

ULUSLARARASI EĞİTİM BİLİMLERİ DERGİSİ

ISSN: 2149-0848

THE JOURNAL OF INTERNATIONAL EDUCATION SCIENCE

Ortaöğretimde İngilizce Yazma Geri
Bildirimini Yeniden Değerlendirmek:
Yapay Zekâ ile Öğretmen Geri Bildiriminin
Karşılaştırılması

Reconsidering English Writing Feedback in
Secondary Education: A Comparison of
Artificial Intelligence with Teacher
Feedbacks

Melissa GRAB*
Zekiye SEİS**



Geliş/Submitted: 26.07.2026
Kabul/Accepted: 28.09.2026

 10.29228/INESJOURNAL.95568

Makale Türü:

Araştırma Makalesi

Article Information:

Research Article

Citation / Atıf

Grab, M. ve Seis, Z. (2026). Ortaöğretimde İngilizce yazma geri bildirimini yeniden değerlendirmek: Yapay zekâ ile öğretmen geri bildiriminin karşılaştırılması. *Uluslararası Eğitim Bilimleri Dergisi*, 13 (48), 20-40.

This article was checked by Intihal.net. Bu makale İntihal.net tarafından taranmıştır.

This article is under the Creative Commons license. Bu makale Creative Commons lisansı altındadır.

*Dr. Öğr. Üyesi, Giresun Üniversitesi, melissagrab@yahoo.com 

** Öğr. Gör. Dr., Giresun Üniversitesi, zekiye.seis@giresun.edu.tr 

ULUSLARARASI EĞİTİM BİLİMLERİ DERGİSİ

ISSN: 2149-0848

THE JOURNAL OF INTERNATIONAL EDUCATION SCIENCE

Öz: Yapay zekâ destekli geri bildirim araçları, ikinci dil yazma öğretiminde sıklıkla kullanılmaktadır. Bununla birlikte, bu araçların, öğretmen geri bildirimleriyle karşılaştırıldığında, ortaöğretim düzeyinde öğrencilerin yazma gelişimini nasıl etkilediğine dair bulgular yetersizdir. Bu çalışma, ortaöğretim düzeyinde ikinci dil (L2) yazma öğretiminde yapay zekâ tarafından üretilen geri bildirimin ve öğretmen geri bildiriminin öğrencilerin yazma performansı üzerindeki etkilerini karşılaştırmayı amaçlamaktadır. Çalışmada karma yöntemli bir tasarım kullanılmış ve müdahale 12 hafta sürmüştür. Katılımcılar, 10. ve 11. sınıflarda kayıtlı 44 öğrenciden oluşmuş ve her biri 22 öğrenciden oluşan iki gruba ayrılmıştır: yapay zekâ geri bildirim grubu ve öğretmen geri bildirim grubu. Çalışmanın teorik çerçeveleri sosyokültürel teori, geri bildirim okuryazarlığı teorisi ve öz düzenlemeli öğrenme teorisine dayanmaktadır. Nicel kısım, dilbilgisel doğruluğun geliştirilmesinde yapay zekâ geri bildirim teknolojisi müdahalesinde yüksek bir etki büyüklüğü ve yüksek bir sıklık oranı göstermiştir. Argümantasyon ve özgünlüğün geliştirilmesinde anlamlı bir fark görülmemiştir. Niteliksel bölümde bazı olumsuz yönler ortaya çıktı; bunlar arasında kalıplaşmış yazım biçiminin gelişme potansiyeli ve yapay zekâ geri bildiriminin eleştirel olmayan bir şekilde kabul edilmesi yer almaktadır.

Anahtar Kelimeler: İngilizce, İkinci dil yazımı, yapay zekâ, öğretmen geri bildirimi, hibrit yaklaşım.

Abstract: Artificial intelligence-powered feedback tools are frequently being used in second-language writing instruction. However, findings on how these tools affect students' writing development, compared with teacher feedback, at the secondary education level remain scarce. This study aims to compare the effects of AI-generated feedback and teacher feedback on students' writing performance in second language (L2) writing instruction at the secondary school level. A mixed-methods design was used in the study, and the intervention lasted for 12 weeks. The participants consisted of 44 students enrolled in 10th and 11th grades, divided into two groups: an AI feedback group and a teacher feedback group, each comprising 22 students. The theoretical frameworks of the study were based on the sociocultural theory, the theory of feedback literacy, and the theory of self-regulated learning. The quantitative part showed a high effect size in the AI feedback technology intervention for the development of grammatical accuracy, as well as a high frequency rate. The development of argumentation and originality did not show any significant differences.

Some negative aspects were revealed via the qualitative part, including the potential for the development of formulaic writing and the uncritical acceptance of AI feedback.

Keywords: English, artificial intelligence, teacher feedback, hybrid approach.

Introduction

Artificial intelligence (AI) is transforming the education field. Among the most dynamic and challenging subjects in the field is writing. AI is seen to have the potential to assist in the resolution of the long-lasting challenge in writing instruction. AI provides immediate individualized feedback. This is seen to have the potential to boost the motivation, self-regulation, and fluency of the learner. AI tools, such as ChatGPT, Brisk Teaching, MagicSchool AI, and the writing tool developed by the Khan Academy, have been shown to have the potential to boost the motivation, self-regulation, and fluency of the learner.

In addition, writing is a critical skill in any field or discipline, but secondary-level students are commonly found to lack coherence, argumentation, and grammatical accuracy (National Center for Education Statistics, 2024). Although feedback from teachers is important, it is time-consuming, making it difficult to respond to all students (Shute, 2008). AI tools are seen as a solution to the problem of delayed feedback, thereby allowing continuous writing and revision (Bai & Stede, 2023). Previous research has proven the effectiveness of AI tools in enhancing grammatical accuracy as well as student engagement during the revision phase (Escalante, Pack, & Barrett, 2023; Ranalli & Hegelheimer, 2022; Zhai & Ma, 2023).

However, it has also been established that over-reliance on AI tools by teachers could result in various pedagogical drawbacks, including formulaic writing, lack of creativity, bias, as well as shallow thinking (Blodgett et al., 2020; Lee, 2018; Ranalli, Link, & Chukharev-Hudilainen, 2020). Teachers have had mixed feelings about AI tools, as they are both beneficial and detrimental (Baidoo-Anu & Ansah, 2023; Holmes & Tuomi, 2022; Kumar, 2023). As a result, a critical research gap has emerged, as there is limited understanding of the interaction between AI tools and pedagogical theories as well as higher-order student writing skills (Escalante et al., 2023). Moreover, there is a lack of empirically supported, theoretically based models of the combined effects of AI tools and feedback from teachers, focusing on the maximization of the benefits while minimizing the drawbacks (Graham, Harris, & Santangelo, 2015; Lo et al., 2026).

In response, the proposed study attempts to fill the research gap by answering the following research questions:

1. How does AI-generated feedback improve secondary students'

grammatical accuracy, coherence, argumentation, and originality compared to teacher-only feedback?

2. How do students engage with and perceive AI-generated feedback during the writing process?

3. What are teachers' experiences and concerns regarding the integration of AI feedback tools in their writing instruction?

4. How may AI and teacher feedback be optimally integrated to support both lower-order and higher-order writing competencies?

This study was conceptualized based on sociocultural theory, feedback literacy, and self-regulated learning to propose an approach of AI-human feedback hybridization to optimize secondary-level L2 writing.

LITERATURE REVIEW

AI-Based Feedback in Writing: Potentials and Limitations

The results of the meta-analytic review of AI feedback tools, including Automated Writing Evaluation (AWE), show a moderate positive effect of AI feedback tools on writing quality, specifically grammatical/mechanical aspects (Lv, Ren, & Xie, 2021; Zhai & Ma, 2023; Zhang, Jantakoon, & Laoha, 2025). AI feedback tools are seen to be effective in enhancing the quality and quantity of writing by providing non-evaluative feedback, which could reduce affective filtering in writing (Fahmy, 2024; Holmes & Tuomi, 2022). Even Large Language Models (LLMs) of AI feedback tools, such as ChatGPT, are seen to be effective in providing context-based feedback on writing organization/argumentation, even going as deep as human feedback (Nazaretsky et al., 2024; Vázquez-Cano et al., 2023).

However, limitations of AI feedback tools are also seen: AI feedback tools are seen as being general or unsuitable for complex/creative writing, promoting standardized writing (Fleckenstein, Liebenow, & Meyer, 2023; Seßler, Stahl, & Breuer, 2025). The use of AI feedback tools is seen as making the rationale of feedback suggestions opaque, making it difficult for learners to comprehend writing principles (Lee, 2017). AI feedback tools are seen as being poor at providing feedback on the deeper structure/creativity of writing (Dai et al., 2025; Wilson & Roscoe, 2020). The quality of AI feedback tools is also seen as being variable, showing wide discrepancies in terms of accuracy/suitability (Escalante, Pack, & Barrett, 2023; Goldman, Smith, & Carreon, 2024).

Perceptions of Students and Teachers towards AI Feedback

The perception of students and teachers towards AI feedback is diverse. Generally, students find the immediacy and clarity of AI feedback useful but limited for

their development as writers. For instance, Barrett & Pack (2023), Charles & Charles (2024), and Henderson et al. (2025) found that students generally find the immediacy and clarity of AI feedback useful but limited for their development as writers. High school students, for instance, find the AI feedback useful for grammatical and mechanical issues, but not useful for argumentation and creative writing, as found by Charles and Charles (2024).

Teachers find the use of AI useful because they can use the saved time for more advanced and personalized work. For instance, Guo et al. (2023) and Yeo (2023) found that teachers find the use of AI useful, as they can use the saved time for more advanced and personalized work. However, there is a growing concern that students might over-rely on the teachers, the development of critical evaluation skills, and issues of equity, as found by Kumar (2023) and Longoni et al. (2019). The use of AI, therefore, could be considered a “double-edged sword” as it may both build students’ confidence and fluency and, at the same time, undermine their voice and critical thinking, as found by Yao et al. (2025). The “double-edged sword” phenomenon, therefore, calls for models that can clearly separate the non-overlapping roles of both, rather than substituting them, as found by Ranalli and Hegelheimer (2022) and Sperber et al. (2025).

Integrative Models of AI and Human Feedback

Currently, there is a rising call for the development of a “hybrid” or “blended” feedback model that leverages the advantages of both AI and the contextual, dialogic nature of the instructor (Graham et al., 2015; Ranalli & Hegelheimer, 2022; Urzúa et al., 2025). Integration of AI with a human feedback system is a form of pedagogical design that focuses on the complementary nature of the work of each actor in the process.

The proposed models include the “Peer and AI Review + Reflection (PAIRR)” model, which combines AI with peer review and reflection (Sperber et al., 2025). In another study, Lo et al. (2026) proposed a hybrid model of AI feedback, where AI is used for mechanical aspects of feedback, and instructors handle aspects of argumentation and rhetorical effectiveness. Gu, Chen, and Yan (2026) also proposed a study on the scaffolding of self-regulated feedback literacy in the context of AI feedback. However, there is a lack of theoretical underpinnings of the proposed models, as well as the development of feedback literacy in the context of AI feedback. Moreover, there is a lack of empirical validation of the proposed models in the context of secondary education (Carless & Boud, 2018; Linh Hoang, 2024; Yin et al., 2025). This study aims to address the gap in the literature through the development of a hybrid model of AI feedback that is empirically grounded.

THEORETICAL FRAMEWORK

The study adopted a combination of three theoretical frameworks.

Sociocultural Theory

Vygotsky's sociocultural theory of learning is a perspective on the nature of learning. He believed that learning occurs through a process of mediational action. In the context of feedback, it occurs within the "Zone of Proximal Development" (ZPD), which is the distance between what the learner can do independently and what the learner can do with assistance. In the context of feedback, the ZPD refers to the level of cognitive development that the learner is currently at. In the context of feedback, the ZPD refers to the level of cognitive development that the learner is currently at. "Cultural tools, like AI, assist with lower-level cognitive tasks, like grammar and spelling. However, if the goal is for the learner to develop higher-order cognitive skills, then the learner must interact with a more competent other, like a teacher" (Engeness & Gamlem, 2025). Thus, the function of the teacher changes from a feedback provider to a "metacognitive guide."

Feedback Literacy

Feedback literacy, as defined by Carless and Boud (2018), is the ability of learners to understand, judge, and use feedback. Feedback literacy includes four dimensions: appreciating, judging, managing, and acting. However, with the integration of AI, the dimensions increase to include the ability to understand the feedback, judge the source of the feedback, recognize the limitations of the feedback, and decide which suggestions to follow or reject (Gu et al., 2026). Although the use of AI can promote feedback literacy through the use of writing conventions, the authoritative tone can promote the acceptance of the feedback without questioning, which can hinder the development of judgmental skills. This hybrid model requires teaching how to critically evaluate the feedback given by the AI.

Self-Regulated Learning

According to Zimmerman (2002), self-regulated learning is a cyclical process of goal setting, strategy use, self-monitoring, and adaptation to feedback or changes. Writing is the primary self-regulated learning process that includes goal setting, self-monitoring, and adaptation through feedback or self-evaluation (Dai et al., 2025). The use of immediate feedback from the AI can sustain the self-regulated learning process through the provision of continuous self-monitoring information, which can result in an increase in the number of revisions made to the writing as well as the use of forethought and performance strategies (Gu et al., 2026). However, without the teacher's scaffolding, the use of the AI can result in shallow self-regulation as opposed to goal-oriented self-regulated learning. The teacher is required to scaffold the use of the AI to ensure the alignment of the revisions with the higher-order writing objectives as well as the metacognitive development to promote autonomous writing.

Integrative Lens

An integrated view of the use of AI is the concept of the mediated learning system, which includes the teacher providing feedback as well as the curation of the AI to promote the use of the feedback to advance the learners to the next stage of autonomous self-regulated learning through the use of metacognitive dialogue.

METHODOLOGY

Research Design

A quasi-experimental convergent mixed-methods design was used. Quantitative data collection methods examined the change in writing and revision behaviors. Qualitative data collection methods triangulated the perceptions and experiences of the participants. The study was preregistered on the Open Science Framework. The study design was approved by the Institutional Review Board at the University of Washington (IRB #2023-0456) and the Research Office at Seattle Public Schools.

Participants

To determine the number of participants, a power analysis was conducted using G*Power (Faul et al., 2009). The effect size was set to 0.5 for a two-tailed independent t-test, alpha set to .05, and power set to .80. The number of participants for the study was calculated to be 64. A convenience sample of 68 high school students was initially recruited to account for potential participant dropout. The final study participants consisted of 44 high school students in grades 10 and 11 from a metropolitan urban comprehensive high school in the Pacific Northwest. The study participants consisted of 22 AI group participants and 22 teacher group participants. The final number of participants limited the study's power to detect small effect sizes and interactions. The study participants consisted of 45% eligible for free/reduced-price lunch, 38% English language learners, and the racial/ethnic composition was 42% Hispanic/Latino, 28% White, 18% Asian, 8% Black, and 4% multiracial. The participants all took college-prep English and scored proficient on state writing assessments and were assigned to the experimental and control groups by random assignment which occurred at the individual participant level, and was stratified by grade level, producing equal-sized groups, each having 22 participants. There was no participant dropout.

Four English teachers (3 female, 1 male) participated, each teaching one section, to control for teacher variables. Teachers had prior experience teaching technology-based instruction, although limited experience teaching AI writing tools.

Intervention Design

Both groups worked on the same four major writing assignments, based on the curriculum requirements set by the school districts. The four assignments included a

personal narrative essay, an expository essay, an argumentative essay, and an analytical essay, conducted in Weeks 2-3, 4-6, 7-9, and 10-12, respectively. The process-writing approach was used for each assignment.

Experimental Group Using AI Tools

The participants worked on the four assignments and submitted drafts via a custom-designed online platform that included four AI writing tools: ChatGPT-4 for organization and argument, Khan Academy Writing Coach for genre-specific structure, Brisk Teaching for coherence and transitions, and MagicSchool AI for grammar, mechanics, and style. The students could use any of the four tools for the assignments. The analytics provided by the platform showed that all the participants used MagicSchool AI for grammar, and the percentages for the other three tools were 86% for ChatGPT-4, 77% for Brisk Teaching, and 68% for Khan Academy Writing Coach. The immediate automated feedback was provided after the submission of each draft, and at least one revision was made before the final submission.

Control Group (Teacher)

The same four assignments and the same platform were used, and the participants received conventional written feedback. To control the quality and content of the written feedback, the teachers participated in a 2-hour training on formative feedback strategies, specifically on the importance of finding a balance between lower- and higher-order concerns. Table 1 represents the Intervention Timeline and Activities.

Table 1. *Intervention Timeline and Activities*

Week	Activity	AI Group	Control Group
1	Pre-test writing assessment	All students complete a baseline writing sample	All students complete a baseline writing sample
2-3	Narrative writing unit	AI feedback on drafts; teacher instruction	Teacher feedback on drafts; teacher instruction
4-6	Expository writing unit	AI feedback on drafts; teacher instruction	Teacher feedback on drafts; teacher instruction
7-9	Argumentative writing unit	AI feedback on drafts; teacher instruction	Teacher feedback on drafts; teacher instruction

Week	Activity	AI Group	Control Group
10-11	Analytical writing unit + Post-test	AI feedback on drafts; post-test writing	Teacher feedback on drafts; post-test writing
12	Surveys and interviews	Student surveys; teacher interviews	Student surveys; teacher interviews

Data Collection and Outcome Measures

Writing Samples and Rubric Development: A pre-test writing sample was taken during Week 1, while a post-test sample was taken during Week 12 using a similar argumentative prompt to avoid test-retest effects. A post-intervention writing sample was taken at the end of Week 12 using a different analytical prompt. An analytic rubric based on Jacobs et al. (1981), modified for this study, was developed to evaluate the writing samples based on four subscales:

- *Grammar/Mechanics* (0-20): sentence structure, usage, mechanics.
- *Coherence/Organization* (0-20): logical sequencing, transitions, structure.
- *Argumentation/Development* (0-20): clarity of thesis, evidence, reasoning.
- *Originality/Creativity* (0-20): presence of distinctive voice, creativity, non-formulaic thinking.

Two expert blind raters scored all writing samples. Experts were provided a set of 10 benchmark samples not part of the actual study. Two raters scored all samples. As a result, the score of inter-rater reliability (Two-way mixed, Absolute Agreement) was .90 (95% CI [.86, .93]) for total scores. ICC values for individual subscales ranged from .84 to .91. Then, the actions were followed in an order:

1. *Revision Frequency*: The online tool recorded the number of drafts per assignment.
2. *Student Surveys*: A 20-item post-intervention survey was conducted using Likert scales (1-5).
3. *Teacher Interviews*: Semi-structured interviews (45-60 minutes) of all four teachers were conducted near the end of the intervention period, covering instructional experiences, benefits, drawbacks, observations of student

engagement, and ethics of AI use.

Data Analysis

For the quantitative data analysis, normality of data (Shapiro-Wilk) and homogeneity of variance (Levene's test) assumptions were checked for all the study outcomes. Baseline equivalence was checked through the independent t-test on the pre-test total writing scores. ANCOVA was applied to each post-test writing subscale with the pre-test scores as covariates. Effect size was reported as partial eta squared with 95% confidence intervals. Cohen's d was calculated to report the effect size for the revision frequency outcome. To adjust for multiple comparisons across the four subscales, the Bonferroni correction was applied to set the alpha level at .0125.

For the qualitative data analysis, thematic analysis was applied to the interview and open-ended survey data. Braun and Clarke's (2006) thematic analysis procedure was adopted, with the deductive coding stage drawing on the theoretical framework of sociocultural mediation, feedback literacy, and self-regulated learning, followed by the inductive coding stage. NVivo v. 15 was used as the data analysis software to aid the analysis process. The reliability of the coding process was checked through the use of kappa correlation with the analysis conducted independently by two researchers on 30% of the data set, which yielded a kappa correlation coefficient of .89. Discrepancies were resolved through discussion.

FINDINGS

Quantitative Results

The independent t-test was applied to the pre-test total writing score to compare the baseline equivalence of the AI group with the teacher group. The result indicated no significant difference between the two groups, with the AI group having a score of 59.5 (SD = 7.1) compared to the teacher group having a score of 59.9 (SD = 7.4) on the pre-test total writing score, $t(42) = -0.18$, $p = .86$. Table 2 presents the results of the ANCOVA analyses for the post-test writing subscales, controlling for pre-test scores.

Table 2. ANCOVA Results for Post-Intervention Writing Subscales by Group

Subscale	Group (n=22 each)	Post- Test M (SD)	F (df)	p	95% CI for η^2
Grammar Mechanics	/ AI	17.6 (1.)	8.94 (1, 41)	.003*	.18 [.04, .35]

Subscale	Group (n=22 each)	Post- Test M (SD)	F (df)	p	95% CI for η^2
	Teacher	16.3 (2.0)			
Coherence Organization	AI	17.8 (2.1)	5.23 (1, 41)	.027	.11 [0.00, .27]
	Teacher	16.9 (2.3)			
Argumentation Development	AI	16.0 (2.5)	2.01 (1, 41)	.16	.05 [0.00, .19]
	Teacher	15.7 (2.4)			
Originality Creativity	AI	14.2 (2.7)	0.28 (1, 41)	.60	.01 [0.00, .09]
	Teacher	14.5 (2.6)			

Note: $p < .0125$
(Bonferroni
correction for 4
comparisons). Pre-
test scores were used
as a covariate.

The ANCOVA revealed a statistically significant effect of feedback type on the Grammar/Mechanics subscale after controlling for pre-test scores, $F(1, 41) = 8.94$, $p = .003$, $\eta^2 = 0.18$. The effect on Coherence/Organization did not reach the Bonferroni-corrected significance level, $F(1, 41) = 5.23$, $p = .027$, $\eta^2 = 0.11$. No statistically significant effects

were found for Argumentation/Development or Originality/Creativity.

Regarding revision frequency, an independent-samples t-test revealed that the AI group made significantly more drafts per assignment ($M = 2.7$, $SD = 0.8$) than the teacher group ($M = 1.6$, $SD = 0.6$), $t(42) = 5.12$, $p < .001$, Cohen's $d = 1.54$ (95% CI [1.00, 2.08]).

Qualitative Themes

Four themes were identified through thematic analysis.

Theme 1: Efficiency versus Depth

The significance of this study lies in the fact that there was a significant decrease in the time spent on error correction, thus allowing more time for argumentation and originality. At the same time, however, there was a growing concern over the level of students' understanding, despite this efficiency.

Teacher 3: *"I estimate saving at least 5–7 hours per week on basic grammar feedback. However, students increasingly view the AI as the primary and sole authority. They submit drafts, receive suggestions, and implement changes without adequately considering the underlying reasons or whether the modification enhances their voice or argument."*

Teacher 2: *"I actually have more time for individual student work, yet I encounter different problem types. Instead of asking me, 'Is this a comma splice?' they now ask, 'Should I make this change?' effectively outsourcing the thinking process."*

Theme 2: The Formulaic Drift

The development of formulaic writing, which saw writing becoming more standardized, was noticed by both teachers and students.

Teacher 1: *"The suggestions yield 'correct' sentences, but they tend to be formulaic and homogeneous. The essays resemble those written by a single author. Correct and polished, but lacking voice. I can no longer distinguish my students."*

Student 8 (AI Group): *"My essays are better with AI assistance; I make fewer mistakes, and they read more fluently. Yet they do not feel like my own writing. After receiving my last essay, I did not recognize it as my own ideas."*

Theme 3: Critical Engagement Deficit

The survey results revealed that while 86% of the students in the AI group found the tools helpful for grammar, only 35% of them were confident about assessing the content or argument implications of the suggestions.

Student 3 (AI Group): *"I can judge the grammar suggestions, but for other recommendations, such as whether to reorganize a paragraph or change my argument, I end up*

following the AI's suggestions because I assume they know best, not me."

Teacher 4: *"They are not learning to evaluate feedback critically. With teacher feedback, they would sometimes push back, question, or argue. With the AI, they simply accept. It is as if the AI has become the ultimate authority."*

Theme 4: Differential Impact

The teachers reported differential impacts of AI feedback based on the students' pre-existing writing proficiency.

Teacher 2: *"My struggling writers... thrived with AI. It gave them confidence to develop ideas without worrying about the basics. For them, it was transformative."*

Teacher 1: *"But the stronger writers, who possessed authentic voice and style, plateaued. AI suggestions smoothed their syntax and diminished distinctiveness. Their work was satisfactory but lacked pizzazz. I had to encourage them to reject some AI suggestions."*

DISCUSSION

The study provides further evidence for the efficacy of AI-based feedback systems in facilitating the development of certain aspects of L2 writing in the context of secondary education. The study's findings on grammatical accuracy and frequency of revisions support the literature on Automated Writing Evaluation (Lv et al., 2021; Zhai & Ma, 2023). The findings on grammatical accuracy and frequency of revisions provide further evidence for the efficacy of AI-based feedback systems in the development of the Zone of Proximal Development (ZPD) in grammar and coherence (Engeness & Gamlem, 2025) and in monitoring and control aspects of Self-Regulated Learning (SRL) (Dai et al., 2025; Gu et al., 2026).

Nevertheless, the qualitative results highlight the limitations of such scaffolding when the contexts are pedagogically isolated. The small effect on originality, the trend towards formulaic writing, and the absence of feedback literacy indicate that without pedagogical mediation, the use of AI-assisted learning may result in mechanical learning rather than deep learning. The paradox of the effectiveness of AI-assisted learning on mechanical learning, originality, and critical feedback literacy is the core challenge to writing instructors in the age of AI (Kumar, 2023; Longoni et al., 2019).

Research Question 1: Writing Competency Gains

The results showed a significant effect of AI-assisted learning on grammatical accuracy compared to teacher feedback alone. Although coherence was positive, the effect was non-significant after adjusting for multiple comparisons. The non-significant effect on argumentation and originality suggests that without pedagogical mediation, the use of AI-assisted learning is not effective in enhancing argumentation and

originality writing competencies, which is supported by the theoretical view of the scaffolding effect of AI on the mechanical learning of writing (Vygotsky, 1978; Ranalli & Hegelheimer, 2022).

Research Question 2: Student Engagement and Perceptions

The results showed student engagement with the use of AI-assisted learning through the increase in the frequency of revisions. However, the engagement was mechanical learning rather than cognitive learning, as shown by the acceptance of the feedback without critical thinking. This extends the concept of feedback literacy to the effect of the authoritative tone of the AI on the judgment process (Carless & Boud, 2018; Gu et al., 2026). Although the participants appreciated the ease of use of the AI-assisted learning tool, the results showed the absence of critical feedback literacy to evaluate the content-related feedback, as shown by the non-significant effect on argumentation and originality writing competencies and the trend towards formulaic writing (Barrett & Pack, 2023; Henderson et al., 2025).

Research Question 3: Teacher Experiences and Concerns

Teachers appreciated the efficiency gains, allowing them to concentrate more on higher-order teaching, but felt uneasy about the students' uncritical reception of AI suggestions and the loss of individuality in writing products. This mirrors the fundamental contradiction of AI integration: efficiency gains are possible but at the possible cost of developing the desired skills (Holmes & Tuomi, 2022; Yao et al., 2025; Yeo, 2023).

Research Question 4: Pathways to Optimal Integration

The results imply that optimum AI integration is more than merely offering AI tools as part of the traditional teaching package but involves specific pedagogical attention to feedback literacy, critical thinking about AI suggestions, and differentiation between AI and human feedback (Lo et al., 2026; Sperber et al., 2025; Urzúa et al., 2025).

Theoretical Implications

From a sociocultural perspective, AI, as a cultural artifact, is seen as mediating the internalization of lower-order writing skills but requiring teacher mediation in the development of higher-order skills. The ZPD of the classroom with AI integration could be seen as two-layered: the AI-mediated ZPD of procedural writing skills and the teacher-mediated ZPD of critical thinking, creativity, and metacognitive skills (Engeness & Gamlem, 2025).

In relation to feedback literacy theory, it was seen that AI integration in the classroom environment demands the extension of feedback literacy to critical thinking about AI feedback. Without specific attention being given to feedback literacy, AI

integration could actually be seen as undermining rather than supporting the development of feedback literacy (Carless & Boud, 2018; Gu et al., 2026).

With respect to SRL theory, the results suggest that while AI may increase the frequency of self-regulation behaviors (more revisions), it does not necessarily enhance the quality of self-regulation without the teacher's involvement in goal setting, strategy selection, and metacognitive reflection (Zimmerman, 2002; Dai et al., 2025).

The Hybrid AI-Human Feedback Approach

In light of the results and the theoretical framework, we propose the Hybrid AI-Human Feedback Approach to optimize feedback task allocation based on the strengths of each agent. Regarding, *sociocultural role*, AI is conceptualized as a cultural artifact that offers initial scaffolding support for basic concerns, while the teacher is conceptualized as the knowledgeable other who helps the student internalize the use of the cultural artifact to facilitate genuine development.

With the aim of Feedback Literacy, the approach locates the development of student feedback literacy as the teacher's responsibility, thereby repositioning AI as a critical assessment tool rather than one offering response-related feedback, prompting students to consider: "Is this a good suggestion? Does it maintain my voice and intent?"

For SRL Enhancement, AI enhances the SRL cycle by offering rapid feedback, while the teacher enhances the quality of the SRL cycle by assisting students in goal setting beyond "error correction" and assessing the argumentative quality and creativity of student feedback.

CONCLUSION

The study suggests that the use of AI tools for providing feedback is beneficial to the development of some L2 writing skills of secondary school students. However, the study results should be viewed with some degree of caution as the use of AI tools was not adequate to promote the development of higher-order writing skills such as critical thinking and creativity. The study suggests that the teacher is indispensable in the writing classroom.

The contribution of the study is the Hybrid Approach of AI-Human Feedback, which is informed by sociocultural theory, feedback literacy, and self-regulated learning. The approach views the teacher as using the AI tool as a teaching aid to implement a given pedagogical framework, allowing the teacher to focus on the core functions of teaching. The future of writing classrooms is no longer a battle between the teacher and the AI tool but the integration of both could strengthen learning and teaching foreign languages. Some practical and policy implications are suggested:

For Classroom Practitioners, effective AI integration requires a targeted

pedagogical approach:

- *Strategic Role Definition*: Teachers should clearly delineate which aspects of writing AI will address (grammar, mechanics, basic coherence) and which remain the teacher's purview (argumentation, voice, critical thinking). This should be transparent to students.

- *Explicit Feedback Literacy Instruction*: Students need direct instruction in how to evaluate AI feedback critically. This includes modeling the interrogation process, providing frameworks for decision-making, and creating opportunities for students to justify their revision choices. Teaching students how to critically evaluate feedback generated by AI, which includes modeling and decision-making tools (Carless & Boud, 2018; Gu et al., 2026).

- *Structured Reflection*: Require students to document their revision decisions, explaining which AI suggestions they accepted, rejected, or modified and why. This metacognitive practice supports feedback literacy development. Documentation of decisions made by the AI system is necessary.

- *Differentiated Approaches*: Recognize that students at different proficiency levels may need different guidance. Struggling writers may benefit from AI's scaffolding; advanced writers may need encouragement to selectively reject AI suggestions to preserve voice.

New approaches to meet the needs of different proficiency levels (Ranalli & Hegelheimer, 2022) could be developed.

For Teacher Education and Professional Development, opportunities to engage with the tools and critically evaluate the AI writing tools should be made available (Holmes & Tuomi, 2022). Teachers' knowledge of how to teach with/through AI (Yao et al., 2025) could be enhanced by promoting pedagogical content knowledge for AI. The ethical implications of the use of AI should be addressed (Blodgett et al., 2020; UNESCO, 2021).

For Policy and Curriculum, new guidelines to inform the appropriate use of AI tools, ethical considerations, and the feedback literacy could be formed

- *Assessment Reform*: Reviewing writing assessment to ensure the integration of AI tools is feasible, which may involve the use of process writing or writing in the classroom (Sperber et al., 2025).

- *Equity Considerations*: The digital divide should be addressed to ensure equitable access.

Future Research

A longitudinal study could be done for the development of feedback literacy and

writing instruction. Also, a design-based research might be developed as the Hybrid AI-Human Feedback Approach proposed. Subgroup analyses of the effects of AI-based instruction for diverse learners (Goldman et al., 2024) could be carried out. A revision quality study could be done to compare AI-informed and human-informed revisions. Some interventions to teach critical evaluation skills for AI-informed revisions could also be useful (Gu et al., 2026). New studies may compare the effectiveness of different AI tool configurations. Very importantly, ethical studies for the impact of AI bias and equity issues might lead how to benefit from AI in language instruction (Blodgett et al., 2020; Lee, 2018).

REFERENCES

- Bai, X., & Stede, M. A. (2023). Survey of current machine learning approaches to student free-text evaluation for intelligent tutoring. *International Journal of Artificial Intelligence in Education*, 33, 992–1030.
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52-62.
- Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20(1), 1-24.
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5454–5476.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
- Charles, R., & Charles, T. (2024). High school students' perceptions of using AI for learning. *Voice of the Publisher*, 10(3), 284–297.
- Dai, X., Wen, Z., Jiang, J., Liu, H., & Zhang, Y. (2025). How students use AI feedback matters: Experimental evidence on physics achievement and autonomy. *arXiv*. <https://doi.org/10.48550/arXiv.2505.08672>
- Engeness, I., & Gamlem, S. M. (2025). Exploring AI-driven feedback as a cultural tool: A cultural-historical perspective on the design of AI environments to support students' writing process. *Integrative Psychological and Behavioral Science*, 59, Article 23.
- 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.

- Fahmy, Y. (2024, July 5). Student perception of AI-driven assessment: Motivation, engagement and feedback capabilities. In *Proceedings of the 41st Twente Student Conference on IT (TScIT)*. University of Twente.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160.
- Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6, 1162454.
- Goldman, S. R., Smith, S. J., & Carreon, A. (2024). Artificial intelligence to support writing outcomes for students with disabilities. *Journal of Special Education Technology*.
- Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-based writing practices and the common core: Meta-analysis and meta-synthesis. *The Elementary School Journal*, 115(4), 498–522.
- Gu, J., Chen, J., & Yan, Z. (2026). Fostering feedback literacy by scaffolding self-regulated feedback: A comparative study of GenAI and human peers. *Frontiers in Education*, 10, 1736446.
- Guo, K., Zhong, Y., Li, D., Kai, S., & Chu, W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education*, 203, 104862.
- Henderson, M., Bearman, M., Chung, J., Fawns, T., Buckingham Shum, S., Matthews, K. E., & de Mello Heredia, J. (2025). Comparing generative AI and teacher feedback: Student perceptions of usefulness and trustworthiness. *Assessment & Evaluation in Higher Education*, 1–16.
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542-570.
- Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Newbury House.
- Kumar, R. (2023). Faculty members' use of artificial intelligence to grade student papers: a case of implications. *International Journal for Educational Integrity*, 19(1), 9.
- Lee, I. (2017). *Classroom Writing Assessment and Feedback in L2 School Contexts*. Springer.
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 1–16.
- Linh Hoang, G. T. (2024). Effects of automated feedback on English as a foreign language learners' writing performance: Evidence from a quasi-experiment. *RELC Journal*.
- Lo, J., Wong, C., Ng, A., Wong, P., Cheung, D., & Lai, P. (2026). Stretching AI's reach: Assessing

- an AI-driven feedback system for extended academic writing. *Computers and Education: Artificial Intelligence*, 10, 100511.
- Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4), 629–650.
- Lv, X., Ren, W. & Xie, Y. (2021). The effects of online feedback on ESL/EFL writing: A meta-analysis. **Asia-Pacific Edu Res*, 30*, 643–653.
- National Center for Education Statistics. (2024). *Report on the Condition of Education 2024* (NCES 2024-144). U.S. Department of Education.
- Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2024). AI or human? Evaluating student feedback perceptions in higher education. In R. Ferreira Mello (Ed.), *Proceedings of the European Conference on Technology Enhanced Learning (EC-TEL 2024)* (Vol. 15159, pp. 284–298). Springer.
- Ranalli, J., & Hegelheimer, V. (2022). Introduction to the special issue on automated writing evaluation. *Language Learning & Technology*, 26(2), 1–4.
- Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2020). Automated writing evaluation for formative assessment of L2 writing: Investigating the criterion validity of essay scores and error feedback. *Assessing Writing*, 44, 100451.
- Seßler, S., Stahl, B. C., & Breuer, J. (2025). Can large language models provide useful feedback? A comparative evaluation of GPT-generated, expert, and peer feedback in science education. *British Journal of Educational Technology*, 56(2), 342–361.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Sperber, L., MacArthur, M., Minnillo, S., Stillman, N., & Whithaus, C. (2025). Peer and AI Review + Reflection (PAIRR): A human-centered approach to formative assessment. *Computers and Composition*, 76, 102921.
- United Nations Educational, Scientific and Cultural Organization. (2021). *Recommendation on the ethics of artificial intelligence*.
- Urzúa, C. A. C., Ranjan, R., Saavedra, E. E. M., Badilla-Quintana, M. G., Lepe-Martínez, N., & Philominraj, A. (2025). Effects of AI-assisted feedback via generative chat on academic writing in higher education students: A systematic review. *Education Sciences*, 15(10), Article 1396.
- Vázquez-Cano, E., Ramírez-Hurtado, J. M., Sáez-López, J. M., & López-Meneses, E. (2023). ChatGPT: The brightest student in the class. *Teaching and Teacher Education: Leadership and Professional Development*, 102, 101380.
- Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences*, 15(3), 343.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

- Wilson, A., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple feedback types and their impact on revision quality. *Journal of Educational Computing Research*, 58(6), 1123–1149.
- Yao, Y., Zhu, X., Xiao, L., & Lu, Q. (2025). Secondary school English teachers' application of artificial intelligence-guided chatbot in the provision of feedback on student writing: An activity theory perspective. *Journal of Second Language Writing*, 67, 101179.
- Yeo, M. A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal*, 14(3), e716.
- Yin, J., Xu, H., Pan, Y., Wang, H., Liu, T., & Qiu, J. (2025). Effects of different AI-driven chatbot feedback on learning outcomes and brain activity. *Npj Science of Learning*, 10(1), 17.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education -- Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39.
- Zhai, N., & Ma, X. (2023). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research*.
- Zhang, J., Jantakoon, T., & Laoha, R. (2025). Meta-analysis of artificial intelligence in education. *Higher Education Studies*, 15(2), 189–199.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64-70.

“COPE-Dergi Editörleri İçin Davranış Kuralları ve En İyi Uygulama İlkeleri” beyanları:

Etik Kurul Belgesi:	Bu çalışma için etik kurul onayı gerekmemektedir.
Çıkar Çatışması Beyanı:	Bu makalenin araştırması, yazarlığı veya yayınlanmasıyla ilgili olarak yazar/ların potansiyel bir çıkar çatışması yoktur.
Finansal Destek:	Bu çalışmanın araştırma ve yazım aşamasında herhangi kişi/kurum veya kuruluşlar tarafından finansal destek alınmadığı bildirilmiştir.
Katkı Oranı Beyanı:	%50-%50
Destek ve Teşekkür Beyanı:	
Sorumlu Yazar:	Melissa GRAB
Çifte Kör Hakem Değerlendirmesi:	Dış-bağımsız
<i>The following statements are made in the framework of “COPE-Code of Conduct and Best Practices Guidelines for Journal Editors”:</i>	
Ethics Committee Approval:	Ethics committee approval is not required for this article.
Declaration of Conflicting Interests:	No conflicts of interest were reported for this article.
Financial Support:	It has been reported that this study did not receive financial support from any

Author Contributions:	person/institution or organization during the research and writing phase.
Statement of Support and Acknowledgment:	%50-%50
Corresponding Author:	Melissa GRAB
Double-Blind Peer Review:	External-independent