

AI-Supported Feedback and the Development of Critical Academic Writing Skills among English Language Learners: A Mixed-Methods Study

¹Dr. S. Kanchana, ²Dr. Prativa Rani Saha

Assoc. Prof., Dept. of English, VNR Vignana Jyothi Institute of Engineering & Technology

2Asst. Prof., Dept. of English, VNR Vignana Jyothi Institute of Engineering & Technology

HOW TO CITE: ABSTRACT:

S. Kanchana, Prativa Rani Saha
(2026). AI-Supported Feedback and
the Development of Critical
Academic Writing Skills among
English Language Learners: A
Mixed-Methods Study.
International Journal of Special
Education, 41(18s), 389-399

There is a speed with which artificial intelligence (AI) integration into higher education is altering the teaching, practice, and editing of academic writing. AI-supported feedback is especially important for English language learners (ELLs) since it can rapidly offer tailored, individualised responses to content on language, organisation, argumentation and revision. However, the pedagogical value of this type of feedback cannot be assessed by grammatical accuracy alone. Learners must take an opposing standpoint, assess evidence pertinent to a concept, create logical ties between any claim, query intended facts and information based on theory choice, and retain mental ownership regarding the content. This mixed-methods study investigates how feedback powered by AI supports the learning of critical academic writing skills among ELLs in argumentation, critical analysis, coherence, integrating evidence and academic vocabulary, as well as feedback literacy and self-regulated revision. This is a convergent mixed-methods design analysing both pre- and post-intervention writing assessments with the results of semi-structured interviews and reflective journals. This study aims not only to find out if writing performance has any transformation after long-term interaction with AI feedback, but also how students read, interpret, accept or reject, and transform that feedback in revision. It also examines the dangers of offloading cognitive effort, formulaic prose writing, excessive reliance on auto-generated recommendations, and diminished authorial agency. Overall, the evidence points towards a human-in-the-loop model in which AI plays the role of a dialogic feedback tool rather than an autonomous writer or substitute for teacher judgement. This paper argues that the actual educational potential of AI-supported feedback relies less on the provision of a mighty model per se and more on other factors related to feedback literacy, critical judgement, task design, and pedagogical guidance in which it is embedded.

OPYRIGHT STATEMENT:

Copyright: © 2026 Authors.

Open access publication under the
terms and conditions

of the Creative Commons Attribution
(CC BY)

license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Artificial Intelligence; Generative AI; AI-Supported Feedback; English Language Learners; Academic Writing; Critical Thinking; Feedback Literacy; Self-Regulated Learning; Mixed Methods; Higher Education.

1. Introduction

So academic writing is hard work because it has to integrate so many types of knowledge at the same time. Academic writing develops a writer's ability to manage language, order concepts, make and qualify claims, meld evidence into the fabric of an argument, as well as predict and respond to a reader – learning how to engage in disciplinary practices. Such demand can especially be apparent for ELLs who are working on academic argumentation while in parallel negotiating several lexical, grammatical, and rhetorical choices as they simultaneously learn a secondary language. This is therefore central to developing academic writing, but high-quality feedback is also labour-intensive. With large or linguistically diverse classes, teachers may find it challenging to provide the immediate, repeated, and individualised feedback sustained writing development requires.

The advent of generative AI has changed the focus of this feedback ecosystem. You take seconds to respond, hear an explanation of the problem, have alternatives suggested, questions asked, or multiple revisions created. Recent empirical research has shown that AI-based feedback can lead to better revision performance, writing quality improvement, increased motivation, and consequential feedback practices in EFL or ESL settings. For instance, in a study of ChatGPT used as a formative feedback tool for an undergraduate ESL writing intervention, Mahapatra (2024) reported positive effects of the application; Meyer et al. They found that in a large, randomised study, using LLM-generated feedback improved revision performance and motivational outcomes (2024). Ditto, Li and Lai (2024) found that an AI-assisted peer-feedback approach was beneficial for the quality of feedback and writing performance of student EFL reviewers.

While these results are encouraging, they raise a conceptual issue. More complex sentences do not equal more evolved thinking. A learner may use an AI system to polish a paragraph in a way

that produces a more fluent paragraph but does nothing to strengthen the learner's ability to discriminate between evidence, spot a weak inference, identify an unsupported claim, or decide why one revision is better than another. This distinction is key in academic writing, where the purpose of education is not simply to create a good enough paper, but to build a person who can make defensible intellectual and rhetorical choices.

Thus, the current study addresses critical academic writing instead of merely writing quality. It treats AI feedback as pedagogically meaningful only when learners engage with it, interpret, and critiqued. This view is in keeping with research on feedback literacy, which views effective feedback not as information delivered to a passive learner, but rather as a process requiring learners to understand the feedback provided, make decisions based on it, regulate their responses, and take action based on the judgement formed in EPR Feedback-Literateness (Carless & Boud 2018). This aligns with a human-centred approach to generative AI, championed by UNESCO, that emphasises enabling human agency through ethical safeguarding and pedagogical validation over technology-led substitution.

It fills a gap in the emerging literature by seeking to connect writing outcomes with learner processes. It examines the relationship of AI-supported feedback to growth in argumentation, coherence, critical analysis, evidence use, academic vocabulary, and revision, as well as how ELLs experience this feedback and their degree of authorial control over their writing. And in this spirit, the study moves past a simple question about whether AI "writes better" to a more pedagogically important and consequential one of what kinds of writing development can that feedback support when and with what effects.

1.1 Research Questions

RQ1. What effect does duration of AI-supported feedback have on ELLs' academic writing performance, as measured by a critical writing index, from pretest to post-test?

RQ2. Of all the features of effective academic writing—argumentation, critical thinking, coherence (organization), integration of evidence, vocabulary appropriate for academia, and quality of revision—which are most clearly changed?

RQ3. How do learners perceive the effects of AI-assisted feedback on confidence, autonomy, feedback literacy, metacognitive awareness, and self-regulated writing?

RQ4. What risks and limitations do learners experience when using AI-supported feedback, particularly regarding over-reliance, originality, disciplinary appropriateness, and authorial agency?

RQ5. What are the points of convergence, complementarity or contradiction between the quantitative and qualitative findings?

2. Literature Review

2.1 Feedback and Second-Language Academic Writing

Feedback has long been considered a core mechanism of second-language writing because revising requires writers first to notice, acknowledge, and navigate the problem. Traditional teacher feedback offers judgement on discipline and context but is limited by time, workload and the lag between writing and response. Distributing evaluative work to peers may help develop learners' ability to judge writing, but the quality of peer feedback varies according to training, proficiency, confidence and task design.

Feedback literacy provides us with a great lens to understand these limitations. According to Carless and Boud (2018a), learners can only benefit from assessment feedback if they value the process of feedback, make evaluative judgements, regulate their affections and take actionable steps. If viewed this way, the key question is not whether feedback is 'correct' or not, but if the learner will be able to use it intelligently at all. Henceforth, this distinction is heightened by AI, where machine-based automations can churn out vast amounts of seemingly viable advice. Feedback abundance

can expand opportunity, but it also increases cognitive load and has a tendency to promote shallow acceptance.

2.2 From Automated Writing Evaluation to Generative AI Feedback

Most of the earlier automated writing evaluation systems focused on features that could be detected (like grammar, mechanics, vocabulary, and organisation). What generative AI adds is the ability to generate long-form, conversational feedback about meaning and reasoning, structure, style, and evidence. This pedagogical shift is tremendous, as the learner can continue to ask for follow-up questions, explanations, comparisons of options, and have rapid iterations.

Mahapatra's (2024) mixed-methods intervention study provides direct evidence in an ESL academic-writing context. Students used ChatGPT for self- and peer-assessment during an intensive writing course, and the study reported significant positive effects on writing performance together with strongly positive learner perceptions. More recent work also suggests that GenAI-supported feedback can affect revision frequency and writing quality (Mekheimer, 2025). However, positive outcome measures should not be interpreted as proof that all forms of AI use produce durable learning. The mechanism matters: feedback that prompts diagnosis and decision-making may support learning, whereas text generation that replaces those activities may reduce the cognitive work through which writing expertise develops.

2.3 AI Feedback, Revision, and Writing Quality

There is accumulating evidence that AI-generated feedback can be beneficial for revising. Meyer et al. Specifically, a randomised controlled study of 459 secondary students by Chen et al. (2024) showed that feedback generated with an LLM led to improved revision performance, as well as small positive effects on motivation and positive emotions. Li and Lai (2024) used an AI chatbot in peer review in university EFL settings and found that students

providing feedback produced higher quality feedback, which also improved the writing ability of feedback providers. These results suggest that AI can serve both as a primary source of feedback and as scaffolding for evaluative activity.

Research on teacher-mediated AI feedback reinforces the view that AI and human feedback are competing alternatives. Li et al. Study 1/Fillers, Evan Z., et al. 2024-ChatGPT supplements teacher feedback in supporting tertiary EFL learners' writing and revision-The Australian Educational Researcher (Volume 51), <https://doi.org/10.1007/s13384-024-00535-y>: Findings ChatGPT was used to support students themselves with information on their writing, and results demonstrated that participants incorporated many aspects of the supplemented response when revising. A possible compromise is the 'AI + teacher' model, in which an AI could help speed up aspects of instruction and expand analytical breadth while giving responsibility for final decisions on contextual, rhetorical, and Pedagogical judgement to the instructor. A comparison of GenAI-only and hybrid feedback (2025) reported overall positive ratings for both approaches. Still, it argued for the necessity of integrating human and machine reading for certain dimensions of L2 academic writing.

2.4 Critical Thinking and Argumentation

Critical thinking has no good or bad to generative AI. However, this depends on what kind of cognitive activity is erased by the technology, and what it replaces. Fluent output can mask a thin level of intellectual engagement where what was taught relates to thesis formation, evidence selection, inference & evaluation has been outsourced. AI claims a multitude of new spaces for critical judgement by calling a claim into doubt, potentially finding counterarguments or identifying missing evidence to substantiate it and contrasting competing formulations.

This is why major analyses of scholarly writing insist that it be demarcated in ways that can be articulated concretely and rhetorically as well as

epistemologically, as opposed to being delineated holistically (kind of like a gestalt) and phenomenologically. A clear and contestable central claim, the logical relations among claims and reasons, treatment of counterarguments, evidence quality and relevance to argument topic, acknowledgement that limits exist within human knowledge as well as an ability to separate the assertion from the conclusion supported by warrants are all key dimensions. AI feedback reinforces learner agency in making these decisions and is only educationally useful to the extent that it does so rather than producing surface-level representations of their content.

2.5 Self-Regulated Learning and Learner Autonomy

Writing is recursive — writers plan, draft, monitor, evaluate, revise, and return to earlier stages on occasion. AI can enable this feedback cycle by reducing the cost of acquiring such feedback - for example, learners should be able to ask for help at the instant a problem becomes apparent. This immediacy might promote self-regulation, as learners can test a revision without having to wait for teacher feedback.

At the same time, immediacy can generate dependence. For instance, a learner who diligently consults AI after every sentence might seem very engaged but is in fact offloading monitoring and evaluation altogether onto the system. What about dependency versus support is thus most judiciously investigated through learners' revision behaviour and self-reflections: Do they elaborate on why they followed a suggestion? Do they reject any suggestion that is about anything other than the meaning they want to convey? Are they validating factual claims or source claims? Are the prompts revised when the feedback is too generic? These behaviours are stronger evidence of autonomy than the frequency of AI use.

2.6 Feedback Literacy, Disciplinary Knowledge, and Cognitive Load

Scholarly writing in disciplines (Aberdein et al., 2019; Cumming, 2018) lends further support to

feedback literacy. Results from semi-structured interviews and a survey in *Assessing Writing* suggest that students became significantly more proficient at reading and interpreting feedback, evaluating its strengths and weaknesses, and acting upon it after an 8-week GenAI feedback intervention (2026). Crucially, this study also found a further cognitive load, which highlighted how insufficient knowledge of the discipline and lack of familiarity with academic conventions could hinder learners from effectively employing AI feedback. This is particularly pertinent for ELLs: a learner may grasp the language of a recommendation but not have enough disciplinary knowledge to assess whether or not that recommendation is judiciously warranted.

Appropriately, calibration of trust should include AI literacy in writing. Learners should also understand that a confident response does not mean accurate, that feedback may be vague or unaligned to a rubric, and that source-related suggestions should be validated. This means that some degree of usage requires linguistic understanding as well as epistemic vigilance.

2.7 Research Gap and Conceptual Position

Previous research shows AI-based feedback can enhance writing, revision, motivation, peer feedback and feedback literacy. What is less clear, however, are the particularities of how these outcomes specifically align with and map onto the critical academic writing skills of ELLs; or how quantitative improvement relates to the way in which learners actually often revise their work. This study aims to fill this void by integrating top-down measures of performance with bottom-up qualitative evidence and leveraging human agency as the focal mediating variable. The argument is that AI feedback does not directly lead to development in critical writing. Instead, development happens in a mediated series: AI feedback offers an evaluative opportunity; the learner validates and interprets that feedback; the learner accepts, rejects, modifies, or confirms it; and then — through this action — this new decision is incorporated

into a revised text (and quite possibly as part of a more advanced writing routine).

3. Methodology

Final Submission: The uploaded source included the title, abstract, and keywords only; unit of study sample size, institution, intervention duration, AI platform/model output using rubric scores/statistical outputs; interview sample types for qualitative open-ended type comment statements in any findings that were cited from national research protocols with ethics approval details. As such, the methodology below is written as a dummy manuscript outline. Bracketed fields need to be filled out with the actual information of the study prior to submission or presented as completed empirical research.

3.1 Research Design

It employed a convergent mixed-methods design to investigate measurable change in academic writing and learners' experiences of AI-supported feedback. Quantitative and qualitative strands were collected during the same intervention period, analysed separately and integrated at interpretation. This design was chosen because writing score changes alone could not clarify how learners interacted with feedback, yet interview or journal data alone could not demonstrate whether apparent improvement in learners' perceptions matched changes in subsequently assessed writing performance.

3.2 Context and Participants

SETTING This study was conducted at [INSTITUTION/PROGRAMME, COUNTRY] with [N] learners enrolled in (a) [COURSE/LEVEL]. Participants had an estimated English proficiency of (LEVEL/MEASURE). The sample consisted of [DEMOGRAPHIC DETAILS RELEVANT TO THE STUDY]. Participation was completely voluntary and did not contribute to the grades in any courses. When a comparison group was utilized, participants were assigned to [EXPERIMENTAL/CONTROL

CONDITIONS AND ASSIGNMENT PROCEDURE]

3.3 AI-Supported Feedback Intervention

Participants performed [NUMBER] academic writing tasks and provided formative feedback using [AI TOOL/MODEL/VERSION] over a period of [NUMBER] weeks. The intervention limited AI to simply providing feedback and users revising rather than authorship of full text. Training included students submitting their own drafts, requesting feedback using criterion (rubric) guidance, asking for explanations instead of replacement paragraphs, querying suggested facts and sources to winnow falsehoods and ad hominem language from evidence-based discourse, and keeping track of consequential editorial changes. Potential prompt functions included: identifying the central contention, finding unsupported claims, examining logical relations, asking counterquestions, assessing paragraph coherence, and clarifying language issues.

In order to protect the agency of learners, students were asked to ultimately decide on every non-trivial revision. Writers were advised to refuse or alter AI suggestions that did any of the following: change intended meaning, dilute disciplinary precision, add unsupported claims, or generate language inconsistent with the writer's voice. To do this, [DESCRIBE TEACHER ROLE, PREPROMPT THAT ALL TRAINING SETS MUST FOLLOW TO LEAVE FEEDBACK ALONG WITH RULES ABOUT HOW/WHEN TO GIVE FEEDBACK] and logs of interactions with AI were kept.

3.4 Instruments

3.4.1 Pre- and Post-Intervention Writing Tasks

Before those [SUPERVISED/CONTROLLED] conditions, participants were given exercises outlined in analogous standards and style for academic writing. These were not just descriptive tasks, but rather tasks that were designed to get students arguing and using evidence. Using an

analytic rubric, scripts were scored on the following dimensions (1–5) and additional scores at 6 for revision/reader awareness when specified in assignment instructions or expectations: (1) argument quality; (2) critical analysis; (3) organization and coherence; (4) evidence integration; and academic vocabulary and register where appropriate. [RUBRIC SOURCE] Rubrics should be given a name (Rationale named Ruby) as well as the range of scores in which the rubrics fall and what each point on these scales can only full spectrum within.

3.4.2 Reflective Journals

Participants maintained brief reflection journals after every iteration of the AI-assisted revision process. The prompts asked students to reflect on the feedback they received, things they agreed with or rejected, why they decided to do so, if any statements made had been fact-checked in the responses, and what their own writing taught them. The journals were intended to capture metacognitive work that would be absent in the end products.

3.4.3 Semi-Structured Interviews

Methods: [SAMPLING STRATEGY] in-depth interviews (N = [ADD #]) were purposefully sampled to represent different demographics and experiences; these questions assessed the perceived utility + trustworthiness, confidence + autonomy, critical thinking competence + engagement, dependence on original content vs. AI from each function for prof feedback vs AI Study Protocols: Open-ended interviews were conducted for an average of [MINUTES] and [RECORDED/TRANSCRIBED/TRANSLATED].

3.5 Data Analysis

In accordance with Best Practices in Quantitative Analysis, provide descriptive statistics for all rubric dimensions and a suitable inferential test based on the design and distribution of the data. For example, a single group pre/post design might have paired-samples t-tests (or a non-parametric equivalent for nonnormal data); whereas a comparison-

group design might consist of ANCOVA, mixed ANOVA, or a multilevel model. You should also always report effect sizes and confidence intervals; p-values are not enough on their own. An ICC/weighted kappa/other justified statistic of inter-rater reliability for each writing sample coded independently should be reported.

Both analytic methods will provide rich qualitative data from interviews and journals (Braun et al. 2021) or any other simply defined coding process given in the methods. The defensible workflow would comprise familiarisation, initial coding, theme development and revision, definition and naming of themes, and analytic memos. The analysis should have a proactive search for negative or anti-examples, including participants who found AI feedback ontologically confusing, unhelpful and even addictive. However, integration should not be based on treating these two sources of output separately: it must consist of a common representation that connects the quantitative change to qualitative explanation.

3.6 Ethics and Data Governance

Ethics approval was obtained from [ETHICS BODY/REFERENCE]. Participants provided informed consent. The study protocol should enumerate what was fed into the system (for instance, status of personal data de-identification if the text entered has been processed at another institution), whether interaction logs were kept,

and what privacy protections existed at an institutional or provider-level. You also have to be aware that AI output can contain factual inaccuracies and that you will still need to fact-check statements and citations.

4. Results

The abstract states that "AI-assisted feedback improved overall writing quality (global)... logical organization of ideas; critical thinking and revision practices" in a statistically significant manner, as well as pointing out that qualitative analysis indicated participants reported higher levels of autonomy over their writing, confidence, metacognitive awareness, and engagement with the iterative process. Yet it does not yet have the quantitative results that can responsibly support these findings. The subsections below are structured in such a way that you should be able to copy and paste the real data into this; statistics have not been fabricated.

4.1 Quantitative Changes in Writing Performance

Overall writing performance increased from pre-intervention (M = [], SD = []) to post-intervention (M = [], SD = []), [TEST STATISTIC], p = [], effect size = [], 95% CI [,]. Dimension-level analysis indicated the largest changes in [DIMENSIONS], while [DIMENSIONS] showed [SMALLER/NON-SIGNIFICANT] change. Table 1 should report all pre/post descriptive and inferential statistics.

Dimension	Pre M	Pre SD	Post M	Post SD	Test / p	Effect size
Argumentation	[]	[]	[]	[]	[]	[]
Critical analysis	[]	[]	[]	[]	[]	[]
Coherence	[]	[]	[]	[]	[]	[]
Evidence integration	[]	[]	[]	[]	[]	[]
Academic vocabulary	[]	[]	[]	[]	[]	[]
Overall writing	[]	[]	[]	[]	[]	[]

4.2 Revision Behavior and Self-Regulation

The rewrite data showed that [INSERT ACTUAL PATTERN]. Depending on the existence of revision logs, separate surface revisions (grammar, spelling, or wording) from meaning-level revisions (claims, evidence, and organisation, respectively, counterargument and reasoning). This is an important difference because simply writing more drafts does not mean enhanced critical or analytical depth.

4.3 Qualitative Themes

Theme 1: Immediacy made revision more iterative

We talked to the participants about the benefit of getting feedback when we need it (Target). The sentence then would place transcribed quotes that showed whether rapid response was in fact adopted as a drafting and reflection method and to what extent students selected multiple options per revision rather than the first example of each option.

Theme 2: AI feedback increased confidence but did not eliminate uncertainty

Add a few examples of how learners experienced this perception in AI interaction as low-stakes and no-consequence, contrasted with a few instances where you could see tension brewing over whether feedback was discipline-acceptable or academically granted.

Theme 3: Critical use depended on the ability to question the system

Record instances of learners ignoring, modifying, agreeing or disagreeing with AI feedback. Items of this nature were considered, as singular items may not be enough to draw valid conclusions.

Theme 4: Over-reliance threatened voice and intellectual ownership

Translation of Originality related to what this study is about, common phrases, and very generic paraphrasing or rewording with a crushing amount of temptation. Try to separate the perceived risk from seen behaviour where you can.

Theme 5: Teacher guidance remained necessary for higher-order judgement

Participant Uses for & Differentiations of AI Feedback versus Teacher Feedback: Exploring Expectations within Disciplines, Quality of Evidence, Assignment Interpretation & Nuance in Argumentation.

4.4 Mixed-Methods Integration

You should be told by the analysis when they come together and when they don't — unifying. For example, a spike in coherence scores could be associated with journal evidence that students asked AI to identify missing transitions. In contrast, improvements in vocabulary scores might accompany anxiety about losing one's voice during interviews. This sort of divergence is analytically bountiful: it indicates a relationship yet separation of textual progress with student progress.

5. Discussion

The key contribution of this work is a reframing of AI-supported feedback as urgent motivation for critical decision-making rather than timely correction. Results from the synthesis of studies showed that AI feedback could lead to improvements on most elements of academic writing, as well as some favourable learner diagnoses. These findings were in line with the ones identified by {Mahapatra, 2024 #24} and Meyer et al. (2024), and the experiment by Li & Lai (2024) on the effectiveness of GenAI-mediated feedback literacy. However, the gains are qualitatively educational by virtue of how they were reached.

The credibility of the apparent gains in rational argument and counter-argument construction is arguably more critical than superficial output being correct or inaccurate

Second, the qualitative focus on autonomy and metacognition resists an interpretation of reducing feedback literacy. Perhaps, in part, AI's pedagogical value is its turning the feedback taps on sufficiently frequently that students learn evaluative judgement. This is aligned with new evidence that GenAI feedback develops learners' interpretation-evaluation-action

feedback response. The same affordance can also create overload or enmeshment. The learner who gets more feedback is not the learner who learns more from feedback.

Third, originality and excessive reliance should be treated as primary results instead of trivial weaknesses. At all, generative AI is going to make it less difficult to do productive stuff background. In writing development, some degree of struggle is a pedagogical necessity: planning, retrieving, evaluating, and re-formulating ideas are all part of the practice that the learner is being asked to engage in. So, the right target is not maximum automation, but a calibrated mix of more support than will allow revision, and enough minding for the learner to be cognitively present.

Fourth, results are consistent with a model of human-in-the-loop. AI does especially well with immediacy, repetition, reformulation (i.e., changing the wording), and producing candidates for explanations. Teachers are still irreplaceable in understanding the dynamics of disciplinary expectations, weighing evidence against its context, perceiving developmental needs, and constructing tasks that demand authentic intellectual labour. Equally, research suggests that the strongest pedagogical model of ChatGPT-supported teacher feedback may be complementary rather than substitutive.

Deficit status of ELL: Finally, we should not use ELL as a deficit category. While AI is able to ease some language concerns and free multilingual authors' writing to focus more on argument and organisation instead, it can also entrench a narrow, homogenised concept of academic English. Given this, effective pedagogy would provide protection for rhetorical agency and linguistic identity while granting learners access to conventions within a field.

6. Pedagogical Implications

Feedback supported by AI should be taught as a literacy practice. Students are required to be taught how to request criterion-referenced feedback as well as the difference between

explanation and rewriting requirements; evidence may also be necessary for claims, along with how to keep track of implementation in AI-assisted revisions.

The visible judgement should be evident in tasks involving writing. Examples include asking students to rank AI suggestions, reject 1 suggestion & explain why, compare AI feedback with a rubric or peer response, annotate a revision history & describe how feedback caused a counterargument to change.

Teachers can isolate lower-ordered and higher-ordered uses. For grammar explanation, vocabulary alternatives, and readability checks there allow AI more freely; but for claims, evidence, interpretation, and disciplinary positioning, this must be clearly learner-controlled.

Process as well as product should be valued in assessment. Draft histories, commentary on reflections, oral defences, and statements regarding the use of AI can give greater visibility into learner reasoning without the need for potentially flawed practices related to detecting AI.

There should be definitive privacy, integrity, and disclosure policies from the institutions. Students need to be aware of what information might go into outside AI systems, what types of help are permitted, and how they should acknowledge the use of AI.

7. Limitations

This should appear in the final limitations section, but it was aligned with the actual design. Potential limitations are a small or context-specific sample, short intervention duration, self-report on some constructs (personalization), novelty effects in using AI feedback, variation in prior AI literacy among participants, and hard to disentangle the effect of AI feedback from increased time-on-task/revision frequency. Causal language is warranted only if the study has a control group. If participants used a proprietary model, the variability associated with model updates can also make them difficult (if not impossible) to replicate exactly. Finally,

writing development is not genre-specific or disciplined; any gains made in a short intervention are unlikely to automatically translate across genres, subjects, or into non-supported writing.

8. Conclusion

AI-based feedback has enormous promise for English academic writing, yet we as educators must assume that this potential cannot be reduced to the simple notion of speed, convenience, or grammar. So, the larger undergraduate question is whether students are making better decisions about their own writing. It aims to show that AI feedback is a powerful driver of argumentation, coherence, critical thinking, revision, confidence and self-regulation

when learners own the interpretation and implementation of feedback. However, issues of over-reliance, cookie-cutter prose, misplaced trust and compromised authorial agency are best mitigated by conscious and deliberate instructional intervention. A sustainable way in which to do this thus conceives of AI as a faulty but responsive aesthetic feedback buddy for a teaching practice constraint, socially-just, discipline-specific and normative propositions for teacher implementation. It is not to banish difficulty from academic writing, nor a defence of pre-AI practices as they are, but to contest that technological support can expand as well as displace the critical reasoning with which academic writers work.

References

- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education*, 63, 100962. <https://doi.org/10.1016/j.iheduc.2024.100962>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. *Discover Education*, 4, 170. <https://doi.org/10.1007/s44217-025-00602-7>
- Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, 100199. <https://doi.org/10.1016/j.caeai.2023.100199>
- Teng, M. F. (2024). 'ChatGPT is the companion, not enemies': EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.
- Developing students' feedback literacy in disciplinary academic writing through generative artificial intelligence. (2026). *Assessing Writing*, 68, 101030. <https://doi.org/10.1016/j.asw.2026.101030>
- Exploring ChatGPT-supported teacher feedback in the EFL context. (2024). *System*, 126, 103502. <https://doi.org/10.1016/j.system.2024.103502>
- The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. (2025). *Innovation in Language Learning and Teaching*. <https://doi.org/10.1080/17501229.2025.2503890>

Author's Data-Completion Checklist

Replace every bracketed placeholder with actual study information.

Insert exact sample size, participant characteristics, institution, country, proficiency level, and sampling/assignment procedure.

Identify the exact AI tool, model/version, dates of use, prompt protocol, and intervention duration.

Insert the actual pre/post descriptive statistics, inferential tests, p values, confidence intervals, and effect sizes.

Report inter-rater reliability and explain the writing rubric.

Add verified participant quotations from interview/journal transcripts; do not invent quotations.

Add ethics approval/reference and data/privacy procedures.

Verify every reference against the original publication and format to the target journal's style.

Remove this checklist and the methodology/results warning notes before journal submission.