

THE ROLE OF GENERATIVE AI IN DEVELOPING IELTS WRITING SKILLS

Khaknazarova Sayyora Abdumannon qizi

Independent Researcher and EFL Educator, MA TESOL

Founder and Director, Lingua Sayyora Language School

Email: abdumannonovnas.h@gmail.com

Abstract. *This exploratory quasi-experimental study compared generative-AI-assisted feedback with teacher-led feedback in IELTS Writing Task 2 practice. Twenty pre-intermediate learners aged 16–18 from one private language school were assigned to two equal groups. Both groups completed a pre-test, a four-week intervention comprising 16 writing activities, and a post-test under comparable 40-minute conditions. The AI-assisted group used ChatGPT-generated comments and model revisions, whereas the teacher-led group received feedback from an instructor on the same broad dimensions: task response, coherence and cohesion, lexical resource, and grammatical range and accuracy. At post-test, the AI-assisted group obtained a mean overall band score of 6.5 and the teacher-led group 7.0. Classroom observations suggested that AI support was especially useful for idea generation and lexical alternatives, while teacher feedback was more consistently acted upon and appeared stronger for grammar and essay organization. Because individual pre-/post-test scores, dispersion measures, and validated perception data were unavailable, the study does not claim statistical significance or causal superiority. The findings support a pedagogically supervised, hybrid use of generative AI rather than its use as a substitute for teacher judgement.*

Keywords: *generative AI; ChatGPT; IELTS Writing Task 2; AI-assisted feedback; teacher feedback; EFL writing; feedback engagement.*

1. Introduction

IELTS Writing Task 2 requires candidates to develop a position, organize ideas coherently, use an adequate lexical range, and demonstrate grammatical control. The official analytic framework reports performance through four equally weighted criteria: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. For learners, the difficulty is therefore not limited to producing error-free sentences; it also involves selecting relevant ideas, sustaining an argument, linking propositions, and revising under time pressure.¹

Generative artificial intelligence has expanded the range of feedback available outside the classroom. A learner can submit a draft and obtain near-immediate comments, explanations, lexical alternatives, outlines, and sample revisions. This accessibility is pedagogically attractive, but immediacy should not be equated with learning. Feedback becomes useful only when the learner understands it, evaluates its relevance, and applies it during revision. Research has consequently shifted from asking whether AI can generate feedback to examining its quality, learner uptake, and relationship with teacher mediation.

Recent evidence presents a balanced picture. Escalante, Pack, and Barrett found that AI-generated feedback could be comparable to human feedback in several respects, while learner preferences and the pedagogical value of comments varied. Guo and Wang similarly positioned ChatGPT as a potential support for teacher feedback rather than an automatic replacement for it. These findings make a direct classroom comparison valuable, particularly in test-preparation contexts where feedback must remain aligned with explicit performance criteria.²³

The present study examines how AI-assisted and teacher-led feedback functioned during a short IELTS Writing Task 2 intervention. It addresses three

¹IELTS. Writing band descriptors: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. <https://ielts.org/organisations/ielts-for-organisations/test-statistics-and-research/test-performance> (accessed 2 September 2026).

²Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>.

³Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, Article 57. <https://doi.org/10.1186/s41239-023-00425-2>.

questions: (1) What post-intervention performance was observed in the AI-assisted feedback group? (2) What post-intervention performance was observed in the teacher-led feedback group? (3) What differences in writing dimensions and feedback engagement were observed between the groups? The wording deliberately emphasizes observed patterns because the available dataset supports descriptive, not inferential, conclusions.

2. Literature Review

Feedback in second-language writing serves at least three connected functions: it identifies a gap between current and desired performance, provides information that can guide revision, and supports the learner in monitoring future writing. Generative AI can operationalize these functions at scale by responding immediately and generating multiple reformulations. Its limitations are equally important: comments may be overabundant, generic, inconsistent with the learner's intention, or accepted without critical evaluation.

Empirical studies indicate that AI can support specific dimensions of EFL writing. Poláková and Ivenz reported improvement in aspects such as concision, grammar, and inclusion of key information in a mixed-methods study of university EFL learners using ChatGPT feedback. Mahapatra also documented the potential of ChatGPT-supported instruction to improve academic writing, while stressing the importance of an intervention design that integrates the tool into teaching rather than treating it as an autonomous tutor.⁴⁵ Comparative evidence points toward complementarity. Soori, Khojasteh, and Javed compared teacher e-feedback, AI feedback, and hybrid feedback among 88 intermediate EFL learners using IELTS writing tasks; the hybrid condition produced the strongest overall gains, while the feedback modes showed different strengths. More recent evidence also cautions that a greater quantity of AI feedback does not guarantee deeper revision. Farrokhnia and colleagues found that generative AI offered extensive comments,

⁴Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, Article 9. <https://doi.org/10.1186/s40561-024-00295-9>.

⁵Poláková, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1), Article 2410101. <https://doi.org/10.1080/2331186X.2024.2410101>.

yet students sometimes revised less than recipients of teacher feedback. Together, these studies suggest that feedback uptake, not output volume alone, is a central mechanism.⁶⁷

The literature therefore supports a mediated model: AI can widen access to rapid explanations and alternatives, while teachers prioritize, contextualize, and verify feedback. The present small-scale study contributes an IELTS-focused classroom observation to this discussion, but it is interpreted within the limits of the available evidence.

⁶Farrokhnia, M., Latifi, S., Papadopoulos, P. M., Hogenkamp, L., Gijlers, H., Khosravi, H., & Noroozi, O. (2026). Generative AI offers more, but students revise less: Comparing the effects of teacher and AI feedback on student essay revisions. *International Journal of Educational Technology in Higher Education*, 23, Article 6.

<https://doi.org/10.1186/s41239-026-00579-9>.

⁷Soori, A., Khojasteh, L., & Javed, F. (2025). Comparing teacher e-feedback, AI feedback, and hybrid feedback in enhancing EFL writing skills. *Technology in Language Teaching & Learning*, 7(3), 1–23.

<https://doi.org/10.29140/iltl.v7n3.1026>.

3. Methodology

3.1. Research design and participants. An exploratory quasi-experimental comparative design was used. The study involved 20 IELTS learners enrolled at one private language school. Participants were aged 16–18, were described as pre-intermediate, and were divided into two groups of 10: an AI-assisted feedback group and a teacher-led feedback group. As random allocation and baseline equivalence were not documented, the design is treated as quasi-experimental and the findings should not be interpreted as definitive causal effects.

3.2. Instrument and assessment criteria. IELTS Writing Task 2 served as the performance task. Participants wrote under a 40-minute limit. Essays were assessed using Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GRA). To strengthen reliability in future replications, scripts should be anonymized, independently double-rated, and accompanied by inter-rater agreement statistics. The current report retains only the aggregate outcomes available in the source document.

3.3. Intervention procedure

Stage	Timing	AI-assisted group	Teacher-led group
Pre-test	Week 0	IELTS Writing Task 2; 40 minutes	Same task format and time
Intervention	Weeks 1–4	16 tasks; ChatGPT comments and model revisions	16 tasks; instructor comments on matched dimensions
Learner action	Throughout	Review and apply AI suggestions	Review, revise, and apply teacher suggestions

Stage	Timing	AI-assisted group	Teacher-led group
Post-test	End of Week 4	IELTS Writing Task 2; 40 minutes	Same task format and time

Table 1. Intervention design

Both groups completed the same number of writing tasks and received feedback on comparable dimensions. The AI-assisted group submitted drafts to ChatGPT and received comments on grammar, vocabulary, organization, and task response, together with modified versions. The teacher-led group received instructor feedback on the same broad areas. Classroom observation indicated that teacher-group learners reviewed and applied comments more consistently, whereas some AI-group learners engaged selectively with lengthy responses and model texts.

3.4. Data analysis and ethical considerations. **The analysis is descriptive. The available evidence comprises aggregate post-test means and classroom observations of writing dimensions and feedback engagement. No individual scores, standard deviations, pre-test means, confidence intervals, or validated questionnaire responses were available. Accordingly, no t test, effect size, or claim of statistical significance is reported. Participant identities are not disclosed. For publication, the author should retain documented consent/assent procedures and institutional permission appropriate to research involving learners aged 16–18.**

4. Results.

The teacher-led group achieved a post-test mean of 7.0, compared with 6.5 in the AI-assisted group. This 0.5-band difference is descriptively relevant but cannot be treated as statistically significant without participant-level data and measures of variability. The available observations further suggest different profiles

of support across writing dimensions.

Writing dimension	AI-assisted group	Teacher-led group
Idea development	More noticeable observed support	Observed improvement
Grammar	Some observed improvement	Clearer observed improvement
Essay organization	Some observed improvement	Clearer observed improvement
Lexical resource	Slight observed improvement	Similar observed improvement
Coherence and cohesion	Observed improvement	Observed improvement
Task response	Observed improvement	Observed improvement
Post-test overall mean	6.5	7.0

Table 2. Descriptive post-intervention outcomes

The AI-assisted condition appeared most useful when learners needed rapid idea prompts or lexical alternatives. Teacher-led feedback appeared stronger in grammar and organization and was more consistently followed by revision. Both groups showed positive classroom observations in task response and coherence and cohesion. These statements describe the reported pattern; they do not establish that either condition caused the changes.

5. Discussion

The descriptive findings align with the view that AI and teacher feedback offer partially different affordances. Generative AI can rapidly expand an idea space and present alternative wording, which may explain its observed value for brainstorming and lexical choice. The teacher, however, can judge what is

developmentally appropriate, connect comments to a learner's recurring difficulties, and limit feedback to a manageable set of priorities. This contextualization may help explain the stronger observed pattern in grammar, organization, and feedback uptake.

The findings also reinforce the distinction between feedback provision and feedback use. Learners in the AI-assisted group did not consistently examine all suggestions, especially when responses were extensive. This classroom observation is compatible with Farrokhnia et al.'s conclusion that abundant generative-AI feedback may not translate into extensive revision. A productive classroom protocol should therefore require learners to classify suggestions, justify which ones they accept or reject, revise selected passages, and compare the new version against IELTS descriptors. A hybrid model is the most defensible pedagogical implication. AI can provide a first-pass response on language and idea options; the learner then evaluates the response; and the teacher verifies alignment with the task, prioritizes higher-order issues, and addresses misconceptions. Such sequencing preserves teacher judgement while using AI to increase practice opportunities. It also reduces the risk that students simply copy a model essay without developing independent control.

6. Limitations and Directions for Further Research

Four limitations constrain interpretation. First, the sample was small and drawn from one language school. Second, random assignment and baseline equivalence were not documented. Third, the report did not retain participant-level pre-/post-test scores, dispersion measures, or inter-rater reliability evidence. Fourth, perceptions were inferred from classroom observation rather than measured with a validated questionnaire or interview protocol. A future study should preregister the design, use a larger sample, report criterion-level scores for both testing occasions, calculate paired and between-group effects with confidence intervals, double-rate anonymized scripts, document the AI model and prompts, and collect systematic learner-engagement data.

7. Conclusion

Within this four-week exploratory study, both feedback conditions were associated with constructive writing practice, but they displayed different observed strengths. AI-assisted feedback was especially useful for idea generation and lexical options; teacher-led feedback was more consistently used and appeared stronger for grammar and essay organization. The post-test mean was 6.5 for the AI-assisted group and 7.0 for the teacher-led group. Because the dataset does not support inferential testing, this difference should be read descriptively. The most responsible conclusion is not that generative AI replaces teacher feedback, but that carefully structured AI use can extend feedback opportunities when learners remain active evaluators and teachers retain a mediating role.

References

1. Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, Article 57. <https://doi.org/10.1186/s41239-023-00425-2>
2. Farrokhnia, M., Latifi, S., Papadopoulos, P. M., Hogenkamp, L., Gijlers, H., Khosravi, H., & Noroozi, O. (2026). Generative AI offers more, but students revise less: Comparing the effects of teacher and AI feedback on student essay revisions. *International Journal of Educational Technology in Higher Education*, 23, Article 6. <https://doi.org/10.1186/s41239-026-00579-9>
3. Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>
4. IELTS. (n.d.). Writing band descriptors. <https://ielts.org/organisations/ielts-for-organisations/test-statistics-and-research/test-performance>
5. Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, Article 9. <https://doi.org/10.1186/s40561-024-00295-9>

6. Poláková, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1), Article 2410101. <https://doi.org/10.1080/2331186X.2024.2410101>
7. Soori, A., Khojasteh, L., & Javed, F. (2025). Comparing teacher e-feedback, AI feedback, and hybrid feedback in enhancing EFL writing skills. *Technology in Language Teaching & Learning*, 7(3), 1–23. <https://doi.org/10.29140/tl.v7n3.1026>