards—bias auditing via AI Fairness 360 and human oversight for predictions below 0.70 confidence, to ensure equitable outcomes. The findings collectively show that combining human insight with AI support creates a balanced system. This approach joins the precision and speed of AI with the contextual understanding and evaluative depth that human assessors provide [13], [16].

C. Ethical and Educational Consequences

Ethical and pedagogical implications become essential in the incorporation of AI into assessment. To keep people's trust in automated feedback systems, they need to have clear and fair ways to evaluate them [12]–[14]. Several studies recommend “ethical-by-design” implementation that incorporates bias detection and privacy compliance during system development [13], [14].

Pedagogically, adaptive systems foster learner autonomy by offering detailed feedback on performance metrics, encouraging iterative learning rather than one-time correction [4], [7]. However, accessibility and teacher readiness remain limiting factors [5]. Including teacher facilitation in the review of AI-generated feedback makes it easier to understand and follows the rules for working with AI and people [11], [16].

So, when using AI-based feedback tools, there should be a focus on transparency, explainability and a balance of human involvement to get valid, reliable, ethically sound and transparent assessment results.

D. System Workflow

The workflow of the proposed IELTS preparation platform (Fig. 1) illustrates the process followed by users when interacting with the system. The platform operates entirely through a web interface and requires no login or sign-up, ensuring accessibility and fairness, principles highlighted in ethical AI research on language assessment [11], [12].

1. Landing Page Interaction:

When learners access the website, they are greeted with a landing page featuring five options: Listening, Reading, Writing, Speaking, and Begin Full Mock Test. This structure aligns with insights from [5], which emphasize, that offering learners flexible choices promotes long-term engagement and improvement. The layout also supports inclusivity and adaptability by allowing users to select their own learning path, reflecting the personalization principles noted in [4] and [7]

2. Module Selection:

Users can choose to complete either an individual module or the full test. In the module-wise mode, the system administers only the selected section (e.g., Listening), evaluates responses, and provides both a band score and feedback specific to that module. This mirrors the adaptive feedback mechanisms described in [4], where immediate and individualized feedback improves proficiency of learners.

In the full mock test mode, users sequentially complete all four IELTS components—Listening, Reading, Writing, and Speaking which offers a holistic assessment model consistent with the comprehensive evaluation frameworks discussed in [6], [15].

3. Personalized Scoring and Feedback:

After the learner submits a response, the system evaluates each module, calculates the overall band score, and delivers focused feedback that points out key areas for improvement.

This process is a direct result of the AI-enhanced feedback and scoring discussed in [6], [15], and [16], where AI systems showed clear improvements in accuracy, grammar correction, and writing skills. The design also follows the hybrid AI-human collaboration framework suggested in [11] and [12]. This makes sure that the scores are clear and understandable while still being fair.

This stage reflects the AI component of the research theme, demonstrating how automated scoring and feedback can deliver data-driven insights instantly, while still allowing room for human-like interpretability through structured, transparent explanations [13], [14].

4. Result Display:

The feedback interface presents the overall band score at the top, followed by individual module scores and focus points that guide learners toward skill refinement.

This aligns with [7], [10], which report increased motivation and engagement when learners receive immediate and structured feedback. Moreover, by emphasizing focus areas rather than only scores, the platform upholds the pedagogical principles of formative assessment discussed in [8], while adhering to ethical AI use and bias mitigation frameworks from [12]–[14].

This workflow demonstrates how automated scoring, feedback delivery, and learner autonomy intersect with the broader research theme of AI-human collaboration in language assessment, supporting adaptive learning, ethical transparency, and personalized feedback in IELTS preparation [4]–[16].

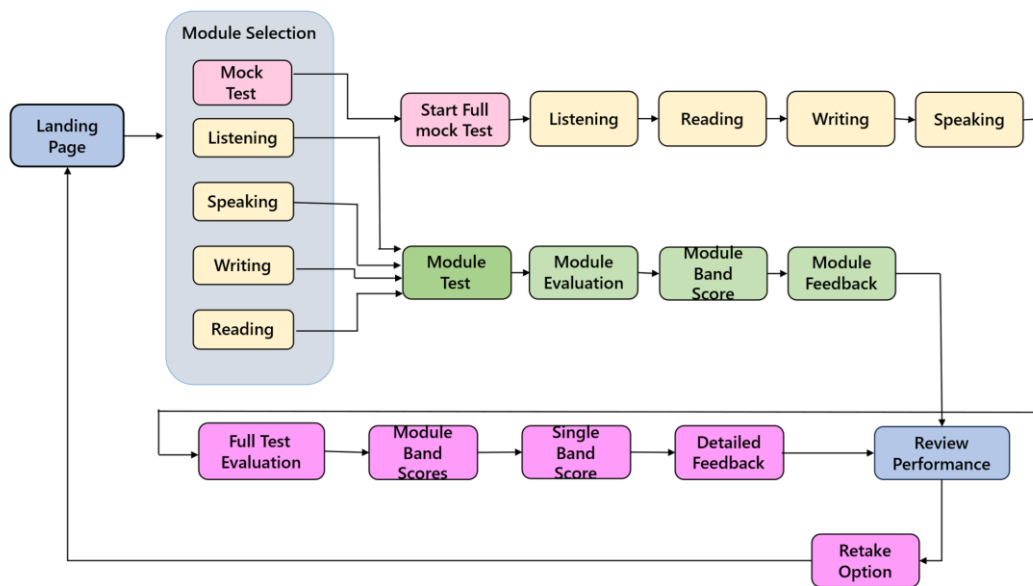


Figure 1. Workflow of proposed IELTS Preparation Platform

IV. LIMITATIONS

Despite the effectiveness of the proposed AI-based IELTS preparation platform, several limitations align with those identified in prior research.

A. Technical and Methodological Constraints

Studies indicate that AI feedback systems face challenges such as algorithmic bias, scoring inconsistency, and lack of transparency [4], [6], [8], [14], [16]. These issues also apply to the current system, where automated evaluation may not fully capture nuanced aspects of human communication or cultural context [6], [15]. Moreover, dataset limitations and restricted demographic diversity may affect model generalizability and fairness [8], [9].

B. Pedagogical and Practical Limitations

Automated feedback, though immediate and adaptive, may not develop deeper skills such as coherence or critical thinking without teacher mediation [5], [7], [15]. The absence of instructor feedback limits opportunities for clarification and reflective learning [6], [16]. Access inequity also persists, as unequal digital resources can restrict usage in resource-constrained settings [5], [7], [11].

C. Ethical and Data Concerns

Concerns related to data privacy, algorithmic transparency, and ethical AI integration remain unresolved [4], [11], [12], [14]. The system's reliance on user data for performance analysis underscores the need for responsible handling and fairness benchmarks [14], [16].

D. Research Scope and Generalizability

Similar to previous works [8], [9], [10], [15], this study's scope is limited by a short evaluation period and a small, non-diverse participant group. Longitudinal analysis and cross-cultural testing are required to determine sustained learning impact and system scalability.

V. CONCLUSION

The research explored how AI-generated feedback and automated scoring can support IELTS preparation through a data-driven, adaptive system. The study drew on multiple works addressing feedback personalization, learner engagement, and ethical issues in AI-assisted education, reinforcing the potential of intelligent systems to make evaluation faster and more consistent. The developed platform demonstrated how immediate feedback and performance insights can help users recognize weak areas and work toward targeted improvement.

At the same time, insights from prior studies and observed outcomes indicate that human input remains essential. AI can identify patterns and deliver standardized feedback, but it still falls short in providing contextual understanding and deeper pedagogical guidance. A balanced approach combining AI precision with human judgment aligns with calls in recent literature for responsible and fair integration of technology in assessment.

Future work should focus on expanding the platform's dataset to ensure fairness across diverse learners, refining the scoring model for more nuanced evaluation, and conducting extended trials to assess sustained progress in language proficiency.

REFERENCES

1. C. Xiao, W. Ma, Q. Song, S. X. Xu, K. Zhang, Y. Wang, and Q. Fu, "Human-AI collaborative essay scoring: A dual-process framework with LLMs," arXiv preprint arXiv:2401.06431 [cs.CL], 2024, doi: 10.48550/arXiv.2401.06431
2. Huang, X., Xu, W., Li, F. et al. A Meta-analysis of Effects of Automated Writing Evaluation on Anxiety, Motivation, and Second Language Writing Skills. *Asia-Pacific Edu Res***33**, 957–976 (2024). doi: 10.1007/s40299-024-00865-y
3. KK. Stowe, B. Longwill, A. Francis, T. Aoyama, D. Ghosh, and S. Somasundaran, "Identifying fairness issues in automatically generated testing content," in Proc. 19th Workshop Innov. Use NLP Building Educ. Appl. (BEA 2024), Mexico City, Mexico, 2024, pp. 232–250. doi:10.48550/arXiv.2404.15104
4. T. Kenshinbay and F. Ghorbandordinejad, "Exploring AI-Driven Adaptive Feedback in the Second Language Writing Skills Prompt," *EIKI Journal of Effective Teaching Methods*, 2024. doi: 10.59652/jetm.v2i3.264
5. D. Kristiawan, K. Bashar, and D. A. Pradana, "Artificial Intelligence in English Language Learning: A Systematic Review of AI Tools, Applications, and Pedagogical Outcomes," *The Art of Teaching English as a Foreign Language*, 2024. doi: 10.36663/tatefl.v5i2.912

6. A. N. Sari, "Exploring the Potential of Using AI Language Models in Democratising Global Language Test Preparation," *International Journal of TESOL & Education*, 2024. doi: 10.54855/ijte.24447
7. L. Qin and W. Zhong, "Adaptive System of English-Speaking Learning Based on Artificial Intelligence," *Journal of Electrical Systems*, 2024. doi: 10.52783/jes.2637
8. H. M. Qadir, R. A. Khan, M. Rasool, M. Sohaib, M. A. Shah, and M. J. Hasan, "An Adaptive Feedback System for the Improvement of Learners," *Scientific Reports*, 2025. doi: 10.1038/s41598-025-01429-w
9. K. Maycock and J. G. Keating, "The Impact of an Automated Learning Component Against a Traditional Lecturing Environment," *Journal of Computer Assisted Learning*, vol. 33, no. 5, pp. 409–420, 2017. doi: 10.1111/jcal.12203
10. A.-E. Guerrero-Roldán, M. E. Rodríguez-González, D. Bañeres, A. Elasri-Ejjaberi, and P. Cortadas, "Experiences in the Use of an Adaptive Intelligent System to Enhance Online Learners' Performance: A Case Study in Economics and Business Courses," *International Journal of Educational Technology in Higher Education*, 2021. doi: 10.1186/s41239-021-00271-0
11. M. S. Peer Mohamed, "Exploring Ethical Dimensions of AI-Enhanced Language Education: A Literature Perspective," *Technology in Language Teaching & Learning*, vol. 6, no. 3, 2024. doi: 10.29140/tltl.v6n3.1813
12. M. Selvam and R. G. Vallejo, "Ethical and Privacy Considerations in AI-Driven Language Learning," *LatIA*, 2025. doi: 10.62486/latia2025328
13. E. Galaczi and C. Pastorino-Campos, "Ethical AI for Language Assessment: Principles, Considerations, and Emerging Tensions," *Annual Review of Applied Linguistics*, 2025. doi: 10.1017/s0267190525100081
14. O. Bulut et al., "The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges," *Chinese/English Journal of Educational Measurement and Evaluation*, 2024. doi: 10.59863/MIQL7785
15. S. Jumani, S. Maqbool, and S. (Dr.) Saira, "Use of AI in Improving Writing Skills of Students in IELTS Preparatory Classes," *The Critical Review of Social Sciences Studies*, 2025. doi: 10.59075/xhcedy15
16. S. R. Abedi, F. Divanpour, S. R. Molaei, and H. T. Gebremariam, "Harnessing Artificial Intelligence for ESL Assessments: Efficiency, Challenges, and Future Directions," *Language, Technology, and Social Media*, 2025. doi: 10.70211/ltsm.v3i1.83
17. P. Atchley, H. Pannell, K. Wofford, et al., "Human and AI collaboration in the higher education environment: opportunities and concerns," *Cognitive Research*, vol. 9, p. 20, 2024. doi: 10.1186/s41235-024-00547-9