THE IMPACT OF GENERATIVE AI TOOLS (E.G., CHATGPT) ON DEVELOPING AUTONOMOUS WRITING SKILLS IN EFL CLASSROOMS

Adashaliyeva Dilshoda
Teacher, Norin District Technical School No. 1
Namangan Region,
https://doi.org/10.5281/zenodo.21022849

Annotation

This study examines how generative AI tools such as ChatGPT influence the development of autonomous writing skills in EFL classrooms, with a focus on balancing scaffolded support and independent performance. The aim is to identify patterns through which AI-mediated feedback affects planning, drafting, revising, and self-regulation in academic writing. A mixed-method design was applied, combining a quasi-experimental comparison of two instructional conditions with discourse-based analysis of student drafts and reflective logs. The novelty lies in operationalizing autonomy as a measurable composite of self-initiated revision density, metacognitive strategy reporting, and source-aware paraphrasing quality under controlled classroom constraints. Findings indicate that guided AI use can accelerate revision cycles and lexical diversification, but may also reduce originality indices when prompting practices are unstructured. The article contributes a pedagogical model for responsible AI integration and proposes assessment indicators that distinguish assisted performance from internalized writing competence in EFL contexts.

Key words: generative AI; autonomous writing; EFL classrooms; revision behavior; metacognitive self-regulation; academic integrity; feedback literacy

Introduction

Generative AI tools in language education can be defined as probabilistic text-generation systems that produce fluent discourse in response to prompts, and their rapid uptake has transformed how EFL learners approach academic writing tasks. The mechanism of influence begins with the redistribution of cognitive effort: when an AI model can propose outlines, topic sentences, or reformulations, learners may shift from generating content to evaluating and selecting among options, which can either strengthen monitoring or weaken authorship depending on instructional framing. For example, a student who asks a tool to refine cohesion devices may learn patterns of discourse markers, while a student who requests a full essay may bypass the struggle that typically consolidates genre knowledge. Classroom surveys across higher education have reported frequent AI trial use, with figures commonly ranging from roughly one third to over half of students depending on policy strictness and access, indicating that the phenomenon is not marginal but structurally present in writing ecologies. From a scientific perspective, the central question is not whether AI "helps" or "harms" but how it modulates self-regulation, strategy choice, and the internalization of genre norms under conditions of partial automation.

Autonomous writing skills can be defined as the capacity to plan, draft, revise, and edit texts through self-directed goal setting and monitoring, using resources strategically while maintaining responsibility for meaning, voice, and evidence. The underlying mechanism typically involves metacognitive regulation: writers evaluate task demands, mobilize linguistic

and rhetorical knowledge, and make iterative revisions informed by feedback, all while preserving alignment with intended argumentation. In EFL classrooms, a common example of autonomy is a learner independently diagnosing overuse of simple sentence patterns and deliberately introducing subordination and stance markers to match academic conventions. Empirical writing research has long associated autonomy-related strategies with improved outcomes; for instance, studies of revision behavior often find that higher-performing writers produce more substantive meaning-level revisions than surface corrections, with differences observable in revision counts and distribution across drafts [1]. Scientifically, autonomy is best treated as a multidimensional construct combining behavioral indicators, such as revision density, and cognitive indicators, such as explicit strategy reporting, rather than as a vague attitude toward independence.

Materials and Methods

The study employed a mixed-method approach that can be defined as the integration of quantitative and qualitative evidence within a single inferential framework to capture both magnitude and mechanism of learning effects. The mechanism of this design is triangulation: numerical comparisons reveal whether differences exist across conditions, while discourse analyses explain how those differences emerge through writing moves, feedback uptake, and strategic choices. For example, improvement in coherence scores may be traced to increased use of framing sentences and explicit transitions, which can be identified through coding of drafts rather than inferred from grades alone. The dataset consisted of iterative writing artifacts collected across a four-week unit, including outlines, first drafts, revised drafts, and reflection logs; in total, multiple drafts per participant yielded a corpus large enough to compute revision density and lexical diversity indices with stable estimates. Scientifically, mixed-method designs are recommended when educational interventions involve complex behaviors and contextual factors that cannot be reduced to test scores without losing explanatory power [4].

Participants were EFL undergraduate students enrolled in an academic writing course, and they were assigned to one of two instructional conditions: a structured-AI condition and an unstructured-AI condition. The structured condition can be defined as guided use under explicit prompting templates and process constraints, while the unstructured condition allowed open-ended tool use with minimal guidance, representing common real-world adoption. The mechanism differentiating the conditions was pedagogical control: in the structured condition, learners were required to document prompts, justify acceptance or rejection of AI suggestions, and perform at least one self-initiated revision pass without AI, whereas the unstructured condition permitted AI-assisted rewriting without obligatory rationales. For example, structured prompts asked for critique of argument logic and identification of cohesion gaps, while unstructured prompts often requested “make it more academic,” leading to opaque transformations. Numerically, both groups completed the same number of writing tasks and received equal instructional time, which supports internal validity by reducing exposure confounds. Scientifically, this operationalization aligns with the notion that tools become educationally meaningful only through mediated activity systems rather than mere availability [2].

Results

Autonomy outcomes can be defined as changes in learners' self-directed control over writing processes, operationalized here through the composite autonomy index and its component indicators across the unit. The mechanism observed in the structured-AI condition was an increase in self-initiated meaning-level revisions, suggesting that guided interaction with AI prompted learners to treat feedback as diagnostic input rather than as a replacement for authoring. For example, students often used AI outputs as a checklist to identify missing warrants or unclear topic progression, then rewrote paragraphs manually to fit their intended claims. Numerically, the structured condition showed a higher average growth in the autonomy index across the unit than the unstructured condition, with the largest contribution coming from meaning-level revision density and strategy specificity in reflections; additionally, the proportion of self-initiated revisions relative to total revisions was higher in the structured group. Scientifically, these patterns indicate that when scaffolding emphasizes evaluation and justification, AI can function as a metacognitive catalyst rather than a shortcut generator.

Writing quality outcomes can be defined as improvements in coherence, lexical choice, and argument structure as evaluated by rubric scores and text analytics. The mechanism in both conditions included enhanced surface correctness and more varied academic phrasing, reflecting AI's strength in providing grammatical and stylistic alternatives that learners can adopt. For example, many revised drafts displayed more consistent tense usage and clearer nominalizations, while introductions more frequently contained explicit thesis statements and signposting phrases. Quantitatively, both groups improved in rubric-based language accuracy and local cohesion, but the structured group demonstrated more consistent gains in global organization and argumentation, whereas the unstructured group's gains were concentrated in sentence-level polish. Scientifically, this distribution suggests that AI readily supports micro-level editing, but macro-level rhetorical competence depends on instructional constraints that force learners to articulate and test their own communicative intentions.

Integrity-related outcomes can be defined as observable behaviors that indicate source awareness, originality maintenance, and appropriate attribution practices under AI assistance. The mechanism distinguishing conditions was the presence or absence of required documentation and reflective justification, which directly influenced whether AI contributions were integrated transparently or absorbed invisibly. For example, in the unstructured condition, some drafts showed abrupt shifts in register and unusually dense academic collocations without corresponding explanations in logs, whereas structured-condition drafts more often contained explicit notes about which AI suggestions were accepted and why. Numerically, the unstructured condition exhibited lower originality indicators in internal similarity checks and a higher rate of flagged paraphrasing weaknesses, while the structured condition maintained more stable authorial voice measures and better paraphrasing scores due to explicit training in source-aware rewriting. Scientifically, these findings support the claim that integrity is not merely a policy issue but a design variable: process requirements can alter how learners appropriate AI output and whether they internalize ethical writing norms [3; 5].

To clarify the quantitative patterns, Table 1 summarizes key indicators by condition as aggregated tendencies rather than individual-level disclosure, reflecting the study's emphasis on pedagogical inference. The mechanism of table-based comparison is normalization across tasks, enabling interpretable contrasts without overstating precision beyond the study's classroom scale. For example, reporting revision density per 100 words controls for differences

in draft length, and reporting proportions of meaning-level revisions distinguishes substantive development from surface cleaning. Numerically, the structured condition shows higher autonomy-index change and higher meaning-level revision proportions, while the unstructured condition shows comparable gains in language accuracy but weaker paraphrasing quality and lower documented strategy specificity. Scientifically, such aligned indicator profiles strengthen construct validity, because multiple measures converge on the same interpretation rather than relying on a single score.

A linguistic micro-pattern emerged in the chemistry of text production, in which AI-assisted revisions often increased nominal density and reduced clause-level explicitness, which can be defined as a shift toward compact academic style. The mechanism involves AI's preference for compressed informational packaging through nominalizations and prepositional phrases, which may raise perceived formality but can also obscure logical relations for developing writers. For example, a learner's clause-based explanation "because teachers give feedback, students revise" might become "feedback provision facilitates revision," which sounds academic but may hide causal reasoning if the learner cannot unpack it. Numerically, drafts in both conditions displayed increased average word length and a higher proportion of abstract nouns, but only the structured condition maintained or improved clarity scores in teacher ratings, implying that compression alone is not a proxy for communicative effectiveness. Scientifically, this pattern resonates with register studies showing that academic style requires controlled complexity, not mere densification, and that learners need explicit instruction to manage the trade-off between concision and transparency [1; 4].

Discussion

The findings can be interpreted as evidence that generative AI's impact on autonomous writing depends primarily on pedagogical mediation rather than tool capability, a claim that can be defined as the primacy of instructional design in shaping learning outcomes. The mechanism is that structured AI use externalizes expert-like questioning, such as "Is my claim supported?" or "Does this paragraph advance the thesis?", and then requires learners to internalize those questions through justification and manual revision, which mirrors classic scaffolding-to-independence trajectories. For example, when students are required to reject some AI suggestions and explain why, they practice evaluative judgment, a core component of autonomy, rather than treating AI as an authority. Numerically, the higher proportion of meaning-level revisions in the structured condition aligns with revision research that links substantive rewriting to deeper learning, while the unstructured condition's surface-focused gains align with editing benefits that do not necessarily translate into independent composing competence [1; 2]. Scientifically, this supports a socio-cognitive view in which tools mediate activity but do not determine learning unless tasks and assessment compel reflective uptake.

Comparative analysis with existing scholarship indicates that the results converge with prior concerns about overreliance and converge with emerging evidence about AI as a feedback amplifier when guided properly. The mechanism of convergence is that both earlier feedback-literacy research and recent AI-in-writing discussions emphasize learners' capacity to interpret and act on feedback rather than merely receiving it. For example, studies of automated writing evaluation have shown that without instruction, students may chase surface corrections and ignore rhetorical issues, whereas structured engagement can foster planning and revision strategies; generative AI appears to intensify this dynamic because it can propose full rewrites,

increasing the temptation to outsource authorship [3; 4]. Numerically, the observed pattern of comparable language-accuracy gains across conditions alongside divergent autonomy gains fits the idea that AI is particularly efficient at local error correction but neutral or negative for higher-order development unless constrained. Scientifically, the implication is that EFL pedagogy should treat AI not as a substitute tutor but as a controlled stimulus for metacognitive practice, aligning with integrity frameworks that stress transparency and process documentation [5].

Conclusion

The study demonstrates that generative AI tools can either support or undermine autonomous writing development in EFL classrooms depending on how their use is structured within pedagogy and assessment. Autonomy increased when AI interaction was constrained by prompting templates, transparency requirements, and reflection tasks that required learners to evaluate suggestions and perform self-initiated meaning-level revisions. Writing quality improved in both structured and unstructured conditions at the level of grammatical accuracy and local cohesion, but higher-order organization and argumentation benefited more reliably from guided integration. Integrity-related behaviors, particularly source-aware paraphrasing and stable authorial voice, were stronger when documentation and justification were mandatory, indicating that ethical writing practices can be designed into the learning process rather than enforced only through policing. Overall, the findings support a process-oriented model in which AI functions as a diagnostic and exploratory tool while learners retain responsibility for claims, evidence, and rhetorical decisions. Sustainable implementation therefore requires assessment criteria that value revision reasoning, prompt literacy, and traceable development alongside final text quality.

References:

1. Hyland K. Second Language Writing. Cambridge, Cambridge University Press, 2019. 320 p.
2. Warschauer M. Learning in the Cloud: How (and Why) to Transform Schools with Digital Media. New York, Teachers College Press, 2011. 176 p.
3. Zawacki-Richter O., Marín V. I., Bond M., Gouverneur F. Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education. Berlin, Springer, 2019. P. 1–27.
4. Creswell J. W., Creswell J. D. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks, SAGE Publications, 2018. 304 p.
5. Flower L., Hayes J. R. A cognitive process theory of writing. College Composition and Communication. Urbana, National Council of Teachers of English, 1981. P. 365–387.
6. Исаков Н. М. Методика обучения письменной речи на иностранном языке в вузе. Москва, Просвещение, 2016. 240 с.
7. Karimov A. Academic writing in EFL contexts: assessment criteria and feedback practices. Tashkent, Uzbekistan National University Press, 2020. 180 p.