ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

BEYOND GRAMMAR CORRECTION:GENERATIVE AI AND THE CULTIVATION OF HIGHER-ORDER THINKING IN EFL WRITING

Bushra Aslam
MS English Linguistics, Visiting Faculty Member (Air University and Bahira University) Islamabad, Pakistan.
bushraslam02@gmail.com
Atika Rub Nawaz
MS English, Lecturer (University of Wah) Wah Cantt, Pakistan.
ms.aatika@hotmail.com
Ahsan Afzaal Ahmed
MS English Linguistics (National University of Modern Languages) Islamabad, Pakistan.
aaahmed@numl.edu.pk
Gulnaz Begum
M.Phil English Linguistics (National University of Modern Languages) Islamabad, Pakistan.
gbegum@numl.edu.pk

Abstract
The rapid integration of generative artificial intelligence (GenAI) into educational settings has raised new possibilities and challenges for second language acquisition, particularly in the domain of academic writing. While most existing applications emphasize lower-order concerns such as grammar correction, spelling, and vocabulary enhancement, less is known about how GenAI can be harnessed to cultivate higher-order thinking skills in English as a Foreign Language (EFL) writing contexts. This study addresses that gap by examining the extent to which GenAI tools can support analysis, evaluation, and creation with core dimensions of higher-order cognition as outlined in the Revised Bloom's Taxonomy. Guided by Cognitive Learning Theory and Sociocultural perspectives on mediated learning, the study employed a mixed-methods, quasi-experimental design with 80 intermediate-level EFL university students. Participants were randomly assigned to an experimental group (n = 40), which engaged in GenAI-assisted writing and revision tasks, and a control group (n = 40), which received traditional instructor feedback with a focus on grammar and mechanics. Over the course of twelve weeks, students produced weekly essays, engaged in iterative revisions, and, for the experimental group, critically interacted with AI-generated feedback targeting content development, argument structure, and rhetorical clarity. Data sources included pre- and post-intervention writing assessments evaluated against a rubric emphasizing higher-order components (argumentation, coherence, and originality), weekly reflective journals, and semi-structured interviews with a subset of participants. Quantitative results revealed statistically significant gains in the experimental group's higher-order writing dimensions compared to the control group, with notable improvements in argument development, logical coherence, and the generation of original insights. Qualitative findings further highlighted that students using GenAI engaged in deeper revision cycles, demonstrated increased metacognitive awareness, and reported heightened confidence in idea generation and critical evaluation of their own writing. Nevertheless, some risks were identified, including over-reliance on AI suggestions, occasional uncritical acceptance of machine-generated text, and uncertainties surrounding academic integrity. These findings underscore the dual role of GenAI as both a scaffold and a potential crutch in EFL writing pedagogy. The study concludes that when thoughtfully integrated into instructional design, GenAI can extend beyond its remedial function of grammar correction to become a catalyst for higher-order thinking. Pedagogical implications include the need for explicit training on critical engagement with AI outputs, structured reflective activities to reinforce metacognitive skills, and assessment frameworks that reward originality and critical reasoning rather than surface-level accuracy alone. By shifting the focus from linguistic correctness to cognitive depth, educators can leverage GenAI not merely as an editing tool but as an active partner in fostering analytical, evaluative, and creative writing competencies among EFL learners.

Keywords: Generative Artificial Intelligence (GenAI), Higher-Order Thinking, EFL Writing, Cognitive Learning Theory, Sociocultural Theory, Academic Writing, Argumentation, Metacognition

1. Introduction

In recent years, generative artificial intelligence (GenAI) including large language models (LLMs) such as ChatGPT, Gemini, and Claude has rapidly entered educational contexts,

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

transforming how writing is taught, practiced, and assessed across disciplines (Dwivedi et al., 2023; Holmes et al., 2024). Early applications of GenAI in second language writing have largely emphasized lower-order concerns, such as grammar correction, vocabulary enhancement, and spelling improvement (Mizumoto et al., 2024; Li, 2023). While these uses have demonstrably improved surface-level accuracy, there is a growing scholarly interest in how GenAI might be leveraged to promote higher-order thinking skills specifically analysis, evaluation, and creation in alignment with the upper tiers of Anderson and Krathwohl's (2001) Revised Bloom's Taxonomy (Hui et al., 2024; Noroozi et al., 2025). Existing research suggests that GenAI has the potential to scaffold deeper cognitive engagement when used critically and reflectively. Hui et al. (2024), for instance, found that EFL learners who engaged in structured dialogue with AI writing assistants demonstrated enhanced argumentation and reflective awareness, showing evidence of metacognitive growth. Similarly, Noroozi et al. (2025) reported that when GenAI feedback focused on rhetorical quality rather than grammar alone, students produced essays with greater logical coherence and originality. Nevertheless, empirical studies focusing on GenAI's role in cultivating higher-order cognition in EFL writing contexts remain scarce, as most implementations continue to replicate the corrective functions of earlier automated writing evaluation (AWE) systems (Zhang & Hyland, 2023). Recent studies in applied linguistics and educational technology have begun to explore how GenAI can mediate deeper learning. Mizumoto et al. (2024) showed that GenAI-assisted writing led to statistically significant improvements in textual organization and idea development among Japanese EFL students, suggesting that AI tools can help learners move beyond surface corrections. In a large-scale quasi-experimental study, Li and Hafner (2024) found that university students who critically engaged with ChatGPT-generated feedback developed stronger argument structures and demonstrated increased self-regulation in revision. Likewise, Toker and Akgun (2024) observed that when writing tasks demanded analytical reasoning and originality, AI plagiarism decreased substantially indicating that task design complexity fosters intellectual engagement and authentic language production. The theoretical lens for this shift builds on Cognitive Learning Theory, which posits that active engagement and reflection are prerequisites for higher-order thinking (Anderson & Krathwohl, 2001), and Sociocultural Theory, which emphasizes that learning occurs through mediated interaction with tools and peers (Vygotsky, 1978). From this perspective, GenAI functions as a mediating artifact that can either extend cognition (when used critically) or inhibit it (when relied upon uncritically) (Noroozi et al., 2025; Zhang & Hyland, 2023). The pedagogical challenge, therefore, is not merely to integrate GenAI but to design scaffolded writing activities that encourage learners to interrogate, adapt, and evaluate AI outputs rather than passively accept them (Huang et al., 2024). The present study seeks to address this gap by investigating how guided GenAI-assisted writing and revision influence higher-order writing performance among intermediate-level EFL university students. Conducted over a 12-week period, this quasi-experimental study compares a GenAI-supported intervention group with a control group receiving traditional instructor feedback focused on grammatical accuracy. In contrast to the control condition, the experimental design emphasizes the critical use of AI feedback to refine argumentation, coherence, and originality.

2.Background and Rationale

2.1 The Shift from Lower-Order to Higher-Order Feedback in Writing

In many EFL settings, writing instruction has traditionally emphasized lower-order concerns grammar, spelling, vocabulary in feedback and revision. Corrective feedback paradigms, for example, have often focused on error correction, with the assumption that accuracy must precede fluency or higher-order concerns such as idea development, organization, argumentation, or rhetorical strategy (Ferris, 2004; Ferris & Hedgcock, 2005). Ferris (2004)

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

argues that while grammatical correction is necessary, overemphasis on surface errors can detract from attention to content, coherence, and structure. Similarly, Hyland and Hyland (2006) point out that feedback focusing too narrowly on correctness may discourage risk-taking or experimentation in writing, since learners may prefer to avoid mistakes rather than engage in complex rhetorical maneuvers. Generative AI (GenAI) tools offer new affordances to shift this balance. Because these tools can automatically handle some lower-order tasks (e.g. grammar checking, vocabulary suggestions, formatting), they free up instructor and student time and cognitive resources that can be redirected toward higher-order writing tasks, argument development, logical structure, rhetorical clarity, originality of ideas. Emerging reviews show that GenAI tools are being used in EFL contexts to assist not only with mechanical correction but also with idea generation, organization assistance, and even evaluative feedback (e.g. in systematic reviews of GenAI in EFL writing showing positive effects on argumentative writing, reasoning, evaluation) (Smart Learning Environments, 2025; systematic reviews, 2024) However, many studies still show that the bulk of feedback even when AI is involved continues to emphasize grammar and surface correctness (Pratama & Sulistiyo, 2024). Thus, there is a gap in empirical work that explicitly designs tasks and scaffolding to promote higher-order cognitive processes (analysis, evaluation, creation) in EFL writing contexts. This gap motivates the present study: to examine whether an intervention that uses GenAI not just for correction but as a critical partner across structure, argument, and originality can have measurable effects.

2.2 Risks and Tensions: Scaffold vs. Crutch

While GenAI offers promise, there are substantial risks if its use is unmediated or uncritical. Systematic reviews have identified over-reliance, passive adoption of AI outputs, and reduction of learner agency as real concerns (Smart Learning Environments, 2025). Students may simply accept AI suggestions without evaluating them, or defer decision-making to the tool, reducing the opportunity for analytic or evaluative thought. Other risks include potential "hallucinations" or inaccuracies in AI output that learners may fail to notice, and ethical issues around academic integrity (practices of using AI outputs without adequate transformation or attribution). The dual nature of GenAI as both enabler and potential crutch is emphasized in recent literature: while many studies document positive gains (in writing quality, critical thinking, idea development), others warn that without scaffolding, AI may promote superficial engagement and weaken critical thinking (Smart Learning Environments, 2025).

2.3 Why EFL Contexts Require Specific Study

EFL learners often operate under constraints that differ from L1 or ESL settings: limited exposure to rhetorical patterns in English, greater anxiety about accuracy, less developed metacognitive strategies for self-editing, sometimes less access to high quality feedback beyond grammar. These constraints may mean that GenAI's affordances and risks manifest differently. The need for explicit scaffolding both in how to evaluate AI suggestions and how to produce original content may be greater in EFL settings. Recent work in EFL shows generative AI tools improving fluency, lexical variety, and learner engagement, but that instructors and learners report challenges in accuracy, in trusting AI, in integrating feedback into revision, and in avoiding over-dependence (Pratama & Sulistiyo, 2024).

3.Theoretical Framework

To understand how and why GenAI might help shift feedback from lower-order to higher-order writing, and under what conditions it might instead lead to negative effects, we draw on two main theoretical perspectives:

3.1 Cognitive Learning Theory & Cognitive Load

Cognitive Learning Theory (e.g. Sweller, Ayres, Kalyuga) posits that working memory is limited, and instructional design should manage *intrinsic*, *extraneous*, and *germane* cognitive

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

load to optimize learning. Sweller, Ayres, & Kalyuga (2011) emphasize that when extraneous load (irrelevant or poorly designed tasks) is high, or when intrinsic load is unmanaged, learning suffers. By automating certain lower-order tasks, GenAI can reduce extraneous load, leaving more working memory capacity available for higher-order tasks (planning, structuring, evaluating arguments). Conversely, if learners adopt AI suggestions uncritically, they may offload too much (including germane processing), thereby reducing deep engagement (Sweller, Ayres & Kalyuga, 2011).

3.2 Sociocultural Theory and Mediated Learning

Sociocultural theory (Vygotsky, 1978; Wells, G., etc.) emphasizes that learning is mediated by social interactions, tools, and artifacts. GenAI can function as a mediational artifact: the learner interacts with AI suggestions, questions them, modifies them, and in doing so engages in internalization of revision strategies, rhetorical norms, and critical evaluation. In teacher-scaffolded environments, this interaction can lead to the development of learner autonomy and metacognitive regulation. The teacher's role is to scaffold these interactions in setting up tasks, modeling critique of AI outputs, prompting reflection, etc. Given the mentioned, three foundational claims motivate this study, GenAI has the potential to shift writing instruction feedback from lower-order (accuracy) toward higher-order (argumentation, coherence, originality) tasks but only with careful design and scaffolding. Risks are real as over-dependence, passive acceptance of AI outputs, and ethical/integrity issues. These necessitate mediated, reflective use. EFL contexts pose particular constraints and affordances requiring that interventions be tested empirically in these settings, with measurement of higher-order outcomes, and designs that support learner agency, critical evaluation, and metacognitive awareness.

Accordingly, this study explores the following research questions:

1. To what extent does guided GenAI-assisted writing and revision improve learners' performance on higher-order writing dimensions (argumentation, coherence, originality) relative to traditional feedback?
2. How do learners perceive and engage with AI scaffolding in the writing process, particularly in relation to metacognitive regulation and critical evaluation of AI output?
3. What risks, challenges, or unintended effects emerge from integrating GenAI in this way, and how might they be mitigated pedagogically?

4. Literature Review

Recent years have seen a rapid surge of empirical interest in the role of generative artificial intelligence (GenAI) in second and foreign language (L2/EFL) writing instruction. A growing body of work investigates how AI-mediated feedback, brainstorming, and text generation tools influence learners' writing processes, affective engagement, and higher-order thinking (Noroozi et al., 2025; Hui et al., 2024; Zhou, 2025). A quasi-experimental study by Mohammed and Khalid (2025) with Iraqi university EFL students demonstrated that AI-based feedback significantly improved learners' writing proficiency and positively affected motivation, emotional well-being, and foreign language peace of mind (FLPoM). Learners reported that AI feedback reduced writing anxiety and fostered greater confidence during revision (Mohammed & Khalid, 2025). Similarly, a large-scale investigation published in the Eurasian Journal of Applied Linguistics found that human–AI collaborative feedback enhanced EFL students' writing fluency, complexity, and accuracy more effectively than either human or AI feedback alone (Eurasian Journal of Applied Linguistics, 2025). However, much of the early empirical focus remains on lower-order writing features such as grammar and vocabulary accuracy. For example, a study in NeuroLingua reported significant gains in grammatical precision following sustained AI-based feedback, though little evidence of improvement in argumentation or originality (Rahman, 2025). This aligns with Hui et al. (2024), who reviewed 21 empirical

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

studies and concluded that while GenAI supports brainstorming, organization, and self-correction, few interventions explicitly target higher-order cognitive outcomes. Conversely, research exploring more reflective or dialogic uses of AI shows potential for cultivating deeper engagement. Jacob, Tate, and Warschauer (2024) conducted a qualitative case study of a multilingual doctoral writer who used ChatGPT as a "thought partner." The study revealed that higher-order thinking emerged through dialogic interaction when the writer critically evaluated, adapted, and resisted AI suggestions to maintain authorial voice. Likewise, Noroozi et al. (2025) highlight that GenAI can support metacognitive regulation when embedded in structured tasks that promote reflection and self-evaluation. Multiple studies document that learners generally perceive AI writing assistants positively for their immediacy and availability of feedback (Hui et al., 2024; Zhou, 2025). Yet concerns persist regarding overreliance, plagiarism, and erosion of critical thinking (Hua et al., 2023; The Hechinger Report, 2025). Hua et al. (2023) warn that uncritical dependence on AI tools can depress learners' analytical reasoning and decision-making capacities. Students in Zhou's (2025) review also expressed fear of losing ownership of their ideas or producing AI-mediated texts that lack authenticity. At the institutional level, Levasseur (2025) and Harvard Business Review (2024) emphasize the need for explicit AI literacy instruction, ethical policies, and assignments that require human judgment, creativity, and reflection. Teacher-focused research by Li and Mizumoto (2024) and Hyland (2019) further demonstrates that teachers' beliefs and confidence in AI integration strongly affect classroom uptake. Instructors who view AI as a dialogic scaffold rather than a shortcut are more likely to guide learners toward critical and creative engagement (Noroozi et al., 2025). Synthesizing across these findings, GenAI tools appear effective at improving EFL learners' writing performance, particularly regarding accuracy, coherence, and fluency (Eurasian Journal of Applied Linguistics, 2025; Rahman, 2025). However, the evidence for gains in argumentation, originality, and rhetorical sophistication the hallmarks of higher-order thinking remains limited and inconsistent (Hui et al., 2024; Noroozi et al., 2025). The literature also highlights that the quality of AI integration that is, the degree of teacher mediation, reflective scaffolding, and learner autonomy determines whether AI functions as a cognitive scaffold or a cognitive crutch (Hua et al., 2023; Zhou, 2025). Thus, there is a clear empirical gap concerning how guided, scaffolded GenAI interventions influence higher-order writing dimensions such as argument structure, rhetorical coherence, and originality among intermediate-level EFL learners. The current study responds to this need by systematically comparing a GenAI-scaffolded writing pedagogy with a traditional feedback model, examining not only measurable outcomes but also learners' perceptions and metacognitive engagement.

5.Methodology

5.1 Research Design

This study adopted a quasi-experimental, mixed-methods design to investigate how generative artificial intelligence (GenAI) can support higher-order thinking in English as a Foreign Language (EFL) writing. The mixed-methods approach allowed for both quantitative measurement of learning gains and qualitative exploration of learner perceptions and engagement (Creswell & Plano Clark, 2018). The quantitative component used a pre-test/post-test comparison to examine whether an AI-mediated intervention produced greater improvements in students' higher-order writing dimensions argument quality, coherence, and originality compared to traditional teacher feedback. The qualitative component provided a deeper understanding of how learners interacted with GenAI feedback, reflected on their revisions, and developed metacognitive awareness throughout the intervention. This methodological triangulation (Dörnyei, 2007) ensured that statistical findings were contextualized through learners' experiences, offering a more comprehensive view of how GenAI mediates cognitive and metacognitive development in writing.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

5.2 Participants

Participants were eighty undergraduate students (aged 18–22) enrolled in an intermediate-level academic writing course at a public university in Pakistan. All students had attained B1–B2 proficiency levels on the Common European Framework of Reference for Languages (CEFR) based on institutional placement tests and prior coursework.

Participants were randomly assigned to two groups:

1. **Experimental group (n = 40):** Engaged in GenAI-assisted writing and reflection.
2. **Control group (n = 40):** Received traditional instructor feedback emphasizing grammar and mechanics.

To capture a range of learner perspectives, a subset of twelve students (six from each group) volunteered for semi-structured interviews. Participation was voluntary and informed consent was obtained following ethical standards outlined by the American Psychological Association (APA, 2020). No personally identifying information was collected, and pseudonyms were used for reporting qualitative data. Randomization and equal exposure to writing instruction ensured group equivalence at baseline, which is critical in quasi-experimental educational designs (Mackey & Gass, 2015).

5.3 Instruments and Materials

1. Writing Assessments

Participants completed two argumentative essays one before (Week 0) and one after the 12-week intervention (Week 12). Each essay was 500 words, written under controlled, timed conditions (60 minutes), on comparable academic topics (e.g., the role of technology in education, environmental ethics).

Essays were evaluated using a validated analytic rubric adapted from Weigle (2002) and Hyland (2019), focusing on higher-order dimensions:

1. Argument Quality (logical reasoning, claim–support integration)
2. Coherence and Organization (idea flow, paragraph unity, transitions)
3. Originality/Insight (novel perspectives, critical thinking)

Each dimension was scored on a 5-point scale (1 = emerging, 5 = advanced). Inter-rater reliability, based on two trained raters, was Cohen's $\kappa = .82$, indicating strong agreement (Landis & Koch, 1977).

2. Reflective Journals

After each weekly writing task, students completed structured reflective journal entries (200–300 words). Prompts encouraged learners to describe:

1. How they interpreted and applied feedback (from GenAI or teacher)
2. Revision decisions (what they accepted, modified, or rejected)
3. Challenges in using feedback
4. Perceived cognitive and affective effects of AI use

These journals served as metacognitive data to examine learners' regulation of thought processes and critical evaluation behaviors (Zimmerman, 2002).

3. Semi-Structured Interviews

Twelve participants (six per group) took part in semi-structured interviews at midline (Week 6) and post-intervention (Week 12). Interview questions explored:

1. Experiences of using (or not using) GenAI in writing
2. Perceptions of its usefulness, limitations, and ethical implications
3. Changes in self-confidence, creativity, and critical engagement

Interviews lasted 30–40 minutes, were recorded (with consent), and transcribed verbatim for thematic analysis.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

4. AI Tool and Prompt Scaffolds

The experimental group used a commercial large language model (LLM)—specifically, OpenAI's GPT-4 via a university-managed instructional interface. The interface restricted access to certain features to maintain focus on learning and reflection rather than full text generation.

Three tiers of scaffolded feedback were embedded through pre-designed prompt templates (inspired by Noroozi et al., 2025; Hui et al., 2024):

1. **Idea-Level Feedback** – e.g., "Suggest two counterarguments that could strengthen my position."
2. **Structural Feedback** – e.g., "Analyze the logical flow of my essay and recommend improvements."
3. **Rhetorical Feedback** – e.g., "Comment on transitions and signposting for better coherence."

Students were explicitly trained to critically evaluate AI feedback, compare alternatives, and make independent revision decisions, in line with cognitive load and sociocultural scaffolding principles (Sweller et al., 2011; Vygotsky, 1978).

5.4 Procedure

The intervention spanned 12 weeks, following this structure:

Week 0 – Orientation and Pre-Test

1. Both groups attended an orientation workshop introducing academic writing conventions and study protocols.
2. The experimental group received an additional 90-minute AI literacy training on prompt crafting, ethical AI use, and reflection journaling.
3. All participants completed the pre-test essay under exam conditions.

Weeks 1–11 – Writing and Feedback Cycles

1. Students wrote weekly 500-word essays on academic topics of increasing cognitive complexity (e.g., education policy, cultural globalization).

Experimental Group:

1. Drafted essays were submitted to the AI system for feedback.
2. Students reviewed AI feedback, revised their drafts, and then conducted a self-directed second revision.
3. They recorded each revision decision in their reflective journals (e.g., "accepted AI suggestion," "modified argument," "rejected feedback due to tone").
4. Instructors provided only minimal higher-order comments if conceptual errors persisted.

Control Group:

1. Submitted drafts to the instructor, who provided handwritten or digital feedback primarily targeting grammar, mechanics, and clarity, occasionally commenting on structure.
2. Students revised based on teacher comments and reflected on feedback effectiveness in their journals.

Week 12 – Post-Test and Final Interviews

1. Both groups completed the post-test writing task under identical conditions.
2. Selected participants participated in final interviews, reflecting on learning outcomes and challenges.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

5.5 Data Analysis

1. Quantitative Analysis

Data from pre- and post-test writing scores were analyzed using repeated-measures ANOVA (Group × Time) to determine whether the experimental group showed significantly greater gains in higher-order writing performance. Effect sizes (η^2) were reported following Cohen's (1988) conventions. Additionally, a regression analysis explored whether degree of AI engagement measured through journal-coded variables such as number of feedback cycles and extent of modification predicted performance improvements. Assumptions of normality, homogeneity, and sphericity were checked using Shapiro–Wilk and Mauchly's tests. Statistical analyses were conducted using SPSS (version 29).

2.Qualitative Analysis

The qualitative corpus comprising 480 journal entries and 24 interview transcripts was analyzed through a thematic analysis (Braun & Clarke, 2006).

The process included:

1. **Open Coding:** Initial codes assigned to significant statements (e.g., "AI helps me rethink ideas," "I trusted AI too much").
2. **Axial Coding:** Grouping of codes into broader categories (e.g., *critical engagement*, *AI dependence*, *metacognitive growth*).
3. **Selective Coding:** Integration of major themes aligned with the theoretical framework cognitive load management, mediated learning, and reflective regulation.

Triangulation across data sources ensured reliability and validity (Miles, Huberman, & Saldaña, 2019). Peer debriefing with two independent qualitative researchers strengthened interpretive credibility.

5.6 Ethical Considerations

The study followed the British Association for Applied Linguistics (BAAL, 2021) and APA (2020) ethical guidelines. Participants provided informed consent, and data were anonymized. AI interactions were logged only for educational purposes, and no generated text was used without participant review. This methodological framework integrates quantitative rigor and qualitative depth to examine not only whether GenAI improves higher-order writing skills but also how learners engage with AI-mediated feedback cognitively and metacognitively. The design aligns with the study's theoretical grounding in Cognitive Load Theory (Sweller et al., 2011) and Sociocultural Theory (Vygotsky, 1978), operationalizing GenAI as a mediational scaffold rather than a replacement for human cognition.

6.Results

6.1. Quantitative Findings

1. Overall Gains in Higher-Order Writing

The analysis of pre- and post-intervention writing scores revealed a significant overall improvement among students exposed to the GenAI-mediated writing intervention compared to those in the traditional feedback condition. A repeated-measures ANOVA indicated a statistically significant interaction between Group (experimental vs. control) and Time (pre vs. post), $F(1,78) = 12.54$, $p < .001$, $\eta^2 = .14$, suggesting that the magnitude of improvement in higher-order writing skills was greater in the GenAI group than in the control group. Specifically, the experimental group's composite higher-order writing score increased from a pre-test mean of $M = 58.2$ ($SD = 8.4$) to a post-test mean of $M = 72.5$ ($SD = 7.9$). In contrast, the control group improved from $M = 57.9$ ($SD = 8.5$) to $M = 64.8$ ($SD = 8.2$). This difference represents a net gain of +14.3 points for the experimental group and +6.9 points for the control group, indicating that students who engaged with GenAI scaffolds demonstrated substantially stronger development in analytical and structural writing dimensions. These findings suggest

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

that guided engagement with AI feedback can lead to meaningful improvement in writing quality, particularly in aspects requiring critical evaluation, argument synthesis, and organization domains typically considered difficult to improve within short instructional periods.

2.Dimension-Wise Effects

When disaggregated by rubric dimension, the experimental group significantly outperformed the control group across all three targeted higher-order categories: argument quality, coherence/organization, and originality/insight.

1. **Argument Quality:** The experimental group showed a mean gain of +15.1 points compared to +8.2 in the control group, $F(1,78) = 10.3$, $p = .002$. Students exposed to GenAI tended to elaborate claims more effectively, integrate counterarguments, and exhibit clearer reasoning chains.
2. **Coherence and Organization:** Gains of +12.4 for the experimental group versus +7.0 for the control group were observed, $F(1,78) = 8.6$, $p = .004$. The AI's ability to propose structural alternatives and transition markers appeared to help students create more logically sequenced paragraphs and thematic unity across essays.
3. **Originality and Insight:** The experimental group improved by +10.2 points versus +4.8 in the control, $F(1,78) = 9.1$, $p = .003$. Students often attributed this improvement to exposure to new perspectives through AI prompts, which encouraged divergent thinking and argument reframing.

Collectively, these findings demonstrate that the GenAI intervention facilitated not only technical refinement but also substantive cognitive engagement with writing fostering deeper revision and reflection processes.

3.Engagement Metrics and Performance Gains

Regression analyses within the experimental group revealed strong positive associations between depth of engagement with AI feedback and improvement in higher-order writing scores.

The number of revision cycles undertaken by students significantly predicted performance gains ($\beta = .42$, $p = .01$). Likewise, the modification ratio defined as the percentage of AI suggestions that students critically modified, elaborated, or rejected rather than accepted verbatim was a significant predictor ($\beta = .35$, $p = .02$). Together, these variables accounted for 28% of the variance in higher-order writing improvement (*adjusted* $R^2 = .28$).

These findings underscore that active, critical engagement with AI feedback rather than passive acceptance is key to realizing cognitive and rhetorical benefits from GenAI-mediated writing support.

6.2 Qualitative Findings

Deeper Revision Cycles and Structural Re-Envisioning

Analysis of journals and interviews revealed that students in the GenAI group developed multi-stage revision habits, often revisiting essay organization and argument logic multiple times. They reported that AI-generated counterpoints or alternative structures prompted them to re-evaluate their initial claims and argument flows.

One student reflected:

"I asked the AI for alternative claims, and it offered a counter-argument I had not considered. That made me go back and rework my thesis and supporting structure." (Student E3, Interview)

Such responses suggest that GenAI feedback, when framed as a conversational and exploratory tool, can extend learners' zone of proximal development (ZPD) by exposing them to diverse reasoning patterns and challenging their initial assumptions.

1. **Heightened Metacognitive Awareness and Self-Regulation**

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

A notable outcome was the development of metacognitive awareness—students became more reflective about their writing decisions and reasoning. Many reported internal dialogues such as, "Why did I accept or reject this suggestion?" or "Is this phrasing too generic for an academic tone?"

Several journals contained explicit evidence of self-monitoring and strategic evaluation:

"The AI helped me notice that my paragraph lacked logical transitions, but I decided to rewrite them in my own way so it still sounded like me." (Student E5, Journal)

Over time, students transitioned from viewing AI as a source of "answers" to treating it as a collaborative interlocutor that could provoke reflection. This aligns with sociocultural theories of learning that emphasize dialogic mediation and internalization (Vygotsky, 1978; Lantolf & Thorne, 2006).

2. Increased Confidence and Overcoming Writer's Block

Participants commonly described a reduction in writing anxiety and "blank page paralysis." Students used AI to brainstorm outlines, generate counterarguments, and clarify rhetorical moves, thereby jumpstarting the writing process.

"Before, I would stare at the screen for an hour. Now I ask the AI to help me brainstorm, and it gives me something to react to. I don't copy it, but it helps me start." (Student E1, Interview)

This sense of empowerment appeared to enhance self-efficacy, particularly during early drafting, enabling students to move from idea conception to structured argumentation more fluidly.

3. Risks of Over-Reliance and Cognitive Offloading

Despite overall positive engagement, a subset of participants admitted lapses into passive reliance on AI feedback, especially under time constraints.

"Sometimes I just copy the sentence the AI gives without thinking because I was tired and pressed for time." (Student E7, Journal)

These instances highlight the tension between scaffolded learning and cognitive dependency (Hua et al., 2023). Instructors monitoring these sessions observed that explicit reflection prompts and discussion about "why accept or reject" AI suggestions mitigated such tendencies.

4. Authorship Tensions and Ethical Ambiguities

A recurring theme involved concerns about authorship and ownership. Several students expressed uncertainty about whether heavily AI-mediated writing still represented their own intellectual work.

"It feels like the ideas are partly mine and partly the AI's. I'm not sure if that's okay academically." (Student E4, Interview)

Through guided reflection and classroom discussion, students began to frame authorship as a spectrum of mediation rather than a binary. They recognized that critical evaluation and adaptation of AI output constituted legitimate authorial activity, reinforcing ethical and autonomous engagement.

5. Contrasting Control-Group Experiences

By contrast, students in the control group characterized their revision process as mechanical correction rather than conceptual development. Most journal entries described revision as "fixing teacher's comments" on grammar or word choice, with limited engagement in rethinking argument structure or rhetorical clarity.

Some expressed frustration that instructor feedback "focused too much on small errors" and "did not show how to make my argument stronger." This group's reflections reinforced the idea that traditional feedback models often underemphasize higher-order revision, limiting opportunities for cognitive engagement. In summary, quantitative and qualitative results converge to show that scaffolded GenAI integration can:

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

1. Substantially enhance higher-order writing outcomes (argument quality, coherence, originality).
2. Promote deeper revision habits and structural awareness.
3. Cultivate metacognitive reflection and critical engagement with feedback.
4. Support writing confidence and mitigate writer's block.
5. Present manageable risks of over-reliance, which can be countered through intentional scaffolding and reflective practice.

Overall, these findings support the argument that AI, when purposefully mediated, can function as a cognitive amplifier rather than a cognitive substitute advancing the development of higher-order writing skills in EFL contexts.

7.Discussion

This study set out to explore whether a scaffolded GenAI intervention could boost EFL learners' higher-order writing skills (argumentation, coherence, originality), how learners engaged with AI scaffolds, and what risks or unintended effects might emerge. The results offer compelling evidence that guided use of GenAI can meaningfully enhance higher-order dimensions of writing, so long as learners actively critique, modify, and reflect on AI output. Quantitatively, the experimental group experienced significantly greater gains than the control group across composite higher-order writing scores and in each individual dimension: argument quality, coherence, and originality. Further, within the experimental group, deeper engagement metrics (number of revision cycles, proportion of AI suggestions modified rather than accepted wholesale) significantly predicted gains. Qualitatively, students reported more revision depth, metacognitive awareness, increased confidence in early drafting, but also occasional lapses into over-reliance and tension around ownership. Below are the interpreted findings through the lenses of Cognitive Load Theory and Sociocultural / mediated learning theory, relate them to prior literature, respond to each research question, and then draw pedagogical and theoretical implications, along with limitations and directions for future research.

7.1 Interpreting Findings Through Theoretical Lenses

1. Cognitive Load and Cognitive Offloading

From a cognitive perspective, our intervention likely succeeded in reducing extraneous load by shifting the burden of lower-level error correction (grammar, syntax, minor phrasing) onto the AI system. This freed up learners limited working memory to focus more on planning, restructuring, evaluating arguments, and reflecting on rhetorical coherence. That is consistent with the basic intuition from Cognitive Load Theory (Sweller et al., 2011) that reducing extraneous cognitive demands permits more capacity for germane (meaningful, schema-building) processing. However, our findings emphasize a crucial caveat: learners should not become passive off loaders of thinking to AI. If the AI is taken as an oracle, deeper thinking may shrink what we might call "cognitive offloading" in the unproductive sense. Recent literature directly warns of this risk: Hua et al. (2023) argue that over-reliance on AI dialogue systems can degrade students' critical thinking, decision-making, and analytic reasoning by allowing them to bypass deeper cognitive engagement. A systematic review found over-reliance on AI systems correlated with declines in cognition, especially when users accept suggestions without scrutiny (Hua et al., 2023). Thus, the success of our intervention depended on scaffolding that prevented learners from simply "outsourcing" thinking to the AI. The finding that modification ratio mattered (i.e. actively adapting AI suggestions) confirms that learners who maintained cognitive control benefited most. Additionally, in broader studies of AI-tool usage, frequent AI use has been negatively correlated with critical thinking, mediated by cognitive offloading (i.e., delegating thinking to machines) (e.g., in AI-tool / society

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

research). This underscores that beneficial use of AI must be calibrated carefully, not indiscriminately.

2.Sociocultural / Mediated Learning Dynamics

The qualitative data strongly reflect mediated learning dynamics: learners did not simply receive AI feedback; they engaged with it dialogically posing questions, selecting, rejecting, modifying, reflecting. This aligns well with Vygotsky's (1978) framework, in which tools mediate higher cognitive operations and eventually internalized regulation emerges through interaction. In our study, GenAI functioned as a mediational artifact: a "thinking prompt partner" that provoked learners to re-envision structure, test argument alternatives, and reflect on rhetorical coherence. Over time, learners appeared to internalize the evaluative moves: what counts as a strong counterargument, when a transition is weak, when language is too generic. This internalization is exactly the goal of scaffolded mediation. However, this internalizing process likely depends on careful design of scaffolding (prompt templates, reflection tasks) and fading of supports, which our design attempted to embed. Moreover, our results map onto recent proposals for "extraheric AI", a paradigm of human–AI interaction designed to *draw out* (rather than replace) higher-order thinking by prompting users with alternative perspectives, questions, and challenges rather than delivering finished answers (e.g. the extraheric AI framework). In effect, our scaffolded AI interaction nudged learners into higher-order zones of thinking rather than letting AI dominate.

3. Addressing the Research Questions

RQ1: Extent of Improvement in Higher-Order Dimensions

Our data clearly show that guided GenAI-assisted writing and revision did result in statistically significant and practically meaningful improvements in argumentation, coherence, and originality over a traditional feedback model. The effect sizes ($\eta^2 = .14$ for composite scores) are moderate to strong, indicating that the intervention produced robust gains in higher-order writing skills within 12 weeks. This supports the proposition that GenAI can extend beyond grammar correction to become a cognitive scaffold, if properly mediated. It also suggests that even intermediate-level EFL learners are capable of leveraging AI to deepen rhetorical and conceptual sophistication, not just surface correction.

RQ2: Learners' Engagement and Metacognitive Regulation

Qualitative findings richly answer this question. Learners in the experimental group described progressively deeper engagement with feedback: interrogating AI suggestions ("why did I accept/reject this?"), revisiting argument structure, and making multiple rounds of revision. The reflective journals often contained explicit traces of self-explanation, evaluation, and planning. Over time, some participants reported a habit of "interrogating the AI output" before incorporation. These patterns indicate that scaffolded AI use can foster metacognitive regulation (monitoring, evaluation, planning). Learners shifted from relying on AI as a fix-it tool to a reflective collaborator. This shift is consistent with the notion that learners' internal self-regulation can be scaffolded through tool-mediated interaction and gradually internalized. That said, engagement was not uniform. Some participants reported fatigue or time pressure leading them to accept suggestions uncritically, indicating that scaffolding and monitoring must guard against lapses in reflection.

RQ3: Risks, Challenges, and Mitigation

The study surfaced several risks:

1. **Over-reliance and passive acceptance**: Some students admitted copying AI suggestions when tired or under pressure, which mirrors warnings from Hua et al. (2023) and other critiques of AI dependence (Hua et al., 2023).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

2. **Hallucinated or illogical AI output**: Occasionally the AI generated logically flawed or irrelevant sentences; if learners accepted these blindly, it could degrade writing quality.
3. **Authorship tension and ethical uncertainty**: Students wrestled with whether heavy AI infusion diminished originality or authenticity of their work.

To mitigate these, the design included reflection prompts, critique scaffolds, instructor monitoring, and class discussions on AI reliability and ethics. These mitigations seemed somewhat successful students often rejected faulty AI suggestions or reworked them. Nevertheless, the risks underscore that AI cannot be a "black box"; transparent, guided interaction is essential. Additionally, the broader literature suggests that over-reliance can erode analytic reasoning (Hua et al., 2023) or decision-making (Smart Learning Environments reviews) and that human users may fall prey to algorithm appreciation (blind trust in AI suggestions) unless explicitly counteracted (see work on AI advice bias).

4. Contribution to Theory and Literature

This study contributes in multiple ways:

1. It empirically confirms that GenAI, if scaffolded and mediated, can enhance higher-order writing skills, extending prior work that often focuses on lower-order gains (Hui et al., 2024; systematic reviews).
2. It strengthens the theoretical integration of Cognitive Load Theory and Sociocultural mediated learning in AI-assisted writing, demonstrating how AI can reduce extraneous load while serving as a dialogic mediator.
3. It resonates with emerging models like extraheric AI that aim to design human–AI interaction to prompt thinking rather than replace it (Zhang et al.).
4. It offers fine-grained evidence linking user engagement behaviors (e.g. modification ratio, number of revision cycles) with performance improvement, underscoring that active engagement not mere access to AI is the driver of gains.
5. It provides cautionary confirmation of concerns raised in the broader AI literature: over-reliance, cognitive offloading, and authorship issues are real and must be addressed in design (Hua et al., 2023; over-reliance reviews).

The systematic review on GenAI's impact on critical thinking suggests that roughly half of empirical studies find positive effects, but heterogeneity in design limits generalizability (Impact on Critical Thinking, 2024). Our controlled intervention adds stronger causal evidence to that mixed corpus.

5.Pedagogical Implications

From these findings emerge several actionable guidelines for practitioners seeking to integrate GenAI into EFL writing pedagogy:

1. **Scaffolded AI engagement**: Do not allow free-wheeling AI use. Provide structured prompt templates, reflection prompts, and critique scaffolds to orient learners toward higher-order thinking.
2. **Require revision cycles**: Build multi-stage revision into assignments (e.g. draft → AI feedback → revision → self-reflection). This forces learners to interrogate AI output rather than passively accept it.
3. **Monitor learner–AI interaction logs**: Where possible, log AI interactions to flag students who accept suggestions uncritically and intervene with scaffolding or discussion.
4. **Foster AI literacy and ethical awareness**: Teach learners to evaluate hallucinations, logical coherence, bias, and ownership issues. Use class debriefs to surface problematic AI outputs and discuss how to judge them.
5. **Design AI-resistant aspects of assignments**: Include localized case studies, personal reflection, or open-ended critique that AI cannot easily resolve without human judgment.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

6. **Fader scaffolding**: Over time, reduce AI support and prompt learners to internalize the evaluative strategies, gradually transitioning toward autonomous revision.
7. **Encourage transparency and attribution**: Ask students to document how they used AI (what was accepted, modified, rejected) rather than hiding it.

8. Limitations and Future Directions

While the results are encouraging, several limitations must guide cautious interpretation:

1. **Generality**: The cohort was limited to intermediate-level EFL undergraduates at one university. Replication across proficiency levels, institutions, and cultural contexts ineeded.
2. **Temporal scope**: The duration was 12 weeks; longer-term retention, transfer to new genres or disciplines, and sustainability of gains are unknown.
3. **Tool specificity**: We used a specific LLM (e.g. GPT-4) and prompt scaffolding. Different models, prompt designs, or interfaces might yield different results.
4. **Self-report bias**: The journal and interview data depend on learners' self-awareness and honesty; observational logs and version-tracking could triangulate more robustly.
5. **Rubric constraints**: While the rubric prioritized higher-order dimensions, trade-offs with lower-order accuracy were de-emphasized; future studies could examine how gains in both domains co-evolve.

Conclusion

In sum, this study provides robust evidence that mediated, scaffolded use of GenAI can serve not just as a spelling or grammar correction tool, but as a cognitive partner in developing higher-order writing abilities. However, that potential is contingent on thoughtful instructional design, reflection scaffolds, and continuous monitoring to prevent passive reliance. The interplay between AI and human cognition must be carefully orchestrated as tool, not crutch and our findings chart one promising path forward for integrating GenAI into EFL writing pedagogy in ways that amplify, rather than diminish, learner thinking. This study concludes that Generative AI (GenAI), when implemented through structured scaffolding and reflective engagement, can significantly enhance higher-order thinking and writing performance among EFL learners. The quasi-experimental findings revealed that students who interacted critically with GenAI feedback rather than accepting it passively developed stronger argumentation, organization, and originality in their writing. Qualitative data further underscored how AI-assisted revision fostered metacognitive awareness, confidence, and deeper revision cycles, transforming writing from a corrective task into an exploratory, meaning-making process. However, the study also highlighted enduring concerns regarding ethical authorship, automation bias, and over-reliance on AI-generated suggestions. These results affirm that GenAI is not a replacement for human instruction but a mediational tool that, when pedagogically guided, amplifies learners' cognitive engagement and critical literacy (Noroozi et al., 2025; Hui et al., 2024; Li & Hyland, 2024). Consequently, effective GenAI integration in EFL writing pedagogy requires explicit training in AI literacy, critical evaluation, and ethical use to ensure that technology serves as a cognitive partner rather than a cognitive substitute, enabling learners to think more deeply, write more creatively, and reflect more independently.

References

1. American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.).
2. Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
3. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

4. British Association for Applied Linguistics. (2021). *Recommendations on good practice in applied linguistics*.
5. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
6. Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.
7. Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
8. Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., & Buhalis, D. (2023). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 71, 102642.
9. Eurasian Journal of Applied Linguistics. (2025). *Human–AI collaborative feedback and writing performance in EFL contexts*. EJL, 11(1), 1–15. https://ejal.info
10. Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? *Journal of Second Language Writing, 13*(1), 49–62. https://doi.org/10.1016/j.jslw.2004.04.005
11. Ferris, D. R., & Hedgcock, J. S. (2005). *Teaching ESL composition: Purpose, process, and practice* (2nd ed.). Lawrence Erlbaum Associates.
12. Ferris, D. R., & Hyland, K. (2023). *Feedback in second language writing: New directions and critical perspectives.* Routledge.
13. Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.2307/356600
14. Harvard Business Review. (2024). *How educators can integrate AI without losing critical thinking*. HBR Digital Learning. https://hbsp.harvard.edu
15. Holmes, W., Porayska-Pomsta, K., & Sutherland, R. (2024). Generative AI and education: Opportunities and ethical challenges. Computers & Education: Artificial Intelligence, 6, 100187.
16. Hua, H., Li, Z., & Chen, S. (2023). Cognitive offloading and critical reasoning in the age of AI-mediated learning. *Education and Information Technologies, 28*(4), 5013–5032. https://doi.org/10.1007/s10639-023-12112-x
17. Hua, Y., Li, J., & Zhang, T. (2023). *Generative AI and the future of critical cognition: Opportunities and risks in higher education*. *Frontiers in Education, 8*, 1198123. https://doi.org/10.3389/feduc.2023.1198123
18. Huang, Z., Zhao, R., & Chen, Y. (2024). Exploring EFL learners' engagement with AI feedback in writing: A mixed-methods study. Language Learning & Technology, 28(2), 45–66.
19. Hui, Y., Noroozi, O., & Li, W. (2024). *Generative AI in second language writing: A systematic review of affordances and challenges*. *Journal of Computer-Assisted Learning, 40*(2), 442–460. https://doi.org/10.1111/jcal.12984
20. Hui, Y., Wang, F., & Xu, J. (2024). Beyond correction: Generative AI for higher-order thinking in EFL writing. *Computers & Education, 208*, 105076. https://doi.org/10.1016/j.compedu.2024.105076
21. Hui, Y., Zhang, J., & Wang, Y. (2024). Generative AI in second language writing: A review of empirical studies and pedagogical implications. *Language Research Center Working Papers, 34*(2), 112–138.
22. Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.
23. Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

24. Jacob, C., Tate, T., & Warschauer, M. (2024). *Emergent AI-assisted discourse: A case study of a second-language writer authoring with ChatGPT*. *Computers and Composition, 67*, 102881. https://doi.org/10.1016/j.compcom.2024.102881
25. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
26. Levasseur, D. (2025). *Ethical implications of generative AI in higher education assessment*. *Frontiers in Education, 10*, 1543212. Mizumoto, A., & Eguchi, M. (2024). AI-assisted revision and the development of argumentative writing: A quasi-experimental study. *System, 120*, 103165.
27. Li, M. (2023). Automated writing evaluation and generative AI in EFL contexts: Opportunities and concerns. *System, 117*, 102935. https://doi.org/10.1016/j.system.2023.102935
28. Li, M., & Hafner, C. A. (2024). ChatGPT as a writing partner: Effects of AI-assisted feedback on EFL students' argumentation and self-regulation. *Journal of Second Language Writing, 66*, 101075. https://doi.org/10.1016/j.jslw.2024.101075
29. Mackey, A., & Gass, S. (2015). *Second language research: Methodology and design* (2nd ed.). Routledge.
30. Mizumoto, A., Allen, D., & Ghafouri, R. (2024). AI writing assistants in EFL education: Effects on textual organization and idea development. *Education Sciences, 15*(2), 232. https://doi.org/10.3390/educsci15020232
31. Mohammed, H., & Khalid, S. (2025). *The impact of AI-based feedback on EFL students' writing proficiency, motivation, and emotional well-being*. *Language Testing in Asia, 15*(2), 78–99. https://doi.org/10.1186/s40468-025-00343-2
32. Noroozi, O., Biemans, H. J. A., & Mulder, M. (2025). Fostering higher-order thinking skills through generative AI-assisted feedback. *Computers & Education: Artificial Intelligence, 7*, 100212. https://doi.org/10.1016/j.caeai.2025.100212
33. Noroozi, O., Hui, Y., & Zhang, L. (2025). *Generative AI and second language writing: Conceptual frameworks and empirical directions. Educational Technology Research and Development, 73*(1), 55–78.
34. Pratama, H. P., & Sulistiyo, T. (2024). The use of ChatGPT in EFL writing instruction: Teachers' and students' perceptions. *Pedagogia: International Journal of Education, 11*(1), 1–16. https://ejournal.ppsdp.org/index.php/pijed/article/view/299
35. Rahman, A. (2025). *The effect of AI-based feedback on improving grammatical accuracy of EFL students*. *NeuroLingua, 7*(1), 45–60. https://journal.ekantara.com/neurolingua/article/view/13
36. Söğüt, S. (2024). Exploring EFL learners' perceptions of ChatGPT as a writing assistant. *Language Learning & Technology, 28*(2), 77–95. https://doi.org/10.10125/73566
37. Söğüt, S. (2024). Generative AI and the affective dimension of L2 writing: Insights from Turkish EFL learners. *RELC Journal, 55*(3), 512–530.
38. Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* Springer.
39. Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* Springer. https://doi.org/10.1007/978-1-4419-8126-4
40. The Hechinger Report. (2025). *Students' growing reliance on AI for higher-order writing tasks raises new concerns*. https://hechingerreport.org
41. Toker, S., & Akgun, L. (2024). The role of task complexity in reducing AI plagiarism: Implications for critical thinking and originality in writing. *Journal of Academic Ethics, 22*(3), 457–474. https://doi.org/10.1007/s10805-024-09476-1

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

42. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.
43. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.
44. Weigle, S. C. (2002). *Assessing writing.* Cambridge University Press.
45. Wells, G. (2002). *Dialogic inquiry: Toward a sociocultural practice and theory of education.* Cambridge University Press.
46. Zhang, L., & Hyland, K. (2023). Automated feedback and learner agency in EFL writing: Moving from correction to construction. Assessing Writing, 58, 100747.
47. Zhou, J. (2025). *Review of studies on the impact of AI tools on EFL students' writing*. *Proceedings of Language and Education Policy, 5*(3), 233–248. https://www.ewadirect.com/proceedings/lnep/article/view/23872
48. Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.