

Fostering Higher-Order Thinking Skills in ESL Classrooms Through Ai-Supported Writing Tasks

Jurayeva Zuhra

Chirchik State Pedagogical University, Uzbekistan

Received: 10 April 2025; **Accepted:** 06 May 2025; **Published:** 08 June 2025

Abstract: The rapid diffusion of generative artificial-intelligence (AI) tools such as ChatGPT is reshaping language-learning ecologies. While early adopters have focused on gains in fluency and accuracy, a more compelling pedagogical question is whether AI can cultivate higher-order thinking skills (HOTS)—analysis, evaluation and creative synthesis—in English as a Second Language (ESL) learners. The present study examines an intervention in which undergraduate ESL students completed a sequence of argumentative and reflective writing tasks mediated by an AI co-writer that provided dynamic scaffolding, metacognitive prompts and automated discourse analysis. Grounded in Bloom’s revised taxonomy and socio-constructivist learning theory, the mixed-methods design combined quasi-experimental pre- and post-testing with thematic analysis of learner journals. Quantitative results showed statistically significant improvements ($p < 0.05$) in students’ HOTS rubric scores compared with a control group engaged in traditional peer-review cycles. Qualitative data revealed heightened metalinguistic awareness and strategic risk-taking in idea development. The findings suggest that, when carefully orchestrated, AI-supported writing can transcend mere linguistic assistance and become a catalyst for deeper cognitive processing. Pedagogical implications and design principles for AI-enhanced assessments that preserve academic integrity are discussed.

Keywords: Higher-order thinking; AI-supported writing; ESL pedagogy; ChatGPT; Bloom’s taxonomy; academic integrity; mixed-methods.

Introduction: Educators have long aspired to move second-language writing instruction beyond sentence-level accuracy toward the cultivation of analytical and creative reasoning. Yet empirical evidence indicates that ESL classrooms often remain anchored in lower-level cognitive activities, partly because teachers must devote considerable time to error correction. The emergence of large language models (LLMs) offers unprecedented opportunities to redistribute this cognitive load. Recent studies conducted in undergraduate writing courses demonstrate that students who receive AI-mediated formative feedback outperform peers on measures of evaluative reasoning and argumentation quality. However, scepticism persists regarding over-reliance on algorithmic text generation and the potential erosion of original thought.

Higher-order thinking skills, situated at the apex of Bloom’s taxonomy, encompass analyzing patterns,

evaluating evidence and producing novel syntheses. Contemporary scholarship urges a re-examination of these taxonomic categories in light of generative AI, positing that strategic prompting can prompt learners to operate at “create” and “evaluate” levels more consistently. Nevertheless, there is limited classroom-based research that operationalises HOTS explicitly within AI-supported task design for ESL populations. Addressing this gap, the present study investigates whether structured human–AI collaboration during writing tasks can measurably enhance HOTS while maintaining language-development objectives.

This investigation adopted an explanatory-sequential mixed-methods design that combined a quasi-experimental component with qualitative process tracing to elucidate how artificial-intelligence mediation affects higher-order thinking during second-language writing. The study unfolded over a fifteen-week semester in two intact sections of “English for

Academic Purposes II” at a public university in Central Asia. Prior to group assignment, students ($N = 54$) completed the TOEFL ITP to ensure baseline comparability ($M = 487$, $SD = 21$); Levene’s test confirmed homogeneity of variance ($p = 0.64$). The experimental cohort ($n = 27$) engaged in AI-supported writing cycles, whereas the control cohort ($n = 27$) followed a traditional peer-review model.

Instructional materials were fully parallel across conditions: three genre modules—argumentative essay, problem–solution report and reflective blog—each allotted four weeks. In the experimental group, students interacted with ChatGPT-4 on a closed intranet at three scaffolded checkpoints within every module: (1) ideation, where an AI prompt encouraged abductive reasoning and analogical transfer; (2) outline refinement, during which students negotiated thesis–evidence alignments through Socratic questioning supplied by the AI; and (3) metacognitive reflection, where learners interrogated the cognitive strategies they employed. All AI dialogues were logged and later subjected to discourse-analytic coding. Teacher facilitation was standardised through a protocol that limited direct linguistic correction, thereby foregrounding cognitive mediation.

Data collection occurred at three junctures. First, pre- and post-intervention essays were evaluated with a four-dimension HOTS rubric adapted from the Critical Thinking Assessment Test; two doctoral-level raters, blind to condition, achieved $\kappa = 0.86$. Second, weekly learning journals elicited introspective accounts of reasoning processes; 312 entries were thematically coded via NVivo, following Braun and Clarke’s six-phase procedure. Third, screen-capture recordings of AI interactions (≈ 27 h) enabled fine-grained analysis of epistemic moves. Quantitative effects were tested with ANCOVA, using initial HOTS scores as covariates; effect sizes were interpreted according to Cohen’s conventions. Credibility of qualitative findings was bolstered through investigator triangulation and member checking in post-semester focus groups.

HOTS were assessed through a validated rubric adapted from the Critical Thinking Assessment Test, encompassing four dimensions: (1) depth of analysis, (2) evidence evaluation, (3) creative integration of sources and (4) reflective self-regulation. Two blind raters scored pre- and post-semester essays with inter-rater reliability of $\kappa = 0.86$. Additionally, weekly learner journals captured perceptions of cognitive engagement. Thematic coding followed Braun and Clarke’s six-phase approach. Quantitative data were analysed in SPSS using ANCOVA, controlling for baseline proficiency.

The adjusted post-test means revealed that experimental-group students achieved higher overall HOTS scores ($M = 22.3$, $SD = 2.1$) than control peers ($M = 18.7$, $SD = 2.8$), yielding a medium effect size ($\eta^2 = 0.21$). Dimension-level analysis indicated the greatest gains in evidence evaluation and creative integration, aligning with earlier findings on human–AI collaborative writing. Journal analysis uncovered three dominant themes. First, students reported a “dialogic push,” noting that AI prompts forced them to articulate warrants for their claims rather than accept surface-level paraphrases. Second, learners described heightened confidence to experiment with unfamiliar disciplinary vocabulary because immediate AI feedback reduced fear of lexical errors. Finally, several respondents articulated a nuanced awareness of authorship, deliberately negotiating which AI suggestions to accept, modify or reject.

The statistically significant improvement in HOTS corroborates arguments that generative AI, when embedded within scaffolding that foregrounds metacognition, can elevate cognitive complexity in L2 writing tasks. This outcome resonates with global research calling for AI-resistant yet AI-enhanced assessments that privilege reasoning over rote output. Beyond quantifiable gains, qualitative insights highlight the role of reflective monitoring—students’ capacity to interrogate AI output emerges as a vital literacy in an era where content generation is no longer the sole domain of humans. Contrary to critiques that AI tools homogenise discourse, the present study observed increased rhetorical originality as learners appropriated or contested AI-generated ideas. Nevertheless, sustainability hinges on pedagogical conditions. Unstructured exposure risks cognitive offloading; thus, teachers must design purposeful prompts aligned with curricular goals and transparently address ethical dimensions of assisted writing.

CONCLUSION

The evidence indicates that strategically-orchestrated collaboration with large language models can recalibrate ESL writing from a predominately form-focused exercise to a cognitively demanding enterprise that nurtures analysis, evaluation and synthesis. Students exposed to AI scaffolding not only outperformed peers on a validated HOTS rubric but also demonstrated metacognitive vigilance in deciding when to appropriate, modify or reject machine-generated text. These outcomes challenge the binary discourse that positions AI either as pedagogical panacea or existential threat, suggesting instead that learning gains hinge on the design of dialogic prompts and the presence of an informed instructor who can re-channel algorithmic affordances toward epistemic

goals.

Several limitations temper the generalisability of the findings: the sample was confined to business majors at a single institution; writing genres were limited to academic expository forms; and the study relied on short-term measures of cognitive growth. Future research should pursue longitudinal multicentre trials, explore creative genres such as digital storytelling and investigate automated analytics capable of issuing real-time HOTS diagnostics without compromising data privacy. Pedagogically, the study underscores the need for explicit instruction in AI literacy, ethical citation of machine assistance and assessment designs that privilege reasoning over surface features. By weaving reflective human judgment into every phase of AI-mediated writing, educators can cultivate the higher-order competencies that underpin lifelong learning in a knowledge economy increasingly shaped by generative technologies.

REFERENCES

- Wu Y., Li S., Chen H. Effects of ChatGPT on students' learning performance // Humanities & Social Sciences Communications. – 2024. – Vol. 11, № 5. – P. 1–14.
- Lu K., Zhang J., Gao M. ChatGPT-assisted teacher training and higher-order thinking outcomes // International Journal of Educational Technology in Higher Education. – 2024. – Vol. 21, № 2. – P. 45–63.
- Khan R., Ahmad N. Impact of ChatGPT as a formative feedback tool on ESL academic writing skills // Smart Learning Environments. – 2023. – Vol. 10, № 31. – P. 1–18.
- Oregon State University E-Campus. Bloom's taxonomy revisited – artificial intelligence tools: faculty guide. – 2024. – Mode of access: <https://ecampus.oregonstate.edu>. – Accessed: 05.06.2025.
- Jackson B. Higher-order prompting: applying Bloom's revised taxonomy to the age of AI // Scholarship of Teaching and Learning in the East. – 2025. – Vol. 4, № 1. – P. 23–38.
- Chen Q., Lee S. Human–AI collaborative writing strategy effects on EFL argument essays // Computer Assisted Language Learning. – 2025. – Vol. 38, № 6. – P. 589–607.
- Rahman J., Lim W. Meta-analysis of ChatGPT in L2 writing instruction // International Review of Research in Open and Distributed Learning. – 2024. – Vol. 25, № 3. – P. 120–139.
- Smith T., O'Connor D. Triad of artificial intelligence, Bloom's taxonomy and critical assessment literacy // Education and Information Technologies. – 2025. – Vol. P. 4115–4134.
- Pérez L., Santos V. AI-resistant assessments in higher education: practical insights from the Global South // Frontiers in Education. – 2024. – Vol. 9. – Article 1499495. – P. 1–17.
- Ali M., Zhang Y. Evaluating ChatGPT in EFL writing assessment through G-theory // Humanities & Social Sciences Communications. – 2024. – Vol. 11. – P. 1–13.
- Brown E. Exploring the impact of AI on critical thinking development in ESL settings. – Rochester, NY: SSRN, 2025. – 28 p.
- Bloom B. S. Taxonomy of educational objectives. Handbook I: Cognitive domain. – New York: David McKay, 1956. – 207 p.