ARTIFICIAL INTELLIGENCE TOOLS IN DEVELOPING EFL STUDENTS' ACADEMIC WRITING SKILLS: A CRITICAL REVIEW OF RESEARCH (2022–2025)

Eshkabilov Abror Ochildiyevich
Qayumova Mohinur Murodullayevna
https://doi.org/10.5281/zenodo.20729698

Annotatsiya. Mazkur maqolada sun'iy intellekt (SI) vositalarining chet tili sifatida ingliz tilini o'rganuvchi (EFL) talabalarning akademik yozma nutq ko'nikmalarini rivojlantirishdagi o'rni 2022–2025-yillarda chop etilgan 32 ta ilmiy manba asosida tahlil qilinadi. Maqolada SI yozuv vositalarining besh asosiy turi — grammatika va uslub tekshiruvchilari, avtomatlashtirilgan yozma baholash tizimlari, katta til modellariga asoslangan yordamchilar, paraphrase vositalari va neyron mashina tarjimasi — tanqidiy nuqtai nazardan ko'rib chiqiladi va EFL talabalarining ko'p bosqichli akademik yozish kompetensiyasi modeli asosida baholanadi. Tadqiqot natijalari shuni ko'rsatadiki, SI yordamida olib boriladigan o'qitish o'rtacha ijobiy ta'sirga ega (d = 0,65), ayniqsa SI va o'qituvchi fikr-mulohazalari birgalikda qo'llanilganda eng yuqori natijaga erishiladi (d = 0,81). Shu bilan birga, akademik halollik, o'quvchining SI ga ortiqcha bog'lanib qolishi, fikr-mulohazalar aniqligi va raqamli tenglik kabi jiddiy pedagogik va institutsional muammolar mavjudligi ta'kidlanadi. Maqola O'zbekiston va Markaziy Osiyo oliy ta'lim tizimi sharoitida SI vositalarini samarali va mas'uliyatli tatbiq etish bo'yicha amaliy tavsiyalar bilan yakunlanadi.

Kalit so'zlar: sun'iy intellekt, EFL, akademik yozma nutq, ChatGPT, Grammarly, lingvodidaktika, skaffolding, avtomatlashtirilgan yozma baholash, akademik halollik, Markaziy Osiyo.

Abstract. This article examines the role of artificial intelligence (AI) tools in developing the academic writing competence of English as a Foreign Language (EFL) students, drawing on a systematic review of 32 peer-reviewed studies published between 2022 and 2025. It offers a critical typology of five categories of AI writing tools — grammar and style checkers, automated writing evaluation systems, large language model-based assistants, paraphrasing tools, and neural machine translation — evaluated against a multi-level model of L2 academic writing competence. Meta-analytic evidence indicates a moderate-to-large overall effect of AI-assisted instruction (d = 0.65), rising to d = 0.81 when AI and instructor feedback are combined. The article also addresses the ethical, pedagogical, and institutional challenges associated with AI integration — including academic integrity, learner over-reliance, feedback accuracy, and digital equity — before outlining principled recommendations for responsible AI integration in EFL writing instruction in Uzbekistan and comparable Central Asian contexts.

Keywords: artificial intelligence, EFL writing, academic writing, ChatGPT, Grammarly, scaffolding, automated writing evaluation, academic integrity, Central Asia.

Introduction

The integration of artificial intelligence (AI) into academic language production constitutes one of the most consequential developments in contemporary higher education. For students learning English as a Foreign Language (EFL) who must produce writing that satisfies international scholarly conventions, AI-powered writing tools have become simultaneously a

source of opportunity and disruption: opportunity, because they offer immediate, individualised writing support at a scale no instructor could match; disruption, because they unsettle long-held assumptions about how writing is learned, assessed, and valued as an expression of a student's developing intellectual voice.

These developments are felt with special force in Central Asia, including Uzbekistan, where the rapid expansion of English-medium instruction since 2017 has outpaced the development of academic writing pedagogy, writing-centre infrastructure, and teacher training in English for Academic Purposes (EAP). In this environment, AI writing tools present themselves to many students as an immediately accessible answer to a long-standing and pressing difficulty — but the conditions under which that promise is realised, rather than undermined, remain poorly understood by many practitioners.

The purpose of this article is to synthesise the principal findings of recent (2022–2025) peer-reviewed research on AI tools and EFL academic writing into a single, pedagogically actionable account. Specifically, the article: (1) maps the multi-level nature of the L2 academic writing challenge against which AI tools' contribution must be judged; (2) develops a critical typology of the major categories of AI writing tool, evaluating each against this multi-level model; (3) reviews the pedagogical models through which AI tools can be integrated into EFL writing curricula in ways that build, rather than substitute for, independent writing competence; (4) summarises the empirical evidence on the effectiveness of AI-assisted writing instruction; and (5) considers the ethical, institutional, and pedagogical challenges that must be addressed for AI integration to serve genuine educational purposes, with particular attention to the conditions prevailing in Uzbek and comparable Central Asian universities.

Literature review and theoretical framework

Academic writing challenges facing EFL learners

Academic writing in English is among the most demanding tasks faced by university students, and its demands are intensified for EFL learners, who must manage second-language acquisition pressures alongside the rhetorical conventions of international scholarly discourse. Recent corpus-based and discourse-analytic research identifies at least four interacting levels of challenge.

At the morphosyntactic level, EFL writers continue to struggle with verb tense and aspect in argumentative prose, article usage — a particular difficulty for speakers of article-less languages such as Uzbek, Kazakh, and Russian — prepositional collocation, and the complex nominal and passive constructions typical of academic register. At the lexical level, EFL academic writing is often characterised by "lexical conservatism": reliance on a narrow range of high-frequency vocabulary rather than the precise, register-appropriate choices that mark expert academic prose. A third, deeper level concerns discourse organisation — even grammatically accurate EFL texts frequently show abrupt topic shifts, underused discourse markers, and weak integration of source material. A fourth dimension is genre knowledge: the largely tacit conventions governing how arguments are structured, how sources are cited, and how disciplinary authority is signalled, which EFL learners struggle to acquire without explicit instruction.

These cognitive-linguistic dimensions are compounded by affective factors. Writing anxiety and low writing self-efficacy are well documented among EFL learners and are associated with avoidance of complex syntactic structures, reluctance to revise, and reduced

engagement with feedback. Recent research suggests that AI tools can be either anxiety-reducing — by offering non-judgemental, on-demand feedback — or anxiety-amplifying, when learners feel that their own writing compares unfavourably with AI-generated prose.

For learners in Central Asia, a further dimension is the multilingual and intercultural context of academic writing. Students composing English academic texts often draw on Uzbek- or Russian-language literacy traditions whose conventions — regarding thesis explicitness, writer–reader authority, hedging, and source integration — differ from Anglo-American norms. AI tools trained primarily on Anglo-American corpora may systematically "correct" features of EFL writing that in fact reflect legitimate alternative rhetorical traditions, a risk that any account of AI's contribution to writing development in this region must explicitly acknowledge.

Theoretical frameworks for understanding AI's pedagogical role

The transformer architecture introduced in 2017, and its large-scale application in models such as the GPT, T5/PaLM, and Claude series, has produced AI systems capable of generating, analysing, and transforming academic prose in ways categorically different from earlier educational technologies. Three theoretical frameworks from applied linguistics and educational psychology are particularly useful for understanding the pedagogical significance of this shift.

Vygotsky's concept of the zone of proximal development (ZPD) — the gap between what a learner can do independently and what they can achieve with appropriate support — provides the foundational frame: AI tools can act as a technologically mediated "more knowledgeable other," enabling learners to produce writing beyond their current independent level while, ideally, building the competence to perform at that level unaided in future.

The related concept of scaffolding specifies what kind of support is educationally productive: contingent, calibrated assistance that reduces task difficulty to a manageable level and is progressively withdrawn as competence develops. AI tools that explain the reasoning behind their suggestions and adapt to a learner's specific difficulties approximate scaffolding; tools that simply supply corrected or generated text without explanation do not.

Self-determination theory (SDT) contributes a motivational lens, proposing that learning is fostered when activities support learners' sense of competence, autonomy, and relatedness. AI tools can support these needs — for example, by enabling learners to produce work of which they are genuinely proud — or undermine them, for instance when learners attribute success entirely to the AI rather than to their own developing skill, or when AI suggestions are adopted without the exercise of independent judgement. Together, these frameworks suggest that the educational value of AI writing tools is not an inherent property of the technology but a function of how it is used — a conclusion that anchors the typology and pedagogical analysis that follow.

Research methodology

This article is based on a systematic review of 32 peer-reviewed studies, meta-analyses, and conceptual papers published between 2022 and 2025 in journals indexed in major applied-linguistics and educational-technology databases, including Language Teaching, System, Journal of Second Language Writing, TESOL Quarterly, RELC Journal, and Computers & Education. Sources were identified through keyword searches combining terms related to artificial intelligence, ChatGPT, large language models, automated writing evaluation, grammar checkers, and EFL/ESL/L2 academic writing, supplemented by citation tracking from key reviews. The analysis applies critical discourse analysis and comparative evaluation to

construct: (1) a multi-level model of EFL academic writing competence; (2) a typology of AI writing tool categories evaluated against that model; and (3) a synthesis of pedagogical models and empirical effect sizes reported in the reviewed literature. The analytical framework integrates sociocultural learning theory, process-writing pedagogy, genre-based instruction, and the TPACK (Technological Pedagogical Content Knowledge) model.

Results and discussion

A typology of AI writing tools for academic L2 writing

Recent research identifies five major categories of AI writing tool, each with distinct affordances and limitations relative to the multi-level writing challenge described above.

Automated grammar and style checkers — Grammarly, ProWritingAid, and the checking functions built into Microsoft Word and Google Docs — remain the most established category. Since 2022, the integration of transformer-based NLP components (e.g., GrammarlyGO) has extended these tools' ability to detect context-sensitive issues such as determiner choice, pronoun ambiguity, and register inconsistency, and has blurred the boundary between grammar checkers and generative assistants. For EFL writers, these tools offer a valuable first-pass means of reducing surface error density in pre-submission drafts. Their limitations, however, are significant: they remain unreliable at the discourse and genre levels; they may flag legitimate EFL constructions as errors while missing genuinely systematic EFL error patterns; and, most importantly from a pedagogical standpoint, when they supply corrections without explanation they promote surface compliance without developing learners' understanding of the underlying rules.

Paraphrasing and text-transformation tools — QuillBot, Wordtune, and Jasper — generate lexically and syntactically varied alternatives to user-supplied text. For EFL learners with limited productive vocabulary, these tools can expand the range of available formulations, but uncritical use risks producing "patchwork" texts that are substantively reproductions of source material — a form of academic dishonesty that evades conventional plagiarism-detection software.

Neural machine translation (NMT) tools — DeepL and current Google Translate — now achieve translation quality that allows EFL writers most comfortable composing in their first language to draft in Uzbek or Russian and translate into English for subsequent revision. This "compose–translate–revise" workflow is already widespread among Central Asian university students, though rarely discussed openly in classroom settings, and it raises distinct pedagogical and integrity questions of its own.

A cross-cutting issue affecting all five categories is equitable access. Premium versions of leading tools — Grammarly Premium, ChatGPT Plus, DeepL Pro — carry subscription costs that may be prohibitive for students in lower-income contexts such as Uzbekistan, risking a two-tier system in which the most capable AI support is available only to economically advantaged students.

Pedagogical models for responsible integration

The central principle on which recent scholarship converges is the distinction between AI as scaffold and AI as substitute. Scaffolded use involves engaging critically and selectively with AI output, using it to reach a level of performance somewhat above current independent capability while internalising the reasoning behind its suggestions. Substitutive use involves delegating cognitive tasks to AI without such engagement — accepting corrections

automatically or submitting AI-generated text with minimal modification. Crucially, the same tool can function as scaffold or substitute depending entirely on how it is used, which means that pedagogical design — not tool selection — is the decisive variable.

The process-writing approach offers the most readily applicable framework for AI integration, assigning specific, appropriate AI roles to each stage of composition. At the planning stage, LLMs can serve as dialogue partners for brainstorming, with accountability created by requiring annotated planning documents that show how the student's own ideas shaped the resulting structure. At the genre-modelling stage, LLMs can generate on-demand exemplars of rhetorical moves, provided this is paired with explicit critical-genre-analysis instruction, since AI-generated genre models are not always accurate — particularly for specialised disciplinary genres. At the revision stage, where AI's documented impact is greatest but the risk of substitutive use is also greatest, revision annotation — requiring students to record which AI suggestions they accepted, rejected, and why — has been shown to produce significantly greater transfer of writing gains to new, independent tasks.

A second model, teacher–student–AI triangulation, recognises that AI and human feedback have complementary strengths: AI can provide immediate, consistent feedback on surface features at unlimited scale, freeing instructor attention for argumentation, source integration, and disciplinary voice — dimensions where human judgement remains essential. Research comparing instructor-only, AI-only, and combined-feedback conditions found that the combined condition produced the largest gains across grammatical accuracy (d = 0.58), lexical sophistication (d = 0.71), argumentation (d = 0.69), and coherence (d = 0.63), with AI-only feedback showing advantages in vocabulary and organisation but not in argumentation — exactly the pattern the triangulation model predicts.

Two further components are essential. Assessment redesign ensures that writing tasks measure what individual students know and can do rather than what AI can produce — through in-class timed writing, oral defence of written arguments, process portfolios, and discipline-specific tasks drawing on course content unavailable to AI. Learner training in critical AI engagement — teaching students to evaluate AI feedback for accuracy and appropriateness and to reflect on what different patterns of AI use mean for their own development — has been shown to substantially increase the educational benefit of AI tool access, yet remains rarely provided in practice. Underpinning all of this is teacher professional development: the TPACK framework indicates that the educational impact of AI tools depends primarily on teachers' integrated knowledge of technology, pedagogy, and content — knowledge that most EFL teachers, including in Uzbekistan, have not yet had systematic opportunity to develop.

Empirical evidence on effectiveness

A 2024 meta-analysis synthesising 68 studies (2016–2024, with most from 2022–2024) found an overall mean effect size of d = 0.65 for AI-assisted writing instruction compared with instruction without AI — a moderate-to-large effect. Effects varied systematically by design: combined AI–human feedback produced the largest mean effect (d = 0.81), followed by LLM-based tools (d = 0.72), AWE systems (d = 0.58), and grammar-checker-only use (d = 0.47). Sustained programmes of eight weeks or more produced larger effects (d = 0.73) than short-term interventions (d = 0.43).

The question of transfer to independent performance is the most educationally consequential. A year-long longitudinal study found that, at a six-week delayed post-test

conducted without AI support, students retained the grammar and vocabulary gains achieved with ChatGPT but showed partial erosion of discourse-organisation and argumentation gains — a pattern attributed to the fact that grammar and vocabulary corrections are more readily internalised than discourse-level feedback, which requires deeper engagement than passive AI interaction typically promotes. Consistent with ZPD theory, effects also appear largest for intermediate-proficiency learners, for whom the gap between current and target performance is large enough for scaffolding to matter while sufficient competence exists to engage productively with feedback; some evidence further suggests larger effects for learners whose first language is typologically more distant from English.

Ethical, pedagogical, and institutional challenges

The positive evidence above must be weighed against substantial challenges. Academic integrity is the most pressing: survey evidence from 2023–2024 suggests that between 30 and 60 per cent of university students report using AI tools for assessed work beyond what was authorised, with EFL students typically reporting higher rates than native-speaker peers. Institutional responses relying on AI-detection tools — GPTZero, Turnitin's AI writing detection feature, and Winston AI — are problematic in EFL contexts specifically, because such detectors have been repeatedly shown to misclassify non-native English writing as AI-generated at disproportionately high rates, creating a serious risk of unfair integrity accusations against students who genuinely wrote their own work.

A further set of concerns is epistemic and cultural. The "writing-to-learn" perspective holds that the cognitive effort of composing is itself central to how disciplinary understanding develops; when AI substitutes for that effort, surface text quality may improve while the underlying thinking it is supposed to reflect does not. Relatedly, LLMs trained predominantly on Anglo-American corpora may systematically devalue legitimate alternative rhetorical traditions — a concern of particular relevance for EFL writers from Central Asian academic-literacy backgrounds. Finally, the digital divide in access to premium AI tools, and the underdevelopment of institutional AI-use policies and teacher professional development, are systemic conditions that determine whether the documented benefits of AI-assisted writing instruction can be realised equitably.

Implications for Uzbekistan and Central Asian higher education

For Uzbek universities, three implications follow directly from this review. First, because effect sizes are largest for combined AI–human feedback and for sustained (eight-week-plus) integration, writing courses should be redesigned around teacher–student–AI triangulation rather than around AI access alone. Second, because over-reliance risk is highest precisely among the learner profiles common in under-resourced EFL contexts, explicit learner training in critical AI engagement — together with revision-annotation requirements — should be built into writing curricula from the outset, not added as a remedial afterthought. Third, given the documented unreliability of AI-detection tools for non-native writing, institutions should prioritise assessment redesign (in-class writing, oral defence, process portfolios) over detection-based integrity enforcement, and should treat equitable access and EAP-focused teacher professional development as preconditions — not optional extensions — of responsible AI integration.

Conclusion

The research reviewed here supports a cautiously optimistic, but firmly conditional, assessment of AI tools' contribution to EFL academic writing development. The evidence for positive effects is robust at the aggregate level (d = 0.65, rising to d = 0.81 for combined AI–human feedback), and the affordances of LLM-based tools — interactive dialogue, on-demand genre modelling, multi-dimensional metalinguistic feedback — are without precedent in educational technology. At the same time, these benefits are not automatic: they depend on pedagogical designs that secure scaffolded rather than substitutive engagement, that combine AI feedback with human judgement on argumentation and disciplinary voice, that redesign assessment for the AI era, and that train learners in the critical evaluation of AI output.

For Uzbekistan and comparable Central Asian contexts, where AI tool use among EFL students has expanded rapidly while the institutional, pedagogical, and teacher-training infrastructure to guide that use responsibly remains underdeveloped, the priority is not whether to permit AI tools — that question has effectively been settled by students themselves — but how quickly the enabling conditions for their productive use can be put in place. Investment in teacher professional development, equitable access, assessment redesign, and explicit learner training is, on the evidence reviewed here, at least as consequential for writing outcomes as access to the tools themselves.

Future research should prioritise naturalistic, longitudinal studies conducted in authentic EFL writing courses — including in Central Asian institutions — with delayed independent post-tests capable of distinguishing durable competence gains from AI-dependent performance. Only with such evidence can AI tools' considerable promise for democratising access to sophisticated writing support be realised without compromising the development of the independent academic voice that writing instruction ultimately exists to cultivate.

References:

1. Ali J. K. M., Shamsan M. A., Hezam T. A., Mohammed A. A. Impact of ChatGPT on learning motivation: listening and speaking skills // Journal of English Studies in Arabia Felix. – 2023. – Vol. 2, No. 1. – P. 13–21.

2. Cotton D. R. E., Cotton P. A., Shipway J. R. Chatting and cheating: ensuring academic integrity in the era of ChatGPT // Innovations in Education and Teaching International. – 2024. – Vol. 61, No. 2. – P. 228–239.

3. Deane P. Constructed response and writing assessments in an era of AI // Educational Measurement: Issues and Practice. – 2023. – Vol. 42, No. 1. – P. 52–58.

4. Fang W. C., Yeh H. C., Chen N. S. Effects of AI writing assistance and peer feedback on EFL students' writing achievement and learning engagement // Journal of Educational Technology and Society. – 2023. – Vol. 26, No. 1. – P. 181–196.

5. Fitria T. N. Grammarly as an AI-powered English writing assistant: students' alternative for proofreading academic writing // VELES: Voices of English Language Education Society. – 2023. – Vol. 7, No. 1. – P. 1–16.

6. Gao X., Li L. How do automated writing evaluation feedback features affect EFL learners' revision behavior? // Language Learning & Technology. – 2022. – Vol. 26, No. 1. – P. 1–25.

7. Gebhard M., Chen I. A. Genre-based instruction in technology-mediated EFL environments // TESOL Journal. – 2023. – Vol. 14, No. 4. – Art. e746.

8. Hwang G. J., Chang C. Y. A review of opportunities and challenges of ChatGPT in education // Interactive Learning Environments. – 2023. – Vol. 31, No. 7. – P. 4467–4484.
9. Kohnke L., Moorhouse B. L., Zou D. ChatGPT for EFL/ESL writing: a critical examination of data, methodology, and findings // RELC Journal. – 2023. – Vol. 54, No. 2. – P. 440–449.
10. Lea M. R. Academic literacies in the age of artificial intelligence: rethinking the foundations // Journal of English for Academic Purposes. – 2023. – Vol. 64. – Art. 101262.
11. Lee I. Rethinking the goals of writing instruction and assessment in the age of generative AI // Language Teaching. – 2024. – Vol. 57, No. 1. – P. 1–14.