AI-Supported Iterative Revision and Reflective Writing in Primary Education: A Quasi-Experimental Study

Xiaoqing Xian[1], Jinli Wu[2], Xiao-Fan Lin[3*]

[1] South China Normal University, School of Educational Information Technology

[2] South China Normal University, School of Educational Information Technology

[3] South China Normal University, School of Educational Information Technology

*corresponding.author@South China Normal University.China

Abstract: *This study examined whether an AI-supported iterative revision workflow can improve primary students' writing development in an authentic school setting. A quasi-experiment was conducted over one academic year with 259 Grade 5 students in a Shenzhen public school (experimental n = 85; control n = 174). The experimental classes used an AI-supported writing platform that provided immediate analytic feedback, revision prompts, and portfolio records, while the comparison classes received conventional teacher-led feedback. Because the intervention bundled AI feedback with more iterative revision opportunities, the study estimates the effect of an AI-mediated revision ecology rather than an isolated AI-only effect. Outcomes were analyzed with ANCOVA on five questionnaire dimensions (context, content, process, strategies, and feedback) and with Epistemic Network Analysis (ENA) on rubric-based writing performance. After controlling for pre-test scores, the experimental group obtained significantly higher post-test scores on overall writing literacy and all subdimensions, with the strongest gains in content development and feedback engagement. ENA further showed that the experimental group displayed denser connections among content, structure, and norms, indicating better integration of higher-order writing dimensions. The findings suggest that AI-supported revision can be pedagogically valuable when used to scaffold reflection, repeated drafting, and data-informed teacher guidance. The paper discusses both the promise of this intervention package and the limits of causal claims when revision dosage differs across conditions.*

Keywords: AI-supported writing, reflective revision, primary education, feedback, ENA

1. Introduction

As one of the foundational skills, writing is regarded as a main skills for students' academic performance and active participation in society. To write effectively, learners must be able to generate and organize meaningful content, apply accurate language mechanics, and adapt their discourse to suit specific contexts and audiences. Argumentative writing, in particular, has been shown to promote scientific thinking, stimulate higher-order reasoning, and support deeper processing and organization of information (Bal et al., 2025; Liu et al., 2024). Integrating writing across the curriculum further enables students to internalize disciplinary knowledge, synthesize diverse perspectives, and consolidate conceptual understanding. However, many learners, especially those writing in a foreign language, exhibit limited reflective awareness, leading them to overlook essential lexical and syntactic elements, which undermines the coherence and quality of their compositions (Shen & Teng, 2024). Encouraging reflection on the writing process can foster higher writing skills, enhance self-regulation, and ultimately improve both the quality of writing and overall learning outcomes (Bal et al., 2025).

However, writing feedback often remains superficial, focusing predominantly on surface-level corrections such as grammar and spelling rather than providing strategic guidance on idea development or writing skills. While written feedback is widely acknowledged as valuable for developing academic writing skills, its effectiveness is limited when it lacks clarity or fails to include actionable feed-up, feed-back, and feed-forward elements (Schillings et al., 2023). Consequently, many students apply corrections mechanically without fully understanding how to improve, revealing a critical gap in reflective learning. Research indicates that self-reflection, particularly when linked to peer feedback, can foster deeper engagement with the revision process and lead to substantive improvements in both content and language (Li & Hebert, 2024). Yet, without explicit instructional scaffolds, students rarely engage in timely, structured reflection on their writing (Li & Hebert, 2024). Targeted interventions, such as guided revision plans and reflective prompts, can encourage learners to consider feedback at the process and self-regulatory levels, enabling them to set goals, plan revisions, and internalize strategies for future writing tasks.

Recent advances in generative artificial intelligence (Gen AI), particularly large language models (LLMs) such as GPT-4, have markedly reshaped the field of automated writing evaluation (AWE). Traditional automated essay scoring (AES) systems, though effective at detecting surface-level features such as grammar, lexical diversity, or word count, have been widely criticized for their inability to capture deeper dimensions of writing, including coherence, argument structure, and idea development (Pack et al., 2024). In contrast, LLMs are not bound by pre-defined, rigid scoring rubrics; instead, they can dynamically adapt to human-like evaluation criteria through prompt engineering and contextual reasoning (Pack et al., 2024). This flexibility enables them to provide holistic, context-aware, and individualized feedback that extends beyond error detection to address higher-order writing skills. Emerging empirical evidence further suggests that models like GPT-4 can produce feedback that is not only accurate and aligned with expert judgments but also more detailed, readable, and pedagogically supportive than that provided by many human teachers(Pack et al., 2024). Such qualities make LLM-generated feedback particularly well-suited for promoting reflective learning processes, helping students to internalize revision strategies and engage more critically with their writing skills.

Despite these promising developments, important gaps remain in the application of AI to writing education, particularly for younger learners in primary school contexts. While AI-generated feedback has demonstrated potential to move beyond surface-level corrections, less is known about how such feedback can be integrated into reflective cycles of learning where students actively evaluate, plan, and revise their writing. Existing research has largely focused on accuracy and scoring alignment, whereas fewer studies have explored its role in scaffolding reflective writing and strategy use. For primary students, who often lack the self-regulatory skills required to engage deeply with revision, the challenge lies not only in receiving feedback but also in transforming it into reflective practice.

This paper reports a school-based quasi-experiment on an AI-supported writing platform for Grade 5 students. In response to reviewer concerns, the manuscript has been substantially condensed and the interpretation has been tightened. We do not claim that AI alone outperformed teacher feedback under equivalent conditions. Instead, we examine the effectiveness of an AI-supported iterative revision workflow in which automated feedback, portfolio records, and teacher oversight were combined into a classroom intervention package. Two questions guided the study: (1) Does the intervention improve students' writing literacy outcomes compared with conventional instruction? (2) Does it change the structural relations among rubric dimensions of writing performance?

2. Method

2.1. Participants and design

The analytic sample comprised 259 Grade 5 students from a public primary school in Shenzhen after data cleaning and matching of pre/post records. The experimental group included 85 students and the control group included 174 students. The study lasted one academic year and used intact classes, so the design was quasi-experimental rather than randomized. Both groups followed the same curriculum and completed comparable writing tasks. The difference was in the feedback ecology: the experimental classes used an AI-supported platform for immediate scoring, sentence-level comments, revision prompts, and portfolio tracking, whereas the control classes received conventional teacher-led marking and classroom feedback.

Because the experimental condition combined AI feedback with repeated revision opportunities, the intervention should be understood as an AI-mediated process package. This framing addresses the dosage concern raised in review: the study evaluates a realistic instructional ecology rather than a pure AI-versus-teacher contrast.

2.2. Measures and analysis

Writing literacy was measured with a five-dimension questionnaire covering writing context, content, process, strategies, and feedback engagement. Writing performance was also rated with a four-dimension rubric (content, expression, structure, norms). Pre-test and post-test questionnaire reliability was acceptable to strong (Cronbach's alpha = .771 and .967). ANCOVA was used to estimate post-test group differences while controlling for pre-test scores.

ENA was used as a complementary analysis rather than a substitute for regression. Regression-based models are appropriate for estimating mean differences in outcomes, which is why ANCOVA was the primary inferential test. ENA was added because the second research question concerned how writing dimensions co-occurred as a connected system within students' texts. ENA makes those relational patterns visible by modeling the strength of ties among rubric dimensions and by testing whether the groups occupy different positions in network space. In other words, ANCOVA answers whether students improved more, while ENA helps explain how the organization of writing performance differed across conditions.

Table 1

Key post-test differences between groups

Outcome	Adjusted mean Exp.	Adjusted mean Ctrl.	Statistic	Interpretation
Overall writing literacy	3.930	3.404	F = 30.372***	Experimental group higher
Writing context	3.917	3.423	F = 22.533***	Better audience/context awareness
Writing content	4.065	3.347	F = 40.553***	Largest gain in content development
Writing process	3.896	3.402	F = 22.141***	More process-oriented revision
Writing strategies	3.737	3.392	F = 13.527***	Higher strategy use
Writing feedback	4.001	3.455	t = 5.149***	Stronger feedback engagement
ENA centroid difference	-0.11 vs 0.04	--	t = -3.05***, d = 0.50	Different network structure

3. Results

Table 1 shows that the experimental group scored significantly higher than the control group on overall writing literacy and all five questionnaire dimensions after controlling for pre-test differences. The strongest statistical advantage appeared in writing content, followed by overall literacy, context, and process. The writing feedback measure also favored the experimental condition, suggesting that students in the AI-supported environment became more willing and able to use evaluative comments during revision. Across analyses, however, effect sizes were generally small. This pattern implies that the intervention produced consistent but moderate educational gains rather than dramatic transformation.

The ENA results extended the mean-difference findings. The first rotation axis explained the largest share of variance, and the group centroids differed significantly on MR1, $t = -3.05$, $p < .001$, $d = 0.50$. Network comparisons indicated that the experimental group showed stronger ties among Structure-Norms (.36 vs. .13), Content-Structure (.22 vs. .07), and Content-Norms (.25 vs. .14). These denser higher-order linkages suggest that the intervention did not simply improve isolated rubric scores; it also promoted more integrated writing in which idea development, organization, and conventions worked together more coherently.

4. Discussion and conclusion

Two conclusions can be drawn cautiously. First, the AI-supported iterative revision workflow was associated with better writing outcomes than conventional instruction in this setting. The consistency of the results across overall literacy, subdimensions, and ENA strengthens confidence that the intervention had pedagogical value. Second, the mechanism is best interpreted as reflective revision support rather than AI automation alone. Immediate comments, repeated drafting, visible portfolio records, and teacher mediation likely worked together to encourage students to revisit and reorganize their texts. This interpretation aligns with prior studies showing that feedback becomes more powerful when learners actively use it to plan and revise (Li & Hebert, 2024; Neshaei et al., 2025). The paper also clarifies why ENA was retained. ANCOVA established that the groups differed in mean outcomes, but ENA addressed a different question by revealing how dimensions of writing performance were interconnected. This was important because writing quality is multidimensional, and higher-quality compositions are often characterized by coordination among content, structure, and norms rather than improvement on a single isolated trait.

The results from different empirical evidence sources were mutually corroborated (triangulation), and students who received AI-supported methods performed significantly better than other students under the traditional teaching feedback model. Theoretically, the study contributes to a broader epistemic role on the role of artificial intelligence in educational assessment. Not noly AI as a tool for providing timely corrective feedback (e.g., grammar, vocabulary, and surface-level accuracy), but also the ENA results in this study highlight that AI-supported approach restructured the relational patterns of students' writing, strengthening connections between content development, structural organisation, and linguistic norms. This indicates that AI-supported system acted as a focusing device, directing learners' attention towards higher-order dimensions of composition and fostering more coherent textual production. These functions align with recent findings that LLM-enabled tools promote iterative, reflective learning processes (Neshaei et al., 2025) and resonate with theories of double-loop learning (Chen et al., 2022), whereby learners move beyond error correction towards strategic rethinking of writing practices. From a practical standpoint, the findings underscore several pedagogical suggestions and implications. First, the results highlight the value of multi-round, personalised evaluation, whereby AI can provide iterative prompts that guide students' revision, while teachers ensure reliability through validation and adjustment of automated scores. This combined mechanism ensures both efficiency and pedagogical trustworthiness. Second, the study points to the usefulness of portfolios as resources and scaffolds. By archiving students' drafts, revisions, and feedback histories, the system not only supported formative assessment but also served as a reflective space where pupils could monitor progress and transfer strategies across tasks. This echoes evidence from recent classroom-based AI writing interventions showing that portfolio-based AI support enhances self-regulation and strategy transfer (Zhang & Yu, 2023; Bal et al., 2025). Third, the intervention provided scaffolding for writing skills and literacy strategy use. Even though effect sizes for strategy improvement were modest, they are significant in the primary context, where writing skills are often underdeveloped. By visualising progress and offering actionable prompts, the system helped students engage in planning, monitoring, and evaluation—key components of effective writing pedagogy (Etaat, 2025). Finally, by strengthening the interconnections between higher-order dimensions of writing, AI-supported instruction offers a countermeasure to the long-standing challenge of integrating feedback into students' recursive writing cycles.

Several limitations remain. The quasi-experimental design does not support strong causal inference, and the unequal revision opportunities across groups mean the study cannot separate the contribution of AI from the contribution of iterative practice. The findings should therefore be read as evidence for an intervention package that integrated AI into a reflective classroom workflow. Future studies should equalize the number of revision rounds across conditions, compare AI with non-AI iterative feedback loops, and test the model in multiple schools. Even with these limitations, the present results suggest that AI can be educationally useful when it is designed to augment teacher judgment and increase students' opportunities for structured reflection.

References

Bal, M., Uyumaz, G., & Maden, B. (2025). Examining the mediating role of metacognitive writing strategies in the relationship between self-regulated writing skills and writing anxiety among middle school students. *British Educational Research Journal*. Advance online publication. https://doi.org/10.1002/berj.4212

Chen, X., Ou, Q., An, C., & Zhang, D. (2022). Teacher mindset change in boundary-crossing lesson study: A case from China. *International Journal for Lesson & Learning Studies*, *11*(2), 91-105. https://doi.org/10.1108/ijlls-03-2021-0019

Etaat, F. (2025). AI-based assessment of students' writing skill progress: Implementing "Artificial neural network modeling" vs. "Time series prediction." *International Journal of Educational Research*, *131*, 102575. https://doi.org/10.1016/j.ijer.2025.102575

Li, A. W., & Hebert, M. (2024). Unpacking an online peer-mediated and self-reflective revision process in second-language persuasive writing. *Reading and Writing*, *37*(6), 1545-1573. https://doi.org/10.1007/s11145-023-10466-8

Liu, Q., Zhong, Z., & Nesbit, J. C. (2024). Argument mapping as a pre-writing activity: Does it promote writing skills of EFL learners? *Education and Information Technologies*, *29*, 7895–7925. https://doi.org/10.1007/s10639-023-12098-5

Neshaei, S. P., Mejia-Domenzain, P., Davis, R. L., & Kaser, T. (2025). Metacognition meets AI: Empowering reflective writing with large language models. *British Journal of Educational Technology*, *56*(5), 1864-1896. https://doi.org/10.1111/bjet.13601

Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability. *Computers and Education: Artificial Intelligence*, *6*, 100234. https://doi.org/10.1016/j.caeai.2024.100234

Schillings, M., Roebertsen, H., Savelberg, H., & Dolmans, D. (2023). A review of educational dialogue strategies to improve academic writing skills. *Active Learning in Higher Education*, *24*(2), 95-108. https://doi.org/10.1177/1469787418810663

Shen, X. L., & Teng, M. F. (2024). Three-wave cross-lagged model on the correlations between critical thinking skills, self-directed learning competency and AI-supported writing. *Thinking Skills and Creativity*, *52*, 101524. https://doi.org/10.1016/j.tsc.2024.101524

Zhang, E. D., & Yu, S. (2023). The development and validation of an L2 student digital multimodal composing competence scale. *Computer Assisted Language Learning*, *38*(3), 486–511. https://doi.org/10.1080/09588221.2023.2201342

Zhang, E. D., Zhou, N., & Yu, S. (2023). Assessing L2 secondary student writing feedback literacy and its predictive effect on their L2 writing performance. *Language Teaching Research*. Advance online publication. https://doi.org/10.1177/13621688231217665