

EFFECTIVENESS OF GENERATIVE AI WRITING ASSISTANTS ON EFL ACADEMIC WRITING IN HIGHER EDUCATION

Ikramova Aziza Aminovna

Associate professor of Bukhara University of Innovation.

<https://doi.org/10.5281/zenodo.17368155>

Abstract. *Generative AI writing assistants are increasingly present in higher education and promise to augment English as a Foreign Language (EFL) academic writing by providing rapid scaffolding across idea generation, organization, language accuracy, and cohesion. This article synthesizes cognitive and sociocognitive theories of writing with emerging evidence on human–AI collaboration to articulate a practice-ready model in which AI acts as a dynamic, feedback-rich partner during prewriting, drafting, revising, and reflecting. A quasi-experimental evaluation blueprint is outlined, emphasizing rubric-anchored ratings, automated discourse indices, process analytics, and learner-reported self-regulation and ethical use.*

Keywords: *generative AI, EFL academic writing, higher education, writing assessment, cohesion, feedback, revision, academic integrity, human–AI collaboration, instructional design.*

ЭФФЕКТИВНОСТЬ ГЕНЕРАТИВНЫХ ПОМОЩНИКОВ ПО НАПИСАНИЮ ТЕКСТОВ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В АКАДЕМИЧЕСКОМ ПИСЬМЕ EFL В ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЯХ

Аннотация. *Генеративные помощники письма в высшем образовании усиливают академическое письмо на английском как иностранном, обеспечивая быстрое «скаффолдирование» идей, структуры, языковой точности и связности. В статье объединяются когнитивные и социокогнитивные модели письма с данными о сотрудничестве человека и ИИ; ИИ рассматривается как динамический источник обратной связи на этапах преднаписания, черновика, редактирования и рефлексии. Предлагается квазиэкспериментальный план оценки, сочетающий рубрикативные рейтинги, автоматические индексы дискурса, аналитику процесса и самоотчёты студентов об авторегуляции и этическом использовании.*

Ключевые слова: *генеративный ИИ, академическое письмо EFL, высшее образование, оценивание письма, связность, обратная связь, редактирование, академическая честность, сотрудничество человек–ИИ, дизайн обучения.*

The accelerating diffusion of generative artificial intelligence within universities compels a re-examination of how English as a Foreign Language academic writing is cultivated, supported, and judged. A central premise of contemporary writing pedagogy is that learners improve when they can iterate rapidly between planning, drafting, and revising while receiving targeted feedback calibrated to genre expectations and disciplinary registers. Generative models now provide an always-available interlocutor that can stimulate idea generation, surface organizational patterns, propose cohesion devices, and render language-level recommendations in near real time. The mere availability of such assistance, however, does not guarantee learning: gains are contingent on how the tools are embedded within instructional designs that preserve authorship, require justification of edits, and make the revision process visible for assessment.

To move beyond intuition and polemics, the present study integrates theoretical, methodological, and practical strands into a coherent account of effectiveness, asking not only whether generative systems enhance EFL academic writing but also which dimensions improve, under what patterns of use, and with which trade-offs for learner agency and integrity.

A theoretical synthesis begins with cognitive process views of writing that treat composition as recursive planning, translating, and reviewing under working-memory constraints. In EFL settings, limited linguistic resources magnify these constraints, making it difficult to hold rhetorical purpose, content development, and sentence-level realization in mind simultaneously. Sociocognitive perspectives extend this view by situating writing within disciplinary discourse communities where genre moves and citation practices signal membership. Generative AI complicates yet potentially relieves these burdens: prompt-driven scaffolds can externalize planning, highlight genre-conventional structures such as problem–gap–claim sequences, and model stance-taking language that is otherwise acquired slowly through exposure. At the same time, the facility with which models produce fluent prose can mask weaknesses in argument development or evidence integration, and their probabilistic nature risks hallucinated citations if guardrails are absent. The conceptual solution advanced here is scaffolded co-authoring, an arrangement in which AI is neither banned nor allowed to ghostwrite but is constrained to roles that foster metacognition and control over discourse: eliciting research questions during prewriting, suggesting alternatives for topic sentences and transitions during revision, and providing language-level options that the learner must adjudicate with reasons.

Building a credible evidence base requires designs that can disentangle the influence of access from the influence of structure. A quasi-experimental arrangement across course sections, with structured AI use, unstructured AI use, and a control condition without AI, permits estimation of effect sizes on targeted outcomes while maintaining ecological validity. Structured use includes prompt templates tied to analytic rubrics, mandatory AI-use statements that attribute contributions, and revision memos in which learners justify acceptance or rejection of suggestions. Unstructured use offers access with integrity expectations but without process scaffolds, modeling the de facto practices that arise when policies are permissive. The control condition protects against confounds by providing equivalent human feedback opportunities, preventing a simple feedback-volume explanation for any observed gains. Outcome measures triangulate human ratings with automated indices of cohesion, lexical diversity, and syntactic complexity while collecting process analytics such as prompt logs and edit histories that reveal how learners navigate suggestions. Learner-reported measures of self-efficacy, cognitive load, and ethical attitudes complement these traces and can illuminate mechanisms of change.

Anticipated results follow from the interaction of scaffolding and learner agency. Under structured use, one expects moderate improvements in organization and cohesion because templates draw attention to claim–evidence–warrant alignment and paragraph unity while AI augments the stock of connectives and reference chains available to the writer. Language accuracy and complexity should show small-to-moderate gains, especially where models propose alternatives that are then filtered by the learner for register and precision. Under unstructured use, heterogeneous patterns are likely: some learners will channel models into surface-level

copyediting that improves fluency without strengthening argumentation, while others may overtrust generated content and thereby import weaknesses in logic or sourcing. In both AI conditions, transparent attribution and citation audits are indispensable to counter hallucinations and to sustain academic values; without such guardrails, any improvements in form could be offset by degradations in source fidelity or authorship clarity.

Institutional adoption must therefore be coupled with policy and pedagogy that transform AI from a black box into an object of inquiry. Course sequences can foreground the craft of prompting by having students construct genre-aware requests—for example, directing the model to produce three competing thesis statements that incorporate a specified counterclaim and that vary in scope. Subsequent activities require students to diagnose cohesion breaks, soliciting the model for multiple transition options and then defending choices with respect to rhetorical purpose. Such practices both harness the model's generative breadth and cultivate metacognitive regulation. Instructors can require before–after excerpts, with margin rationales that tie revisions to rubric criteria. These artifacts anchor assessment in visible learning processes rather than solely in polished outcomes. At the program level, alignment with learning outcomes ensures that AI use advances, rather than distracts from, competencies such as argument development, evidence synthesis, and disciplinary register control.

Ethical considerations are not exogenous to effectiveness; they shape it. An AI-use statement appended to each submission can document where assistance occurred, with tags for brainstorming, outlining, cohesion, lexis, and mechanics. Prohibitions on fabricated citations and requirements for human-verified sources mitigate risks while teaching students to treat generated text as conjecture to be checked rather than authority to be accepted. Data-privacy guidance must delineate what content may be shared with tools and when institutionally hosted instances are required. Assessment practices should incorporate integrity checks that are formative rather than merely punitive, inviting students to reflect on where AI nudged them toward clearer reasoning or where it tempted them to shortcut synthesis. When policy, pedagogy, and tooling cohere, effectiveness is not a narrow score lift but a broader maturation of writers who can appropriate AI outputs critically and transparently.

Measurement strategies deserve elaboration because they determine what counts as improvement. Analytic rubrics that partition content, organization, language, cohesion, and source use enable targeted feedback and analysis of differential impacts. Automated indices drawn from discourse processing can quantify referential cohesion and connective density, supporting or challenging human judgments. Yet automated signals are proxies; alignment with genre expectations remains a human interpretive task. Process data add a third line of evidence: prompt sequences reveal whether students progress from vague, outcome-agnostic requests to specific, criterion-linked ones; edit logs disclose whether AI suggestions are adopted wholesale or transformed through paraphrase and synthesis. Together, these layers permit a nuanced portrait of effectiveness that includes not only where the text ended but also how the writer got there.

Equity and access shape both opportunity and risk. Learners with lower initial proficiency may experience the largest subjective relief from AI-enabled fluency, but without scaffolds they may also cede too much control to generated text, suppressing voice and transfer.

Conversely, higher proficiency writers might leverage models to explore structural alternatives and rhetorical stance, gaining in flexibility. Structured pedagogy can narrow gaps by teaching all students to interrogate outputs, to cross-check sources, and to map suggestions onto rubric categories. Faculty development is necessary to normalize these practices; short workshops that model prompt design, show examples of effective revision memos, and demonstrate citation audits can shift instructor confidence from skepticism or uncritical enthusiasm to informed guidance.

The classroom enactment of these principles can proceed within standard timelines. A two-week cycle begins with instruction on argument structure and genre moves, followed by a prompting workshop where learners craft claims and counterclaims and anticipate objections.

Initial drafts, produced with minimal generative assistance beyond brainstorming, establish a baseline of authorial intent. A dual-feedback phase then combines AI-based diagnosis with human peer review keyed to a shared checklist, minimizing substitution effects while enhancing total feedback. Revision planning memoranda capture intent, and instructor conferences target the highest leverage issues surfaced by the combined feedback. Finalization includes a citation audit and an AI-use statement. This rhythm neither abdicates authorship to tools nor burdens the course with extraneous complexity; it reorders time toward the locus where learning happens—deliberate practice in revising for argument strength and cohesion.

Limitations of the approach should be acknowledged to prevent overgeneralization.

Tool variability means that results with one system may not replicate with another, so institutions should validate locally and update practices as models evolve. Automated text metrics are imperfect mirrors of rhetorical quality and must be interpreted alongside human readings. Differences in digital literacy and access can widen outcome variance unless scaffolding is equitable and explicit. Disciplinary diversity further complicates matters: argument moves in engineering differ from those in sociology, and prompt templates should be tuned accordingly. Finally, persistence of gains beyond a single course is an empirical question; longitudinal studies are needed to see whether metacognitive habits and rhetorical control transfer across genres and contexts.

Bringing these strands together clarifies what effectiveness means in the specific context of EFL academic writing in higher education. It is not merely a rise in grammatical accuracy or reduced error density, though these are welcome. It is the demonstrable capacity of learners to orchestrate claims, evidence, and warrants into coherent wholes; to control cohesion through purposeful deployment of reference, substitution, ellipsis, and connective signaling; to calibrate stance and hedging to disciplinary norms; and to do so through processes that render decision-making inspectable. Generative AI can catalyze this capacity when its role is constrained to scaffolding, when students must defend their choices, and when institutions align policy, pedagogy, and assessment around transparency and growth. In such conditions, the technology becomes a means to cultivate not just more fluent texts but more reflective writers who understand how to shape knowledge claims responsibly.

In conclusion, generative AI writing assistants can realistically enhance EFL academic writing when integrated as structured collaborators within a transparent pedagogy that preserves authorship and enforces integrity.

The scaffolded co-authoring approach, supported by mixed-method evaluation that triangulates human ratings, automated discourse indices, and process analytics, enables programs to document gains in organization, cohesion, and rhetorical control without relinquishing academic values. Rather than entrenching a binary between prohibition and permissiveness, institutions can adopt evidence-seeking practices that transform AI from a source of anxiety into an engine of deliberate practice. With clear attribution norms, citation fidelity checks, privacy safeguards, and assessment of process alongside product, higher education can realize the promise of these tools: not to write for students, but to help them learn to write with greater clarity, coherence, and confidence in the disciplines they are entering.

References

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of FAccT*.
2. Bitchener, J., & Knoch, U. (2010). Raising the Linguistic Accuracy Level of Advanced L2 Writers with Written Corrective Feedback. *Journal of Second Language Writing*, 19(4), 207–217.
3. Ferris, D. (1999). The Case for Grammar Correction in L2 Writing Classes. *Journal of Second Language Writing*, 8(1), 1–11.
4. Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365–387.
5. Hyland, K., & Hyland, F. (Eds.). (2006). *Feedback in Second Language Writing*. Cambridge University Press.
6. Kellogg, R. T. (1996). A Model of Working Memory in Writing. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing* (pp. 57–71). Lawrence Erlbaum.
7. McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press.
8. Truscott, J. (1996). The Case Against Grammar Correction in L2 Writing Classes. *Language Learning*, 46(2), 327–369.
9. UNESCO. (2023). *Guidance for Generative AI in Education and Research*. UNESCO.
10. Ziegler, N. (2016). Taking Technology to Task: Technology-Mediated TBLT and Second Language Learning. *Annual Review of Applied Linguistics*, 36, 136–163.