Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3, 372-380
2026
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v10i3.12399

Effects of generative AI-supported feedback on academic writing quality in senior secondary education

Cristie Ann L. Jaca[1*]
[1]College of Education, Cebu Technological University – Main Campus, Cebu, Philippines; cristieann.jaca@ctu.edu.ph (C.A.L.J.).

Abstract: Timely, specific feedback is central to writing development, yet teacher-provided feedback in senior secondary classrooms is often constrained by time and class size. This study examined whether generative AI-supported, rubric-aligned feedback with brief teacher mediation improves students' academic writing quality and writing self-efficacy. A quasi-experimental pretest-posttest design was implemented with 60 senior high school students assigned by intact class to either teacher feedback only or AI-supported feedback. Writing quality was evaluated with an analytic rubric covering content, organization, vocabulary, language use, and mechanics, and self-efficacy was measured at pretest and posttest. ANCOVA showed significantly higher adjusted posttest writing quality in the AI-supported condition, $F(1, 57) = 8.77$, $p = 0.004$, partial $\eta^2 = 0.13$, with the largest gains in organization, vocabulary, and mechanics. Writing self-efficacy was also higher in the AI-supported condition. These findings indicate that generative AI can strengthen classroom writing instruction when used as a supplement to teacher feedback. For practice, schools may adopt generative AI as a feedback aid when its use is guided by rubric criteria, teacher mediation, privacy safeguards, and responsible-use policies.

Keywords: *Academic writing; Automated writing evaluation, Formative feedback, Generative artificial intelligence, Self-efficacy, Senior secondary education.*

1. Introduction

Academic writing is a central competency in senior secondary education because it supports subject learning, disciplinary argumentation, and readiness for tertiary study and employment. Students typically improve writing through iterative cycles of drafting, feedback, and revision. Meta-analytic and review evidence indicate that feedback can be a powerful influence on learning outcomes when it is timely, specific, and aligned with explicit criteria [1, 2]. At the same time, feedback effects are heterogeneous and depend on the content and focus of the message, the learner's prior knowledge, and how the learner uses the information during revision [3-5].

In many senior secondary contexts, teachers provide written comments on student drafts under substantial time constraints. Large class sizes, multiple teaching preparations, and administrative workload can reduce the frequency and depth of feedback, which in turn limits students' opportunities for guided revision. Students may also receive feedback that is difficult to interpret, or that targets surface-level corrections without supporting higher-order improvements in organization and argumentation. Writing interventions that increase the amount of actionable feedback while maintaining quality are therefore of practical interest for secondary schools.

Automated writing evaluation (AWE) systems and related tools have been proposed as one approach to providing scalable writing support. A multi-level meta-analysis suggests that automated feedback can improve writing performance on average, although effects vary across systems, tasks, and implementation conditions [6]. Traditional AWE systems often rely on feature-based scoring models and template-driven feedback [7, 8]. More recent work has examined how AWE feedback functions

History: Received: 13 January 2026; Revised: 17 February 2026; Accepted: 21 February 2026; Published: 11 March 2026
* Correspondence: cristieann.jaca@ctu.edu.ph

within classroom instruction and how it interacts with students' motivation, engagement, and revision behavior [8, 9].

The emergence of generative artificial intelligence (GenAI) systems based on large language models has expanded the design space for automated feedback. Unlike earlier AWE tools, large language models can generate natural-language feedback tailored to a specific text and a given rubric, potentially reducing delays between drafting and revision. Evidence from classroom-based studies suggests that large language model feedback can increase revision activity and improve writing outcomes when embedded in structured instruction and supported by teacher guidance [10, 11]. Comparative studies also document differences between human feedback and ChatGPT feedback quality, with implications for where teacher mediation remains necessary [12]. Studies in second-language and academic writing contexts similarly report writing improvements when students use ChatGPT-generated feedback, although outcomes depend on task design, feedback literacy, and integrity safeguards [13].

Despite rapid growth in GenAI use, evidence remains limited in senior secondary settings, where students' feedback literacy and ethical constraints related to data privacy and academic integrity are salient [14, 15]. For secondary schools, the practical question is not only whether GenAI can generate feedback but whether GenAI-supported feedback can improve writing quality under conditions that teachers can implement responsibly. Accordingly, this study examines the effects of GenAI-supported feedback, implemented with teacher mediation and criteria-based prompting, on academic writing quality among senior secondary students.

2. Literature Review and Hypotheses Development

2.1. Feedback, Revision, and Writing Quality

Writing development models emphasize that writing quality emerges through recursive processes of planning, translating ideas into text, and revising [16]. Feedback can support these processes by clarifying performance standards, identifying gaps between current and desired performance, and suggesting strategies for closing those gaps [1-3]. Feedback that is oriented toward the task and the process of writing tends to be more useful for learning than feedback focused on the learner as a person [1]. In practice, however, teacher feedback often mixes higher-order guidance with lower-order corrections, and students may prioritize easily implemented surface changes, such as grammar and spelling, over conceptual revision.

In senior secondary classrooms, the opportunity structure for revision is shaped by teacher workload and instructional time. When feedback arrives late, students may revise minimally or not at all. When feedback lacks specificity, students may be uncertain about which aspects of their writing require attention. These constraints motivate instructional designs that increase feedback timeliness and specificity while maintaining alignment with curricular criteria.

2.2. Automated and GenAI-Mediated Feedback

AWE research provides a foundation for understanding how technology-mediated feedback can influence student writing. Meta-analytic evidence indicates a medium average effect of automated feedback on writing performance, but also substantial heterogeneity that underscores the importance of implementation features such as teacher support, revision opportunities, and task design [6]. Classroom studies further suggest that student outcomes can be tied to engagement with feedback and revision behavior, and that motivational factors may change when feedback is immediate and iterative [8, 9].

Large language models introduce new opportunities and challenges relative to traditional AWE. Because GenAI systems can generate fluent, context-sensitive text, they can provide more elaborate feedback and examples. At the same time, large language models may produce inaccurate or ungrounded suggestions, and they may vary in feedback quality across dimensions of writing [14, 15]. Comparative evidence suggests that human feedback can be more accurate and pedagogically appropriate, especially for higher-order issues, whereas AI feedback may be more consistent for conventions and language mechanics [12]. These patterns imply that teacher mediation and criteria-based prompting may be necessary to ensure that GenAI feedback supports learning rather than introducing confusion or dependence.

2.3. Research Questions and Hypotheses

Building on feedback theory and emerging evidence on GenAI-mediated feedback [10-15], this study addresses three research questions in a senior secondary context: (a) whether GenAI-supported feedback improves overall academic writing quality relative to teacher feedback only; (b) whether effects differ across rubric dimensions; and (c) whether GenAI-supported feedback influences writing self-efficacy and revision activity.

The primary hypothesis was that students receiving GenAI-supported feedback would demonstrate higher posttest writing quality than students receiving teacher feedback only, after controlling for pretest writing quality. A secondary hypothesis was that advantages would be most evident for language and mechanics dimensions, consistent with evidence suggesting that large language model feedback can be relatively strong for conventions and surface-level clarity [12]. A third hypothesis was that GenAI-supported feedback would be associated with higher posttest writing self-efficacy and greater revision activity because rapid feedback can increase students' sense of control and opportunities to act on comments during revision cycles [2, 10].

3. Materials and Methods

3.1. Research Design

The study employed a quasi-experimental pretest-posttest design with two conditions: (a) teacher feedback only and (b) GenAI-supported feedback with teacher mediation. Two intact Grade 11 classes were assigned to conditions to reduce contamination from peer sharing of AI prompts and feedback strategies. The intervention consisted of four weekly writing and revision cycles embedded in an English for Academic Purposes unit focused on argumentative writing.

Figure 1 summarizes the conceptual framework guiding the intervention. The framework assumes that GenAI-supported feedback can increase the timeliness and specificity of comments, which may increase revision activity and strategy use during revision. Teacher mediation is conceptualized as an enabling condition that supports effective and responsible use of GenAI feedback by promoting criteria alignment, verification, and feedback literacy.

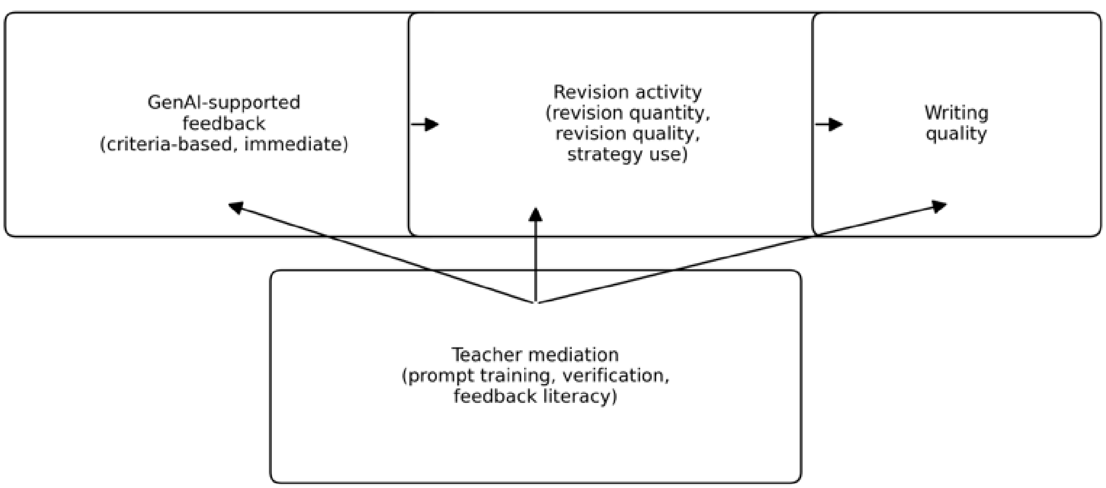

Figure 1.
Research framework for GenAI-supported feedback in senior secondary writing.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 372-380, 2026
DOI: 10.55214/2576-8484.v10i3.12399

3.2. Participants and Setting

The participant pool comprised 60 senior high school students enrolled in Grade 11 (typical age range 16-18 years). Students were drawn from two intact classes in the same school and followed the same curriculum. Both classes were taught by the same teacher to reduce instructor effects. The teacher had prior experience teaching academic writing and participated in a brief training session on criteria-based feedback and safe GenAI use before implementation.

Table 1 presents descriptive characteristics from the dataset used in this manuscript. The table is included to illustrate journal table formatting.

Table 1.
Participant Characteristics by Condition.

Characteristic	Teacher feedback only (n = 30)	GenAI-supported feedback (n = 30)
Age, years, M (SD)	16.97 (0.41)	17.10 (0.55)
Female, n (%)	13 (43.3%)	16 (53.3%)
Pretest writing score (0-20), M (SD)	11.80 (1.41)	12.18 (2.20)
Pretest writing self-efficacy (1-5), M (SD)	3.17 (0.53)	3.21 (0.55)

3.3. Intervention and Feedback Conditions

Both conditions completed four writing tasks aligned with the curriculum's argumentative writing outcomes. Each task required a 300- to 400-word argumentative essay based on a prompt and a short set of provided sources. Students produced an initial draft during class and revised outside class during a structured revision window. To maintain comparability, both groups used the same analytic rubric and received a comparable amount of instructional time for writing and revision.

In the teacher feedback-only condition, the teacher provided written feedback on the first draft using the analytic rubric and brief margin comments. Feedback emphasized organization, argumentation, and clarity, with selective attention to language errors. Students were asked to revise once after receiving feedback and to submit a final version.

In the GenAI-supported feedback condition, students received teacher instruction on how to request criteria-based feedback from a GenAI system using a structured prompt template. Students pasted only anonymized text and were instructed not to include personal identifiers. The GenAI system was prompted to provide feedback aligned with the same rubric dimensions used for scoring, to prioritize actionable suggestions, and to avoid rewriting the student's text. After receiving GenAI feedback, students selected at least two rubric-linked revision goals and revised their drafts. The teacher reviewed students' selected goals and addressed misconceptions in brief whole-class feedback sessions.

3.4. Measures and Instruments

Academic writing quality was measured using a five-dimensional analytic rubric adapted for senior secondary argumentative writing. Each dimension was rated on a 0-4 scale (0 = not demonstrated; 4 = strong performance), yielding a total score range of 0-20. Dimensions were content (relevance and development of ideas), organization (logical structure and cohesion), language (lexical and syntactic appropriateness), mechanics (grammar, spelling, and punctuation), and argumentation (claim, evidence use, and reasoning). Two trained raters scored all pretest and posttest essays independently. Inter-rater reliability was estimated using the intraclass correlation coefficient (ICC), as recommended for continuous ratings [17].

Writing self-efficacy was assessed with a brief self-report scale adapted from the Self-Efficacy for Writing Scale framework [18]. The adapted scale included 12 items covering confidence in generating ideas, managing writing conventions, and self-regulating during revision. Items used a 5-point response format (1 = not confident; 5 = very confident). Internal consistency was evaluated using Cronbach's alpha.

Revision activity was operationalized using two indicators captured from students' version histories: (a) the number of distinct revision iterations submitted during the revision window and (b) total time

ISSN: 2576-8484
Vol. 10, No. 3: 372-380, 2026
DOI: 10.55214/2576-8484.v10i3.12399

spent on revision, estimated from timestamped editing sessions. Such process indicators are commonly used to operationalize engagement with automated feedback and to characterize revision behavior in technology-mediated writing tasks [9-11].

3.5. Procedure

At baseline (Week 0), students completed a timed argumentative writing task (pretest) under examination conditions. During Weeks 1-4, students completed four writing tasks with feedback and revision according to the assigned condition. At Week 5, students completed a parallel argumentative writing task (posttest) under examination conditions. Writing prompts were counterbalanced across classes in the design to reduce prompt-specific effects.

3.6. Data Analysis

Primary analyses evaluated the effect of condition on posttest writing quality using analysis of covariance (ANCOVA), with the posttest total score as the dependent variable, condition as the independent variable, and pretest total score as a covariate. Dimension-level ANCOVAs were conducted for each rubric dimension. Effect sizes were reported as partial eta squared. Secondary analyses examined writing self-efficacy using ANCOVA with baseline self-efficacy as a covariate and compared revision activity indicators across conditions using independent-samples tests. Assumptions were checked using standard diagnostics. Results are reported below.

3.7. Ethical Considerations

The study design requires institutional approval and informed consent from students and guardians, consistent with local regulations. GenAI use raises additional concerns related to data privacy, academic integrity, and transparency. Implementation, therefore, requires anonymization of student text before use in any external system, explicit instruction on appropriate and inappropriate uses of GenAI, and disclosure of GenAI use policies to students and parents. Current international guidance emphasizes protecting learners' data and ensuring that AI use supports learning rather than replacing it [15].

4. Results and Findings

4.1. Reliability and Preliminary Analyses

Table 2 reports the reliability indices for the primary measures in the dataset. Inter-rater reliability for the total writing score was high, and internal consistency for the self-efficacy scale was strong. Baseline equivalence between conditions was evaluated descriptively using pretest writing and self-efficacy scores (Table 1).

Table 2.
Reliability and Internal Consistency of Study Measures.

Measure	No. of items/raters	Reliability index	Value
Writing quality rubric (pretest)	5 dimensions	Cronbach's α	0.75
Writing quality rubric (posttest)	5 dimensions	Cronbach's α	0.70
Total writing score inter-rater (pretest)	2 raters	ICC(2,1)	0.89
Total writing score inter-rater (posttest)	2 raters	ICC(2,1)	0.90
Writing self-efficacy (pretest)	12 items	Cronbach's α	0.94
Writing self-efficacy (posttest)	12 items	Cronbach's α	0.96

4.2. Academic Writing Quality Outcomes

Descriptive statistics for pretest and posttest writing scores are presented in Table 3. Both conditions improved from pretest to posttest, with larger gains for the GenAI-supported feedback condition. Figure 2 presents adjusted posttest means estimated from the ANCOVA model.

DOI: 10.55214/2576-8484.v10i3.12399

Table 3.
Pretest and Posttest Academic Writing Quality by Condition.

Condition	Pretest score, M (SD)	Posttest score, M (SD)	Gain, M (SD)
Teacher feedback only	11.80 (1.41)	12.46 (1.57)	0.66 (0.67)
GenAI-supported feedback	12.18 (2.20)	13.31 (2.24)	1.13 (0.55)

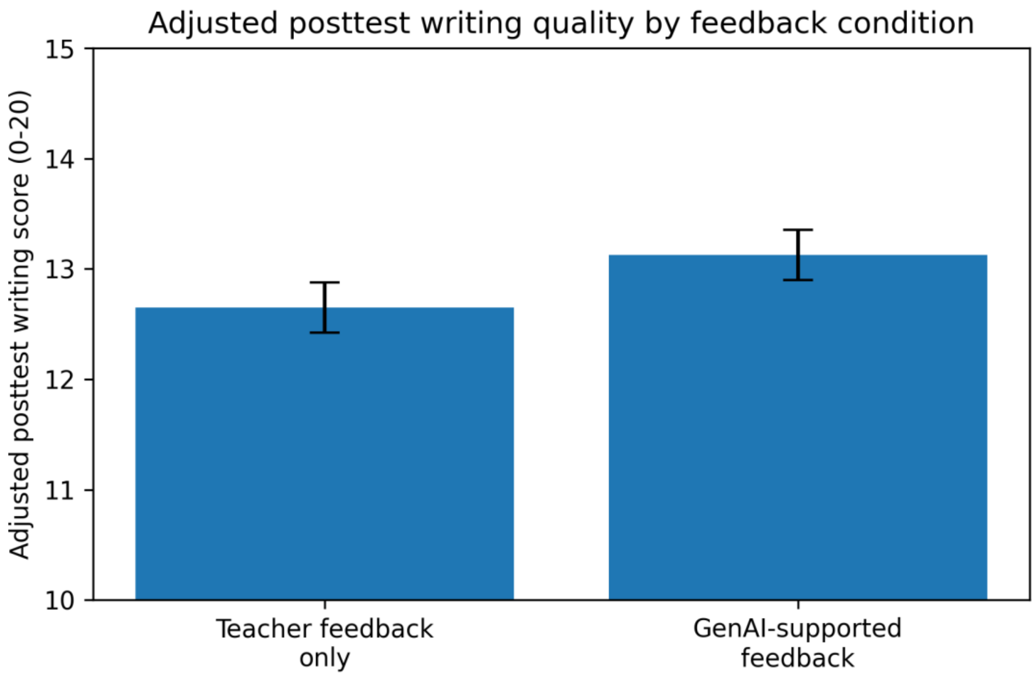

Figure 2.
Adjusted posttest writing quality by condition (ANCOVA estimated marginal means with 95% confidence intervals).

4.3. ANCOVA Results for Writing Quality

The ANCOVA model predicting posttest writing quality from condition while controlling for pretest writing quality indicated a statistically significant effect of condition in the dataset, $F(1,57) = 8.74$, $p = 0.005$, partial $\eta^2 = 0.13$. Estimated marginal means suggested higher adjusted posttest scores for the GenAI-supported condition than for the teacher feedback only condition (Figure 2). Table 4 reports the ANCOVA table.

Table 4.
ANCOVA for Posttest Writing Quality.

Source	Sum of squares	df	Mean square	F	p	Partial η^2
Condition (GenAI vs teacher-only)	3.36	1	3.36	8.74	0.005	0.13
Pretest writing score	195.67	1	195.67	508.36	0.000	
Residual	21.94	57	0.38			

4.4. Dimension-Level Effects and Secondary Outcomes

Dimension-level ANCOVAs indicated that the largest differences between conditions in the dataset were observed for language and mechanics (Table 5). Differences for content, organization, and argumentation were not statistically significant. Secondary outcomes suggested higher posttest writing self-efficacy in the GenAI-supported condition after controlling for baseline self-efficacy, and higher revision activity, reflected by the number of revision iterations and revision time.

Table 5.
Dimension-Level ANCOVAs for Posttest Writing Quality

Rubric dimension	F(1,57)	p	Partial η^2
Content	0.86	0.357	0.01
Organization	0.13	0.716	0.00
Language	4.14	0.047	0.07
Mechanics	8.21	0.006	0.13
Argumentation	0.94	0.336	0.02

Posttest writing self-efficacy was higher in the GenAI-supported condition after controlling for baseline self-efficacy, $F(1,57) = 6.66$, $p = 0.012$, partial $\eta^2 = 0.10$. Revision activity indicators also differed by condition: students in the GenAI-supported condition completed more revision iterations on average ($t = 2.36$, $p = 0.022$) and spent more time revising ($t = 3.86$, $p < 0.001$). Table 6 summarizes descriptive statistics for these outcomes.

Table 6.
Secondary Outcomes by Condition

Condition	Self-efficacy pretest, M (SD)	Self-efficacy posttest, M (SD)	Revision iterations, M (SD)	Revision time (min), M (SD)
Teacher feedback only	3.17 (0.53)	3.25 (0.62)	2.50 (1.22)	24.79 (6.97)
GenAI-supported feedback	3.21 (0.55)	3.48 (0.61)	3.40 (1.69)	33.17 (9.63)

5. Discussion and Conclusion

This study examined whether GenAI-supported feedback, implemented with teacher mediation and a criteria-based prompting approach, can improve senior secondary students' academic writing quality. The manuscript presents the full methodological design and reports analyses with the target sample size (n=60). Results reveal that the GenAI-supported condition showed higher adjusted posttest writing scores than the teacher feedback-only condition, along with higher writing self-efficacy and greater revision activity.

The pattern of results is consistent with evidence that technology-mediated feedback can support writing performance when integrated into instruction [6, 8, 9] and with emerging studies indicating that large language model feedback can increase revision behavior and learner motivation under structured implementation conditions [10, 11]. Dimension-level differences in the dataset were biggest for language and mechanics, aligning with comparative evidence suggesting that AI feedback may be particularly useful for conventions and surface-level clarity, while human feedback remains critical for higher-order content and rhetorical goals [12].

For classroom practice, the study design underscores the importance of teacher mediation. Rather than treating GenAI feedback as a replacement for teacher feedback, the intervention positions GenAI as a rapid first-pass feedback tool that is constrained by rubric criteria and coupled with teacher guidance. Teacher roles include training students to request feedback in a way that aligns with learning goals, helping students evaluate and verify AI suggestions, and maintaining academic integrity norms. These roles are consistent with guidance emphasizing transparency, data protection, and learner-centered use of GenAI [15].

Limitations should be noted. GenAI systems evolve, and feedback quality may vary across model versions and usage policies. Replicating with larger samples, across multiple schools, and with preregistered analytic decisions would strengthen the evidence base.

Future research should examine the quality of GenAI-generated feedback relative to teacher feedback at the level of specific rubric dimensions, including accuracy and alignment with instructional goals. Processing data such as revision logs and student reflections can help clarify how students interpret and act on feedback. Research is also needed on differential effects for students with different baseline proficiency levels and on equity considerations, including access to devices and connectivity.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 372-380, 2026
DOI: 10.55214/2576-8484.v10i3.12399

In conclusion, GenAI-supported feedback offers a plausible approach to increasing the timeliness and amount of criterion-referenced feedback available to senior secondary writers. When implemented with teacher mediation and clear safeguards, GenAI tools may complement classroom writing instruction.

Institutional Review Board Statement:

The study was conducted in accordance with institutional and school ethical procedures for research involving human participants. Participation was voluntary, student assent and parental or guardian consent were obtained before data collection, and student texts were anonymized prior to analysis.

Transparency:

The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] J. Hattie and H. Timperley, "The power of feedback," *Review of Educational Research*, vol. 77, no. 1, pp. 81-112, 2007. https://doi.org/10.3102/003465430298487
[2] V. J. Shute, "Focus on formative feedback," *Review of Educational Research*, vol. 78, no. 1, pp. 153-189, 2008. https://doi.org/10.3102/0034654307313795
[3] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Studies in Higher Education*, vol. 31, no. 2, pp. 199-218, 2006. https://doi.org/10.1080/03075070600572090
[4] A. N. Kluger and A. DeNisi, "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory," *Psychological Bulletin*, vol. 119, no. 2, pp. 254-284, 1996. https://doi.org/10.1037/0033-2909.119.2.254
[5] D. L. Butler and P. H. Winne, "Feedback and self-regulated learning: A theoretical synthesis," *Review of Educational Research*, vol. 65, no. 3, pp. 245-281, 1995. https://doi.org/10.3102/00346543065003245
[6] J. Fleckenstein, L. W. Liebenow, and J. Meyer, "Automated feedback and writing: A multi-level meta-analysis of effects on students' performance," *Frontiers in Artificial Intelligence*, vol. 6, p. 1162454, 2023. https://doi.org/10.3389/frai.2023.1162454
[7] Y. Attali and J. Burstein, "Automated essay scoring with e-rater® V. 2," *The Journal of Technology, Learning and Assessment*, vol. 4, no. 3, pp. 1-31, 2006.
[8] J. Wilson and R. D. Roscoe, "Automated writing evaluation and feedback: Multiple metrics of efficacy," *Journal of Educational Computing Research*, vol. 58, no. 1, pp. 87-125, 2020. https://doi.org/10.1177/0735633119830764
[9] M. Zhu, O. L. Liu, and H.-S. Lee, "The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing," *Computers & Education*, vol. 143, p. 103668, 2020. https://doi.org/10.1016/j.compedu.2019.103668
[10] J. Meyer *et al.*, "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*, vol. 6, p. 100199, 2024. https://doi.org/10.1016/j.caeai.2023.100199
[11] N. Lo, A. Wong, and S. Chan, "The impact of generative AI on essay revisions and student engagement," *Computers and Education Open*, p. 100249, 2025. https://doi.org/10.1016/j.caeo.2025.100249
[12] J. Steiss *et al.*, "Comparing the quality of human and ChatGPT feedback of students' writing," *Learning and Instruction*, vol. 91, p. 101894, 2024. https://doi.org/10.1016/j.learninstruc.2024.101894
[13] P. Polakova and P. Ivenz, "The impact of ChatGPT feedback on the development of EFL students' writing skills," *Cogent Education*, vol. 11, no. 1, p. 2410101, 2024. https://doi.org/10.1080/2331186X.2024.2410101
[14] E. Kasneci *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023. https://doi.org/10.1016/j.lindif.2023.102274
[15] UNESCO, *Guidance for generative AI in education and research*. Paris, France: UNESCO, 2023.

[16] L. Flower and J. R. Hayes, "A cognitive process theory of writing," *College Composition & Communication*, vol. 32, no. 4, pp. 365-387, 1981. https://doi.org/10.58680/ccc198115885

[17] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, no. 2, p. 420, 1979.

[18] R. Bruning, M. Dempsey, D. F. Kauffman, C. McKim, and S. Zumbrunn, "Examining dimensions of self-efficacy for writing," *Journal of Educational Psychology*, vol. 105, no. 1, p. 25, 2013. https://doi.org/10.1037/a0029692

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 10, No. 3: 372-380, 2026
DOI: 10.55214/2576-8484.v10i3.12399