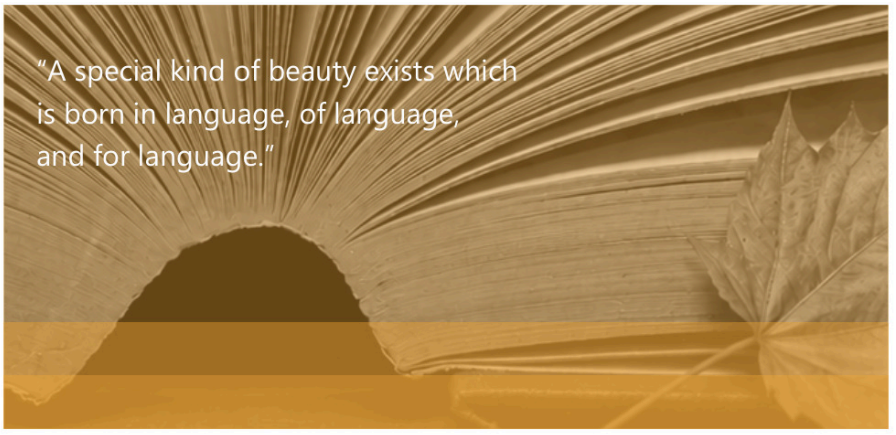
VOLUME 10, OCTOBER 2026
ISSN: 3060-4958







The Lingua Spectrum



UZBEKISTAN STATE WORLD
LANGUAGES UNIVERSITY

MORE INFO:

www.lingvospektr.uz

t.me/thelinguaspectrum

Nazarova N.B.	The Literary Representation of Women's Accomplishments and Charm in Jane Austen's Novels	319
Karimova S.S.	Connotative meaning and pragmatic interpretation in literary texts	325
Radjabova G.G.	Corpus evidence and the limits of intuition in the study of collocation	331
Adinaev O.J.	Onomastic, lexico-semantic and linguocultural foundations for the study of celestial body names	339
Muxitdinova X.S.	Sociocultural and diachronic characteristics of personal names in the Uzbek language	349
Amanov V.B.	Universal values as an object of linguocultural research in Spanish and Uzbek	358
PEDAGOGY		**378**
Kambarova D.I.	The methodology of developing oral speech competence of future foreign language teachers in a digital educational environment	379
Allayarov T.Ya.	Developing communicative competence of public health college students using the standardized patient method	389
Silanteva E.F.	Preliminary training for ESP teachers for effective integrated teaching	396
Omonova E.B., Musoyeva A.B.	Digital revision portfolios for developing English writing skills in Uzbek secondary schools	411
Paluanova Kh.D.	The relationship between pragmalinguistic analysis and linguodidactic design in preservice English teacher education	419
Adamboyeva N.Q.	Age-related development of phonological skills in children aged 7–10 and their role in reading competence	427
Alimardonov L.A.	Theoretical foundations and conceptual model of spiritual technologies in ensuring the information security of young people	435
Pazilova N.M., Abdulhakimova R.A.	Didactic potential and limitations of generative artificial intelligence in teaching Russian	443
Kadirova N.M., Tahirova Sh.Sh.	Stages of working with texts in Russian language classes at higher education institutions	450
Pulatova F.U.	Pragmadidactic features of teaching implicature and indirect speech acts in English language education	456
Sulaymonova M.A.	AI-Assisted Feedback in EFL Writing: Implications for Writing Accuracy and Learner Autonomy	464
Khodjieva Z.M.	Theoretical and Methodological Foundations and Practical Stages of Criterion-Referenced Assessment in Foreign Language Teaching at Higher Education Institutions	489
Musayeva F.M.	The role of specialized vocabulary in developing the lexical competence of oil and gas students	498

The Lingua Spectrum
Cogito, ergo sum

Jurnal Oliy Attestatsiya Komissiyasi tomonidan 2025-yil 4-yanvar kuni 366/5-sonli qaror bilan tavsiya etilgan milliy ilmiy nashrlar ro'yxatiga kiritilgan

AI-Assisted Feedback in EFL Writing: Implications for Writing Accuracy and Learner Autonomy

Sulaymonova Madina Amrullo qizi
sulaymonova1309@gmail.com
Graduated Bachelor student,
Uzbekistan State World Languages University

Annotation *This structured thematic literature review synthesizes peer-reviewed empirical research published between 2020 and 2026 on AI-assisted feedback in English as a Foreign Language (EFL) writing, examining its dual implications for writing accuracy and learner autonomy. Drawing on foundational theories of written corrective feedback, self-regulated learning, and learner autonomy, the review analyzes studies comparing AI-generated, teacher-provided, and hybrid feedback models. Findings indicate that AI feedback appears to improve surface-level accuracy in the reviewed studies, although its effectiveness remains inconsistent regarding higher-order concerns, including coherence, argumentation, and organization. Furthermore, its influence on learner autonomy depends largely on pedagogical design, learners' critical engagement, and teachers' guidance. The review argues that accuracy and autonomy are interdependent outcomes shaped by instructional integration rather than technological tools alone. It identifies critical research gaps, particularly the lack of longitudinal evidence and limited research involving lower-proficiency learners, and proposes pedagogical recommendations for human-centered AI implementation in EFL writing classrooms.*

Keywords *AI-assisted feedback, EFL writing, writing accuracy, learner autonomy, automated written corrective feedback, generative AI, ChatGPT, self-regulated learning*

Ingliz tilini xorijiy til sifatida o‘qitishda sun’iy intellekt yordamida fikr-mulohaza berish: yozma nutq aniqligi va o‘rganuvchilar mustaqilligiga ta’siri

Sulaymonova Madina Amrullo qizi
sulaymonova1309@gmail.com
Bakalavriat,
O‘zbekiston davlat jahon tillari universiteti
Bitiruvchisi

Annotatsiya *Ushbu tizimli tematik adabiyotlar sharhi 2020–2026-yillarda ingliz tilini xorijiy til sifatida o‘qitishda yozma nutqqa sun’iy intellekt yordamida beriladigan fikr-mulohazaga oid taqrizdan o‘tgan empirik tadqiqotlarni umumlashtiradi. Unda bu yondashuvning yozma nutq aniqligi va o‘rganuvchilar mustaqilligiga ta’siri ko‘rib chiqiladi. Yozma tuzatuvchi fikr-mulohaza, o‘zini o‘zi boshqarib o‘rganish hamda o‘rganuvchi mustaqilligi nazariyalariga tayangan holda, sharh sun’iy intellekt, o‘qituvchi va aralash fikr-mulohaza modellarini taqqoslovchi tadqiqotlarni tahlil qiladi. Natijalar sun’iy intellekt fikr-mulohazasi ko‘rib chiqilgan tadqiqotlarda yuzaki darajadagi aniqlikni oshirishi mumkinligini ko‘rsatadi, biroq matn izchilligi, dalillash va tuzilishi kabi murakkab jihatlardagi samarasi bir xil emas. Uning o‘rganuvchilar mustaqilligiga ta’siri pedagogik dizayn, o‘quvchilarning tanqidiy ishtiroki va o‘qituvchi yo‘l-yo‘rig‘iga bog‘liq. Sharh aniqlik va mustaqillik faqat texnologik vosita bilan emas, balki uning ta’limga qanday integratsiya qilinishi bilan belgilanadigan*

o'zaro bog'liq natijalar ekanini ta'kidlaydi. Shuningdek, uzoq muddatli dalillar yetishmasligi va til darajasi pastroq o'rganuvchilar ishtirokidagi tadqiqotlar cheklanganini aniqlab, ingliz tilida yozishni o'qitishda insonga yo'naltirilgan sun'iy intellektdan foydalanish bo'yicha pedagogik tavsiyalar beradi.

Kalit so'zlar *Sun'iy intellekt yordamida fikr-mulohaza berish, ingliz tilini xorijiy til sifatida o'qitish, yozma nutq aniqligi, o'rganuvchilar mustaqilligi, avtomatlashtirilgan yozma tuzatuvchi fikr-mulohaza, generativ sun'iy intellekt, ChatGPT, o'zini o'zi boshqarib o'rganish*

Обратная связь с использованием искусственного интеллекта в письменной речи на английском языке как иностранном: влияние на точность письменной речи и автономию обучающихся

Сулаймонова Мадина Амрулло кизи
sulaymonova1309@gmail.com
Выпускница бакалавриата,
Узбекский государственный университет мировых языков

Аннотация *В данном тематическом обзоре обобщаются рецензируемые эмпирические исследования 2020–2026 годов, посвящённые обратной связи с использованием искусственного интеллекта при обучении письму на английском языке как иностранном. Рассматриваются её последствия для точности письменной речи и автономии обучающихся. Опираясь на теории письменной корректирующей обратной связи, саморегулируемого обучения и автономии учащихся, обзор анализирует исследования, сравнивающие обратную связь ИИ, преподавателей и гибридные модели. Результаты показывают, что обратная связь ИИ, по-видимому, повышает точность на поверхностном уровне, однако её эффективность непоследовательна в отношении более сложных аспектов, включая связность, аргументацию и организацию текста. Её влияние на автономию обучающихся зависит от педагогического дизайна, критического участия учащихся и руководства преподавателей. В обзоре утверждается, что точность и автономия взаимосвязаны и определяются интеграцией ИИ в обучение, а не только инструментом. Выявлены пробелы: отсутствие долгосрочных данных и ограниченное количество исследований с учащимися низкого уровня владения языком. В завершение предлагаются педагогические рекомендации по человеко-ориентированному внедрению ИИ в обучение письму на английском языке как иностранном.*

Ключевые слова *Обратная связь с использованием искусственного интеллекта, письменная речь на английском языке как иностранном, точность письменной речи, автономия обучающихся, автоматизированная письменная корректирующая обратная связь, генеративный искусственный интеллект, ChatGPT, саморегулируемое обучение*

Introduction

Writing in a foreign language rarely improves without feedback. Yet in most EFL classrooms, the demands of response are considerable: a teacher with thirty students who each draft two essays per term is looking at sixty texts to correct, often on top of speaking, reading, and administrative duties. The result is a familiar compromise-either the teacher burns out, or the students receive thin, delayed comments that arrive too late to shape the next draft. Ferris (2006) documented this tension years ago, and despite digital tools, the fundamental scarcity of detailed, timely feedback has not disappeared. If anything, it has grown more visible as communicative and task-based approaches push students to produce more text, not less.

Automated tools promised relief as early as the 1990s, but the first generation of software-rule-based grammar checkers and early automated writing evaluation (AWE) systems like Criterion-offered little beyond surface-level error flagging. They could spot a missing article or a subject-verb mismatch, yet they said almost nothing about whether an argument held together or whether a paragraph actually supported its topic sentence. Students received a score and a list of mechanical fixes, not a conversation about their writing. The gap between what these tools measured and what good writing actually requires left many teachers skeptical, and rightly so.

What changed was the arrival of large language models. When ChatGPT appeared in late 2022-alongside emerging competitors such as Claude and Gemini-it brought something genuinely different: the capacity to generate extended, contextual commentary, to explain why a sentence sounded awkward, and even to suggest alternative phrasings while preserving the student's original meaning. Suddenly, the tool did not merely mark errors; it talked about them. For EFL teachers drowning in drafts, this looked like a lifeline. For researchers, it raised a set of questions that earlier automated systems never quite provoked.

The most urgent of these questions cluster around two outcomes that pull in slightly different directions. The first is writing accuracy-whether AI-generated feedback actually helps students produce grammatically, lexically, and discursively better texts, and whether it does so as effectively as a trained teacher. The second is learner autonomy-whether access to instant, unlimited feedback encourages students to monitor their own language, reflect on their choices, and gradually take charge of their writing development, or whether it simply trains them to outsource every doubt to an algorithm. These two outcomes matter because they define what successful writing instruction ought to achieve: not merely a correct final draft, but a writer who can eventually produce correct drafts alone.

The tension between them is real. A student who pastes every paragraph into ChatGPT and accepts every suggestion without question may end up with a cleaner essay, but not with any clearer understanding of why the changes were needed. Accuracy improves on paper; autonomy may quietly erode. Conversely, a student who interrogates every AI suggestion, checks it against what they know, and revises deliberately is practicing exactly the kind of self-directed learning that Little (2007) and others have long argued sits at the heart of language education. The tool, in other words, does not determine the outcome. The surrounding pedagogy-and the student's own engagement-does.

This review enters that conversation not by adding another small-scale experiment, but by stepping back to ask what the existing evidence, taken together, actually suggests. The empirical literature on AI-assisted feedback in EFL writing has expanded rapidly since 2023,

but it remains fragmented. Some studies compare AI and teacher feedback on grammatical accuracy; others examine student attitudes or self-efficacy; a smaller number look at self-regulation or revision behavior. Rarely do they examine accuracy and autonomy side by side, as interdependent concerns rather than separate variables. That is the gap this review addresses.

Specifically, the following questions guide the analysis:

- What does current peer-reviewed research indicate about the effects of AI-assisted feedback on EFL writing accuracy?
- What does it suggest about the effects of such feedback on learner autonomy and self-regulated learning?
- How do these two outcomes relate to one another, and what pedagogical conditions appear to shape that relationship?
- What risks, limitations, and unanswered questions remain?

The scope is deliberately bounded. The review concentrates on empirical studies conducted in EFL contexts-where English is learned as a foreign language rather than acquired through immersion-and on research published between 2020 and 2026, with foundational theoretical work drawn from earlier scholarship where necessary. Sources were selected from openly accessible databases and evaluated for methodological rigor, not merely for topical relevance. The aim is not to declare AI feedback either beneficial or harmful, but to synthesize what is known, identify what is not, and offer a clear-eyed account of where the field stands as EFL programs worldwide begin integrating these tools into their writing classrooms.

Review methodology

A literature review is only as trustworthy as the path taken to assemble its sources. If the selection process is opaque, the synthesis becomes a personal anthology rather than an examination of evidence. This section describes exactly how the studies underlying this review were located, filtered, and organized, and it acknowledges the constraints that shaped those decisions.

This review is best described as a structured thematic literature review. It employs systematic search procedures and explicit inclusion criteria, but it does not claim the exhaustive retrieval and statistical synthesis of a full systematic review. That choice is deliberate. The field of generative AI in language education is evolving too rapidly for a frozen systematic protocol to capture its current state, and many early studies exist only as advance-online publications without final pagination. A structured thematic approach allows the review to be rigorous without pretending to a completeness that the current literature does not yet permit.

Search Strategy. The search began with databases selected for their coverage of applied linguistics, education technology, and open-access scholarship: Google Scholar, ERIC, the Directory of Open Access Journals (DOAJ), and ResearchGate. Scopus and Web of Science were consulted where institutional access permitted, but the primary pool was built from sources freely available to researchers without subscription barriers. That decision carries a limitation-some empirical work published behind paywalls in *System*, *Language Learning & Technology*, or *TESOL Quarterly* may have been missed-yet it also reflects the reality of many EFL practitioners and early-career researchers who rely on open repositories. Acknowledging that constraint upfront is preferable to pretending a comprehensive sweep was possible.

Search strings were constructed in logical combinations rather than single keywords, to avoid drowning in irrelevant hits. The following combinations were entered across platforms:

- "AI-assisted feedback" AND "EFL writing"
- "ChatGPT" AND "EFL writing" AND "feedback"
- "generative AI" AND "writing accuracy" AND "EFL"

- "automated written corrective feedback" AND "EFL"
- "AI feedback" AND "learner autonomy"
- "large language model" AND "L2 writing" AND "revision"

Each string was run with quotation marks to preserve phrase integrity, then modified by dropping or adding terms when the initial yield proved too narrow or too broad. For instance, replacing "EFL" with "L2 writing" occasionally surfaced relevant studies from ESL contexts that included EFL sub-samples, though these were later screened against the inclusion criteria described below.

The search was conducted between July and August 2026, with a publication window of 2020 to 2026. The lower bound was set at 2020 because that period marks the rapid expansion of empirical research on generative-AI-assisted feedback, following the public release of ChatGPT in late 2022. Earlier foundational work on automated writing evaluation, written corrective feedback, self-regulated learning, and learner autonomy was retained where necessary to establish theoretical continuity, but the empirical core of the review is anchored in the period when large language models became available to classrooms.

Inclusion and Exclusion Criteria. Not every article mentioning "ChatGPT" and "writing" deserved a place in this review. A two-stage screening process was applied. First, titles and abstracts were read for relevance. Second, full texts were examined to confirm that the study met all inclusion criteria and violated none of the exclusion rules.

Inclusion criteria were as follows:

- Peer-reviewed empirical research, published in journals or edited volumes with clear review processes. Conference proceedings were considered only when they reported original data with transparent methodology.
- Studies situated in EFL or mixed EFL/ESL contexts where English was explicitly taught as a foreign language.
- Studies involving AI-assisted, automated, or generative-AI feedback on student writing.
- Studies reporting at least one measurable outcome related to writing accuracy, revision quality, learner autonomy, self-regulation, self-efficacy, or engagement with feedback.
- English-language publications, to ensure consistent interpretation of methodological descriptions.
- Studies whose bibliographic details-including volume, issue, page numbers, or stable DOI-could be independently verified. Advance-online publications without final pagination were excluded unless they contained sufficient metadata to confirm peer-reviewed status.

Exclusion criteria removed studies that:

- Discussed AI in education generally without specific attention to writing or feedback.
- Were purely technical papers describing algorithmic architecture without educational application or learner data.
- Relied exclusively on student opinion surveys without any writing sample, accuracy measure, or behavioral indicator.
- Duplicated data already reported in another included study.
- Focused on speech recognition, translation, or plagiarism detection rather than writing feedback.
- Could not be verified through database records, publisher metadata, or stable institutional repositories.

This process yielded an initial pool of roughly twenty-five potentially relevant studies. After full-text screening and bibliographic verification, nine empirical studies met all criteria and form the evidentiary core of this review. A literature matrix was constructed to record author, year, context, participant profile, AI tool used, feedback type, main findings on accuracy, main findings on

autonomy or related constructs, and noted methodological limitations. That matrix-available in the Appendix-served as the organizing backbone for the thematic synthesis that follows.

Analytical Framework. The review was structured thematically rather than chronologically. A chronological arrangement would have produced a parade of studies-"Author A Found X, Author B found Y"-which is the habit of an annotated bibliography, not a synthesis. Instead, studies were grouped according to the conceptual questions they addressed: error identification and accuracy, revision behavior, self-regulation and autonomy, and the conditions under which AI feedback is mediated by teachers or task design.

Within each theme, the analysis proceeded by identifying points of convergence, contradiction, and ambiguity. Where two studies reported similar accuracy gains using different AI tools, the pattern was noted. Where one study found AI feedback comparable to teacher feedback on grammatical accuracy and another found teacher feedback superior, the moderator variables-participant proficiency, feedback timing, error type, measurement instrument-were examined to understand why the discrepancy might exist. This interpretive work is necessarily provisional; the field is young, and the empirical base is small. But provisional synthesis is still preferable to mere inventory.

Finally, methodological quality was treated as part of the analysis, not as an afterthought. Studies with control groups, pre-post designs, and multiple measures were weighted more heavily in the synthesis than one-shot surveys or uncontrolled pilot reports. Self-report measures of autonomy were read with appropriate caution, since what learners say about their own self-direction does not always match what they do when faced with a blank page. That skepticism is woven into the discussion in Sections 4 and 5 rather than confined to a perfunctory "limitations" paragraph at the end.

Theoretical background

Before the empirical evidence can be interpreted, the concepts that organize it need to be clear. Writing accuracy, learner autonomy, and AI-assisted feedback do not exist in a theoretical vacuum; each carries assumptions about how people learn languages, how they respond to correction, and what it means to become an independent writer. This section traces the intellectual lineage behind those assumptions, moving from the mechanics of error correction to the broader question of what feedback is ultimately for.

Written Corrective Feedback. The practice of correcting student writing has been debated in second-language research for decades, and the debate has never been simply about whether teachers should mark errors. The deeper question is which errors to mark, how to mark them, and what students are supposed to do with the information. Bitchener and Ferris (2012) provided the most influential taxonomy for thinking about these choices, distinguishing between focused feedback-targeting a limited set of error types such as articles or past tense-and unfocused feedback, which responds to whatever errors appear in a text. They also drew a line between direct correction, where the teacher supplies the right form, and indirect correction, where the teacher merely indicates that an error exists and leaves the repair to the student. Each combination carries different cognitive demands. Direct correction may help a student produce a cleaner draft in the moment, but indirect correction potentially forces the kind of problem-solving that leads to longer-term retention, provided the student has the linguistic knowledge to solve the problem in the first place.

Ferris (2006) approached the same territory from a slightly different angle, asking whether error feedback helps student writers at all. Her answer, grounded in a survey of existing

research, was cautiously affirmative-but with important conditions. Feedback works best when it is clear, consistent, and tied to revision opportunities. A comment scrawled in red ink and returned two weeks after submission, with no chance to rewrite, is unlikely to change much. A targeted note that a student can act on immediately, within a task structure that encourages reflection, stands a better chance. The implication for this review is straightforward: when we evaluate AI-assisted feedback, we are not really evaluating the tool itself. We are evaluating whether the tool creates conditions that satisfy these established principles of good correction.

What AI tools add to this landscape is scale and speed. A teacher can give focused, indirect feedback to perhaps fifteen students before the quality degrades; an algorithm can do it for fifteen thousand. But scale does not automatically preserve the cognitive conditions that make feedback effective. If an AI system corrects every error directly, instantly, and without requiring the student to think about why the change was necessary, it may violate the very principles Bitchener and Ferris identified as crucial for learning. That tension-between the efficiency of automation and the cognitive demands of effective correction-runs through much of the empirical literature reviewed in Section 4.

Formative Feedback and Self-Regulated Learning. If corrective feedback is about fixing what is wrong, formative feedback is about steering what comes next. The distinction matters because AI-assisted feedback can operate in either mode. A grammar checker that simply underlines errors is essentially corrective. A system that explains why a sentence is awkward and suggests strategies for restructuring it is moving into formative territory, even if it never uses the word.

Nicol and Macfarlane-Dick (2006) argued that the most powerful feedback is the kind that closes the gap between current performance and intended standards while simultaneously building the learner's capacity to monitor that gap independently. They proposed seven principles of good feedback practice: it should clarify what good performance looks like; facilitate self-assessment; deliver high-quality information to students about their learning; encourage teacher and peer dialogue; promote positive motivational beliefs; provide opportunities to close the gap between current and desired performance; and provide information to teachers that can be used to shape their own instruction. The list is deceptively practical. Behind each principle sits a theory of learning in which the student is gradually transferred from external guidance to internal control.

That transfer is what self-regulated learning describes. When a student receives a comment-from a teacher, a peer, or an algorithm-the immediate response is reactive: read, understand, fix. The deeper response is regulatory: the student begins to anticipate the comment before it arrives, checking their own text against the criteria they have internalized. Nicol and Macfarlane-Dick called this the shift from external feedback to internal self-monitoring, and they treated it as the defining purpose of formative assessment. For this review, their framework raises a critical question about AI tools. *Do they accelerate that shift, giving students enough practice and explanation that they eventually need less external help? Or do they arrest it, creating a permanent dependency on an external checker?* The empirical studies examined in Section 5 speak to that question directly, though they do not yet answer it conclusively.

Learner Autonomy. The concept of autonomy has been defined in various ways across educational research, but in language education it carries a specific weight. Little (2007) argued that autonomy is not merely the ability to study alone, detached from a classroom or a teacher. It is the capacity to take charge of one's own learning, which includes setting goals, selecting methods, monitoring progress, and evaluating outcomes. That

capacity is not innate; it develops through practice, reflection, and - crucially - interaction with the target language itself. Little insisted that learner autonomy and target language proficiency are not separate achievements but mutually reinforcing ones. The more a learner can do with the language independently, the more autonomous they become; the more autonomous they are, the more they seek out opportunities to use and refine the language.

This has direct implications for writing instruction. An autonomous writer is not someone who simply writes without help. It is someone who, when stuck, knows what kind of help to seek, can evaluate whether that help is reliable, and can integrate it into an independent compositional process. Little, Dam, and Legenhausen (2017) expanded this view by emphasizing three interconnected elements: learner involvement in decision-making, reflection on the learning process, and use of the target language as the primary medium of thought and communication. A student who switches mentally to their first language every time they encounter a difficulty, or who outsources every compositional decision to a tool, is not using the target language as a medium of learning. They are using it as a surface to be polished by something else.

That distinction becomes urgent when generative AI enters the classroom. If a student uses ChatGPT to brainstorm ideas, check grammar, and revise structure, the activity could be interpreted as highly autonomous-the student is making choices about when and how to use a resource. But if the same student accepts every suggestion uncritically, without reflection, without evaluating whether the change matches their intended meaning, the activity begins to look less like autonomous learning and more like delegated production. Little's framework helps us see that autonomy is not about the presence or absence of a tool; it is about the quality of the learner's engagement with it.

From Automated Writing Evaluation to Generative AI Feedback. The theoretical shift from earlier automated writing evaluation to contemporary generative AI is not merely a story of better technology. It is a story of changed relationships between the writer, the text, and the source of feedback.

Early automated systems-Grammarly, Criterion, Pigai, and their predecessors-operated on rule-based or statistical models. They compared student text against databases of correct forms and flagged deviations. The feedback was immediate and consistent, but it was also thin. It could tell a student that an article was missing; it could not explain why the article mattered for the coherence of the paragraph, nor could it engage in a dialogue if the student asked for clarification. These tools represented an important step in the automation of writing support, and research from the 2010s established both their utility for surface-level error detection and their limitations for rhetorical and organizational feedback. Theoretically, they sat comfortably within a behaviorist tradition: writing was a set of patterns to be reinforced, and feedback was a stimulus designed to elicit a corrected response.

Generative AI, by contrast, is built on large language models trained to predict and produce human-like text. When such a model responds to a student essay, it is not simply matching patterns against a database. It is generating a novel response conditioned on the specific content, style, and apparent proficiency level of the draft it has read. That capacity opens the door to feedback that is dialogic, explanatory, and adaptive-the kind of responsive interaction that Vygotskian theory associates with learning within the zone of proximal development. In principle, an AI that can ask "What did you mean here?" and adjust its advice based on the student's reply is functioning as a scaffold rather than as a corrector.

But the theoretical promise comes with theoretical risks. A scaffold, in the Vygotskian

sense, is temporary and designed to be withdrawn as the learner's competence grows. An AI feedback tool, however, is always available, always patient, and never too busy to respond. There is no natural pressure for the learner to internalize the support because the support does not go away. The question, then, is whether generative AI functions as a scaffold that fades-which would support both accuracy and autonomy-or as a permanent prosthesis, which might improve surface-level accuracy while leaving the underlying writer unchanged. The empirical sections that follow examine what the evidence suggests about that question, but the theoretical framework makes clear that the answer depends less on the algorithm than on how the classroom around it is structured.

AI-Assisted Feedback and Writing Accuracy

The most immediate question teachers ask about AI feedback is also the simplest: *does it make student writing more accurate?* The evidence from the studies included here suggests a carefully qualified yes. AI-assisted feedback appears to improve accuracy under certain conditions, but the improvement is uneven across error types, and the comparison with teacher feedback is far from settled. This section unpacks that qualification using the nine empirical studies that met the review's inclusion criteria.

What AI Detects and What It Misses. At the level of surface errors, large language models appear to perform remarkably well. Missing articles, subject-verb disagreement, wrong prepositions, and awkward collocations-the mechanical layer of language that occupies much of a teacher's marking time-are generally identified with reasonable consistency. Özdere (2025), working with EFL students on academic writing tasks, found that AI tools handled these mechanical issues reliably, often flagging errors that busy instructors might have overlooked in a long draft. That pattern aligns with what Mizumoto, Eguchi, and Oyama (2024) observed when they compared ChatGPT's error coding against human raters working with the Cambridge Learner Corpus: the model caught a high proportion of grammatical deviations and matched human judgment more often than earlier automated systems managed.

But accuracy is not only mechanical, and this is where the evidence becomes more tentative. When Mizumoto and colleagues examined nuanced cases-sentences where grammaticality depended on context, intention, or rhetorical effect-the alignment with human judgment weakened. ChatGPT would sometimes smooth out a phrasing that a human reader might have preserved because it carried the student's voice. The model, trained to predict the most statistically likely next word, appears to have a bias toward the conventional. That bias may be useful for standardized test essays; it is less clear whether it serves writing that attempts something stylistically unexpected.

The distinction matters because writing accuracy in EFL instruction is not supposed to mean merely "grammatically perfect." It means the student can control the language well enough to convey intended meaning without distraction. An AI that systematically irons out stylistic variation in the name of correctness may improve a surface metric while damaging the writer's developing sense of voice. That risk is not always captured in accuracy scores, and it suggests that researchers need to clarify what kind of accuracy they are actually measuring before claiming improvement.

AI Feedback versus Teacher Feedback: The Mixed Record. Whether AI feedback produces accuracy gains comparable to those from a trained teacher remains unresolved. Two of the included studies found that teacher feedback produced stronger accuracy outcomes than AI, at least under the conditions tested. Khodaie Alvar and Sahebkheir (2025), working with forty pre-intermediate Iranian EFL learners, reported that students who received teacher corrective feedback made greater gains in grammatical accuracy than those using ChatGPT-4, even though both groups

improved relative to a control condition. Wang (2024), comparing teacher feedback against AI-generated corrective feedback delivered through the Poe application, likewise found that teachers produced stronger accuracy outcomes, while the AI excelled instead on lexical diversity-suggesting that the two sources may be good at different things.

Yet the pattern is not universal. Alnemrat et al. (2025), using sixty Jordanian undergraduates working on argumentative essays, found no statistically significant difference between AI and teacher feedback on overall writing quality and accuracy measures. Both groups improved; neither source held a clear advantage. That finding is important because it came from a context with a larger sample and a different task type than the Iranian study, hinting that the AI-versus-teacher question may not have a single answer. It may depend on who the students are, what they are writing, and how the feedback is delivered.

One possible moderator is proficiency. Khodaie Alvar and Sahebkheir worked with pre-intermediate learners, while Alnemrat et al. used university undergraduates who likely possessed a higher baseline proficiency. The discrepancy between these findings raises the possibility that lower-proficiency students may benefit differently from AI feedback than advanced students, though too few studies have explicitly tested this variable to treat it as established. The safer interpretation is that AI feedback does not consistently outperform teacher feedback, nor does it consistently underperform; the comparison appears to depend on proficiency level, error type, and how the feedback is integrated into the writing process.

From Feedback to Revision: Does It Stick? Accuracy on a final draft is one thing; accuracy that persists into the next essay is another. The empirical studies included here are mostly short-term-interventions lasting a few weeks, with post-tests administered shortly after the feedback cycle. Within that narrow window, the evidence on revision behavior is cautiously positive. Yang, Hussin, and Hashim (2025), working with eighty-two university EFL students, found that ChatGPT feedback promoted active revision: students engaged strategically with the suggestions, rewrote sentences rather than simply accepting replacements, and reported a sense of ownership over the revised text.

Still, the gap between successful revision and lasting learning remains largely unmeasured. A student who corrects an article error because ChatGPT highlighted it may produce a cleaner paragraph today without necessarily remembering the rule next month. The same concern applies to teacher feedback, of course, but teachers have informal mechanisms for reinforcement-returning to the same error type across assignments, explaining patterns in class, building a relationship that encourages the student to pay attention-that an algorithm does not automatically replicate. Whether AI feedback can be embedded in a pedagogical structure that produces retention, rather than just immediate correction, is one of the open questions this review returns to in Section 8.

Synthesis: What Accuracy Research Actually Shows. Taken together, the included studies suggest a profile rather than a verdict. AI-assisted feedback appears to be a competent mechanical proofreader. It identifies grammatical, lexical, and collocational errors with a hit rate that rivals' human raters on clear-cut cases and falls short on ambiguous or context-dependent ones. It improves accuracy on the draft in front of the student, and sometimes on the next draft as well, though the evidence for long-term retention is thin. It does not consistently outperform teacher feedback, but it does not consistently underperform either; the comparison depends heavily on the population, the error type, and the pedagogical conditions.

Most importantly, accuracy gains from AI feedback appear to be bounded by what the

tool can see. It sees sentences well and paragraphs less well. It sees conventions clearly and intentions dimly. It sees what is statistically probable and may miss what is intentionally deviant. For EFL programs that treat writing as the production of grammatically correct text, these boundaries may not matter much. For programs that treat writing as the development of an independent writer who can make conscious choices about language, the boundaries are significant. They point toward a conclusion that runs through the rest of this review: the value of AI feedback depends on what the teacher asks the student to do with it.

AI-Assisted Feedback and Learner Autonomy

If accuracy is about what the text looks like, autonomy is about what the writer becomes. The distinction is easy to state and hard to measure. A student who submits a flawless essay may have learned a great deal, or may have learned only how to delegate the difficult parts to an algorithm. Conversely, a student who struggles through a messy draft, evaluates conflicting suggestions, and revises with partial success may have advanced further as a writer even if the final product is less polished. This section examines what the included literature suggests about where, on that spectrum, AI-assisted feedback tends to place EFL learners.

Confidence, Engagement, and the Affective Shift. One of the more consistent findings across the included studies is that students often prefer using AI feedback, and that this preference translates into measurable changes in how they feel about writing. Yang, Hussin, and Hashim (2025), following eighty-two university EFL students through a ChatGPT-assisted revision cycle, reported that participants found the feedback less intimidating than teacher commentary. An algorithm does not carry the implicit evaluation of a grade, does not remember last week's mistakes, and does not convey disappointment. For students who have learned to associate teacher feedback with anxiety, that emotional neutrality is a structural change in the writing environment.

The affective shift appears to have behavioral consequences. Nguyen and Doan Thi Hue (2025), measuring learner autonomy and self-efficacy as distinct constructs within a self-regulated learning framework, found that the ChatGPT group showed significantly greater gains in autonomy scores than a comparison group. The researchers interpreted this as evidence that the tool was not merely making students feel better but was changing how they conceptualized their own role in the writing process.

But feeling in control is not the same as being in control, and studies that rely on self-report measures leave that distinction unresolved. A student who reports greater autonomy may simply be reporting greater comfort with a tool that handles the decisions for them. The questionnaire does not always capture the difference. That is why the behavioral evidence-what students actually do when they sit down to write-matters more than the attitudinal data. Autonomy measured through self-report instruments may reflect perceived rather than exercised self-direction, a distinction that the current empirical base does not fully resolve.

Self-Regulation and the Mediation of Writing. When researchers look at behavior rather than attitude, the picture becomes more textured. Zhai (2025), working with intermediate Chinese EFL learners using Grammarly alongside teacher feedback, found that the combination significantly enhanced self-regulation. Students began to monitor their own output more actively, pausing to check predicted errors before submitting drafts, and they reported increased confidence in their ability to identify problems independently. The researchers described this as a shift from reactive correction to proactive monitoring-the kind of internal feedback loop that Nicol and Macfarlane-Dick (2006) identified as the ultimate purpose of formative assessment.

Han and Li (2024) approached a similar question through classroom observation rather than individual measurement. In their study with four instructors and 102 Chinese EFL students, the most successful outcomes occurred when AI-generated comments were staged before teacher review. The AI identified surface-level issues; the teacher added meta-level guidance about why certain errors recurred and used the saved time to conference with students about argumentation. Students in that condition reported feeling more in control of their revision process than students in either pure-AI or pure-teacher conditions. The finding suggests that autonomy may grow not from the volume of feedback but from the space created for the student to act on it deliberately.

The difficulty is that this agency is not automatic. A poorly designed mediating tool can deskill the user rather than develop them. A GPS that dictates every turn does not teach someone to read a map; it teaches them to follow instructions. The same risk applies to AI writing feedback. If the interface is designed for one-click acceptance-if the path of least resistance is to approve every suggestion and move on-then the tool may reduce the cognitive work of revision rather than redistribute it toward the learner.

The Dependency Risk. That risk is not merely theoretical. AI feedback is fast, fluent, and syntactically authoritative. A student who is tired, pressed for time, or uncertain about their own English proficiency faces a strong temptation to accept suggestions rather than interrogate them. The algorithm does not ask "Are you sure?" in a way that forces reflection. It presents a correction and moves on. Without explicit pedagogical scaffolding-without someone teaching the student to pause, compare, and justify-the default behavior is likely to be acquiescence. That is not autonomy. It is the opposite: a transfer of editorial control from the learner to the machine.

The contrast with teacher feedback is instructive. A teacher who marks an error and the student does not understand can ask for clarification. The interaction is dialogic, and the dialogue itself is a learning event. Most AI feedback systems, even generative ones, are not genuinely dialogic in practice because the student does not typically initiate the exchange. The AI comments; the student reads. The power to question, to push back, to say "I think my original version was better because..." resides with the student in theory but is rarely exercised in practice. Rebuilding that critical stance is, as this review will argue in Section 9, a pedagogical task rather than a technical one.

Synthesis: What Autonomy Research Concludes. The included evidence on autonomy is less settled than the evidence on accuracy, but it points in a clear direction. AI-assisted feedback can support learner autonomy when the surrounding conditions are right. It can reduce anxiety, increase engagement, and create opportunities for self-monitoring and strategic revision. It can function as a mediating tool that shifts responsibility toward the learner, particularly when the learner has enough proficiency to evaluate the feedback and enough time to do so deliberately.

But those conditions are not met automatically. They require intentionality from the teacher, the task designer, and the student. Without that intentionality, the same tool that promises autonomy can deliver dependency-a writer who produces accurate text by outsourcing every decision to an algorithm that is always available, always patient, and never asks the student to think for themselves. The question for EFL instruction is not whether AI feedback supports autonomy. It is whether we are structuring the writing classroom so that it must.

The Relationship Between Accuracy and Autonomy

So far this review has examined writing accuracy and learner autonomy as separate outcomes. But they are not separate. Treating them as independent variables may be the most common conceptual weakness in the

emerging field of AI-assisted writing research. A study that measures accuracy gains without asking how they were achieved, or that measures autonomy through self-report without checking whether the student's writing actually improved, risks celebrating an illusion. This section argues that the relationship between the two outcomes is not fixed; it is constructed by the pedagogical choices that surround the technology. It then proposes a conceptual model, derived from the synthesis of evidence reviewed here, that makes that construction visible.

The Virtuous Pathway. The most productive relationship between accuracy and autonomy can be described as a reinforcing loop. A student produces a draft with predictable errors-article omissions, verb tense slips, awkward transitions. The AI identifies these problems and explains them in accessible language. The student reads the explanation, compares it against what they intended to say, and decides whether the suggestion preserves their meaning. They revise, but they do not simply paste. They rewrite the sentence, sometimes imperfectly, sometimes with a new error, but with an awareness that something needed changing. The next draft is more accurate not because the AI corrected it, but because the student understood enough to correct it themselves. Over time, the student begins to anticipate the feedback before it arrives. They pause while drafting to check their own articles. They have internalized a standard that was initially external.

This pathway connects accuracy to autonomy in a cycle that strengthens both. The feedback is accurate enough to be trustworthy; the student is autonomous enough to engage with it critically; the engagement produces accurate revision; the accurate revision builds confidence; the confidence encourages further autonomous engagement. It is a virtuous cycle, and it is not accidental. It depends on what Han and Li (2024) observed in their classroom study: a structure in which the algorithm handles the analytical labor of error identification while the teacher maintains the relational and interpretive work that turns suggestions into learning. In their study, students who received staged AI-and-teacher feedback produced more accurate texts and reported feeling more in control of their revision process than students in either pure-AI or pure-teacher conditions.

Zhai (2025) reported a related pattern. Students using Grammarly alongside teacher feedback showed enhanced self-regulation-they monitored their own output more actively-while also demonstrating accuracy gains. The combination suggests that when AI feedback is embedded in a pedagogical structure that requires independent judgment, the two outcomes can move together.

The Dependency Pathway. But the loop can run backward. Consider the same student with the same draft, but now the task is due in two hours, the classroom culture treats the draft as a product to be submitted rather than a process to be understood, and the AI interface offers a single button that applies all suggested corrections at once. The student clicks it. The essay is now grammatically accurate, perhaps more accurate than anything they could have produced alone. They submit it, receive a decent grade, and learn that accuracy can be achieved with minimal effort. The next time they face a blank page, their first thought is not "What do I want to say?" but "How quickly can I get the AI to fix this?"

This is the dependency pathway. The text gains accuracy; the writer loses autonomy. The two outcomes have become antagonists rather than allies. The student has learned to produce accurate surface language without developing the underlying competence to judge whether that language is accurate-a form of learned helplessness that technology makes efficient.

The distinction between the two pathways is not determined solely by the AI tool itself. Both scenarios could involve the same version of ChatGPT. The difference lies in what the student is asked to do between receiving the feedback and submitting the

revision. If the intervening space is filled with evaluation, reflection, and independent rewriting, accuracy and autonomy tend to improve together. If that space is collapsed-if the feedback flows directly into the final text without passing through the student's own judgment-accuracy may still rise, but autonomy will likely fall. The tool is not entirely neutral; interface design can nudge behavior toward acquiescence or reflection. But the decisive factor is pedagogical.

A Conceptual Model of AI-Mediated Writing Development. Drawing together the evidence reviewed in Sections 4 and 5, this review proposes the following conceptual model as a framework for understanding how AI-assisted feedback operates in EFL writing instruction. The model is not empirically tested here; it is a theoretical synthesis derived from the patterns observed across the included studies. Its purpose is to make the relationship between accuracy and autonomy explicit, and to identify the pedagogical levers that determine which pathway a given instructional design will produce.

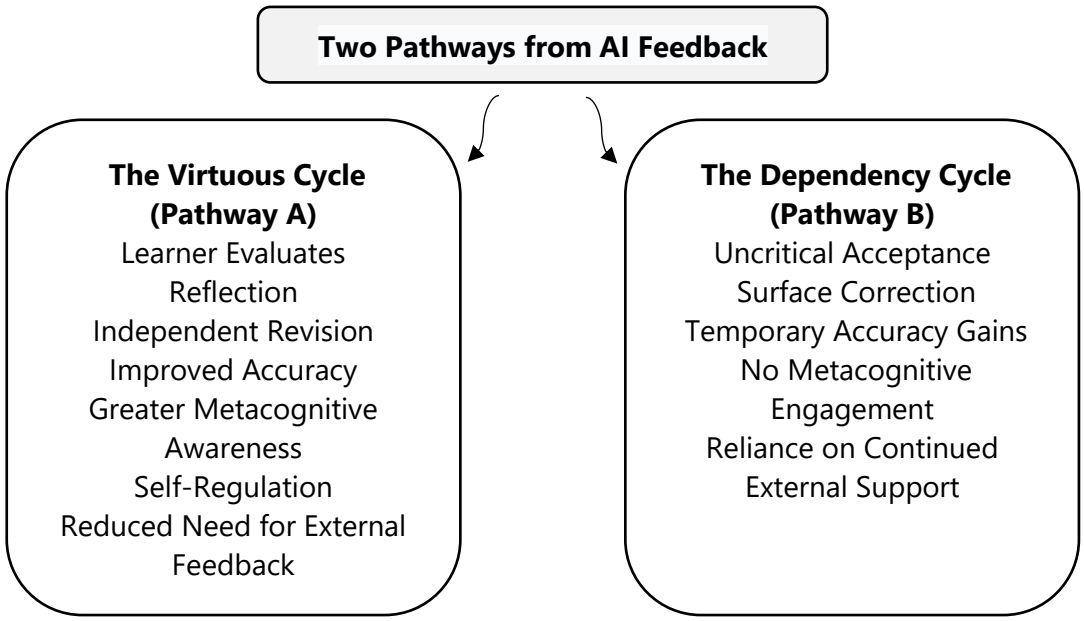


Figure 1. *Conceptual Model of AI-Mediated EFL Writing Development*

Note. *Pathway A and Pathway B represent ideal-types, not empirical certainties. Actual learner behavior may fall between these extremes.*

A flowchart showing two branching pathways from "AI Feedback." Pathway A (Virtuous Cycle): Learner Evaluates → Reflection → Independent Revision → Improved Accuracy → Greater Metacognitive Awareness → Self-Regulation → Reduced Need for External Feedback.

Pathway B (Dependency Cycle): Uncritical Acceptance → Surface Correction → Temporary Accuracy Gains → No Metacognitive Engagement → Reliance on Continued External Support.

The critical node in both pathways is the learner's evaluation of the feedback. When the student must decide whether to accept, modify, or reject a suggestion-and when that decision is supported by sufficient proficiency, time, and pedagogical guidance-the pathway tends toward autonomy. When the student bypasses that decision, the pathway tends toward dependency. This aligns with Little's (2007) argument that autonomy is not the absence of help but the quality of engagement with it, and with Nicol and Macfarlane-Dick's (2006)

principle that effective feedback must build the learner's capacity for self-monitoring.

The model also suggests why the empirical literature has produced mixed results. Studies that find AI feedback beneficial for autonomy (Nguyen & Doan Thi Hue, 2025; Zhai, 2025; Han & Li, 2024) tend to embed the feedback in tasks that require active evaluation. Studies that find no benefit, or that express caution, tend to examine AI feedback as a standalone intervention without specifying the pedagogical conditions. The model predicts that future research will continue to produce contradictory findings until studies routinely report not just whether AI feedback was used, but how students were expected to engage with it.

Pedagogical Design as the Deciding Factor. This insight shifts the focus from the algorithm to the assignment. Teachers cannot change how a large language model processes language, but they can change the conditions under which students encounter its output. Several concrete approaches emerge from the included studies.

One approach is to make the evaluation of AI feedback an explicit part of the task rather than an assumed skill. Instead of asking students to "revise your essay using ChatGPT's suggestions," a teacher might ask them to produce two documents: the revised essay and a reflective memo explaining which suggestions they accepted, which they rejected, and why. This simple change turns the revision process from a passive reception into an active judgment. It forces the student to articulate criteria- "I rejected this suggestion because it changed my meaning" - which is exactly the kind of metacognitive practice that autonomy theorists describe as central to self-directed learning.

Another approach, emerging from Han and Li (2024), is to stage the feedback so that the AI does not have the last word. Students first receive AI-generated comments, then discuss selected comments in small groups or with the teacher, and only then revise. That staging introduces a social layer between the algorithm and the final text. It creates space for disagreement, for a student to say "I don't think this suggestion is right," and for a peer or teacher to validate that skepticism. The AI becomes one voice in a conversation rather than the final authority.

A third approach is to teach a protocol rather than a tool. The sequence proposed in this review is worth restating here not as a slogan but as a pedagogical architecture: **Generate, Evaluate, Explain, Revise, Reflect.** Generate the draft independently. Evaluate the AI feedback against your own knowledge and intention. Explain your decision to accept or reject each significant suggestion. Revise the draft yourself, not by cutting and pasting. Reflect on what you learned about your own typical errors. That sequence treats AI feedback as a resource within a self-regulated process, not as a replacement for the process itself. It is demanding-more demanding, in some ways, than simply receiving teacher corrections-but that is precisely the point. Autonomy is not the absence of effort. It is the willingness to exert effort on one's own behalf.

Synthesis: The Interdependence Argument. The research reviewed here does not support the conclusion that AI-assisted feedback is either good or bad for EFL writers. It supports the more specific conclusion that the effects of such feedback are mediated by how the classroom is organized around it. Writing accuracy and learner autonomy are not separate levers that teachers can pull independently. They are interdependent outcomes of the same underlying process: the student's gradual transition from external support to internal control.

When that transition is supported-when feedback is accurate enough to be useful, dialogic enough to be questioned, and embedded in tasks that require independent judgment-both accuracy and autonomy tend to improve. The student becomes a more accurate writer because they are becoming a more autonomous one, and vice versa. When

the transition is bypassed-when feedback is treated as a delivery system for correct answers rather than as a stimulus for independent thinking-accuracy may still improve on the page in front of the teacher, but it will be fragile. It will depend on the continued presence of the tool, and it will leave the student no better equipped to write without it than they were before.

That interdependence has implications that extend beyond the individual classroom. It suggests that the most important research question in this field is not "Does AI feedback improve writing?" but "Under what conditions does AI feedback produce writers who no longer need it?" The empirical literature is only beginning to address that question. The next two sections examine why that is-what the current studies cannot yet tell us, and what risks they have identified along the way.

Limitations and Risks

The preceding sections have described what AI-assisted feedback can do under favorable conditions. This section turns to what it cannot do, what it does poorly, and what it may do harmfully if adopted without caution. No technology that enters the classroom does so without trade-offs, and generative AI is no exception. The risks are technical, pedagogical, and ethical, and they overlap in ways that make simple mitigation difficult.

Technical Limitations. The most immediate problem is that AI feedback is not always correct. Large language models are trained to predict plausible text, not to verify truth, and that distinction matters when the text in question is a student's emerging English. Mizumoto, Eguchi, and Oyama (2024) documented this clearly: while ChatGPT aligned well with human raters on straightforward grammatical errors, its reliability dropped on nuanced cases-sentences that were technically non-standard but contextually appropriate, or errors that required understanding the student's intended meaning to judge correctly. An AI might smooth out a deliberate stylistic choice, flattening a voice that a human teacher would have recognized and preserved. It might miss an error entirely if the mistake happens to be statistically common in its training data, or it might generate an explanation for a correction that has no linguistic basis beyond the model's confidence in its own output.

This last problem-often called hallucination-appears to be a structural feature of how these models operate rather than an incidental glitch. They generate text by predicting the most probable next token, which means they can construct fluent, authoritative-sounding explanations for suggestions that are simply wrong. A student who receives such an explanation has no easy way to verify it unless they already possess the grammatical knowledge that the feedback was supposed to provide. For lower-proficiency EFL learners, who are precisely the population most likely to rely heavily on AI assistance, this creates a dangerous asymmetry: the tool sounds confident, the student is not, and the student's default response is to trust what sounds authoritative.

The problem is compounded by inconsistency. Mizumoto and colleagues noted that ChatGPT's error coding could shift with slight rephrasing of the prompt or different sampling settings. For a teacher, that unpredictability is manageable because human judgment is itself variable and can be discussed. For an algorithm, the variation is invisible. The student cannot ask the model why it changed its mind; they can only accept the latest output as given. That opacity undermines one of the core principles of effective feedback identified by Nicol and Macfarlane-Dick (2006): that students should receive information they can use to close the gap between current and desired performance. If the information fluctuates without explanation, it is hard to use systematically.

Pedagogical Risks. Beyond technical failure lies a set of risks that develop slowly, across semesters rather than single assignments. The most significant is overdependence-the gradual atrophy of the

student's own editorial judgment as they learn to defer to the algorithm. Over time, that pattern can produce what might be called *competence inflation*: a student whose submitted work exceeds their actual ability, and who struggles increasingly when the tool is unavailable-during a timed exam, for instance, or in a speaking task where no algorithm is present. The risk is not that the student cheats, but that they mistake the tool's competence for their own.

A related risk is the erosion of voice. Writing in a foreign language is not only about correctness; it is about developing the capacity to express a personal perspective through an unfamiliar medium. An AI trained on statistically probable text has a bias toward the conventional, the safe, the middle of the distribution. It may smooth out idiosyncrasy, normalize unusual phrasing, and nudge every student toward the same blandly competent style. For advanced EFL learners, this is particularly damaging, because it is precisely at the advanced level that writers begin to make conscious choices about register, tone, and rhetorical effect-choices that may deliberately violate a rule for stylistic reasons. An AI that "corrects" such choices is not teaching writing; it is teaching conformity.

There is also the risk of reduced teacher-student interaction. One of the advertised benefits of AI feedback is that it frees teachers from the mechanical labor of marking surface errors, allowing them to focus on higher-order concerns. Han and Li (2024) suggest that this benefit is real when the freed time is reinvested in dialogue. But in under-resourced programs, the danger is that AI feedback becomes a substitute for teacher attention rather than a supplement to it. A department facing budget pressure might see AI as a way to increase class sizes while maintaining the illusion of individualized feedback. The student receives detailed comments on every draft, but never speaks to a human about their writing. The efficiency gain is real for the institution; the educational cost is borne by the learner.

Ethical and Equity Concerns. The risks extend beyond the individual classroom to broader questions of fairness and human agency. UNESCO's guidance on generative AI in education (2023, 2024) frames these concerns within a human-centred approach that insists on protecting learner agency, ensuring ethical validation of AI systems, and maintaining age-appropriate use. The 2023 guidance is particularly explicit: AI should support human decision-making, not replace it; it should be transparent in its operations; and its deployment should be governed by principles of inclusion and non-discrimination.

Those principles are not always met in practice. Access to AI tools is uneven. Students with reliable internet, modern devices, and institutional subscriptions to premium AI services have a significant advantage over peers who rely on free, rate-limited versions or who lack connectivity altogether. The EFL context amplifies this disparity because it often involves learners in regions where technological infrastructure is less developed. A literature review that celebrates AI feedback without acknowledging this access gap risks treating a privilege as a universal.

Data privacy presents another ethical layer. When students paste their drafts into a commercial AI platform, they are uploading their intellectual work to servers they do not control, often in jurisdictions with different data protection standards. For minor classroom assignments, this may seem trivial. But for students working on personal narratives, reflective essays, or eventually thesis drafts, the exposure is real. Institutions that mandate AI feedback without addressing data governance are asking students to trade privacy for convenience without a clear mechanism for informed consent.

Finally, there is the question of academic integrity, which sits uncomfortably at the boundary between assistance and substitution. If a student uses AI feedback to correct every error in their essay, at what point does the final text cease to be their own work? Current

policies struggle with this boundary because it is not fixed. A grammar checker that fixes articles is widely accepted; a model that rewrites entire paragraphs is generally considered cheating. But the space between those extremes is large and gray, and AI feedback tools operate throughout it. A student who receives a suggestion to restructure a paragraph, accepts it, and rewrites accordingly has done something different from a student who copies a generated paragraph wholesale-but both have relied on algorithmic input in ways that traditional notions of independent authorship did not anticipate. The field has not yet developed consistent frameworks for distinguishing acceptable assistance from unacceptable substitution, and that uncertainty creates anxiety for students and assessment challenges for teachers.

Synthesis: Risks as Design Constraints. These limitations and risks are not reasons to reject AI-assisted feedback. They are reasons to treat it as a conditional tool rather than a universal solution. The technical inaccuracies mean that AI feedback should be checked, not trusted blindly. The pedagogical risks mean that its use should be embedded in tasks that require critical engagement rather than passive acceptance. The ethical concerns mean that its deployment should be governed by explicit policies on access, privacy, and integrity. None of these conditions are met automatically. They require deliberate design-by teachers, by program administrators, and by the researchers who study how these tools function in real classrooms.

That design imperative leads directly to the next question: *given what we know and what we do not, what should an EFL teacher actually do?* Section 8 examines those gaps, and Section 9 addresses that practical question.

Research Gaps

Every literature review reaches a point where the evidence runs thin. The temptation is to treat that thinness as a minor inconvenience-a paragraph of boilerplate about "more research being needed" before moving to the conclusion. That would miss the point. In a field as young as AI-assisted feedback for EFL writing, the gaps are not incidental; they define the boundaries of what can responsibly be claimed. This section identifies the most significant holes in the current evidence base, drawing them from the specific limitations of the studies reviewed here rather than declaring them in advance.

- *The Absence of Longitudinal Evidence*. The most glaring gap is time. Every empirical study in the included pool operated within a compressed window-interventions lasting between four and ten weeks, with post-tests administered shortly after the final feedback cycle. That design can detect immediate effects: *Does the student's next draft improve? Does their self-efficacy rise?* What it cannot detect is whether those effects last, or whether they represent genuine learning rather than temporary performance enhancement. A student who scores higher on a post-test after eight weeks of ChatGPT feedback may have internalized new patterns, or they may simply have become adept at using the tool during the study period. Once the tool is removed, the gain may evaporate.

This matters because the central argument of this review-that accuracy and autonomy are interdependent-depends on development over time. Autonomy, in Little's (2007) sense, is not a state that appears after a single semester. It is a disposition that consolidates across months and years of practice. If we have no longitudinal studies tracking EFL writers across multiple semesters, we cannot say whether AI feedback accelerates that consolidation or merely simulates it. The few included studies that followed students beyond the immediate intervention relied on self-report measures administered once, which tells us what the student remembers feeling rather than what they can actually do. What the field needs are delayed post-tests-writing

samples collected three to six months after the intervention ends, under conditions where AI assistance is not available, to see whether the student can still produce accurate text independently. None of the studies reviewed here included that design element.

- *The Proficiency Blind Spot*. A second gap concerns who is being studied. The included empirical literature is heavily weighted toward intermediate and advanced university-level EFL learners-students who already possess enough English to produce coherent drafts and enough metacognitive awareness to engage with feedback, however imperfectly. Lower-proficiency learners, including many secondary school students and adult beginners, are underrepresented. Khodaie Alvar and Sahebkheir (2025) worked with pre-intermediate learners, but theirs was a small sample at a single institute, and their findings on teacher superiority may not generalize to other proficiency bands.

This is a significant omission because the theoretical predictions about AI feedback point in different directions for different proficiency levels. A beginner might benefit enormously from the explicitness and patience of an algorithm that never tires of explaining articles. Or they might be overwhelmed by the volume of suggestions, unable to distinguish critical errors from minor ones, and tempted to accept everything just to make the feedback stop. Without studies designed specifically for lower-proficiency populations-and without fine-grained analysis of how beginners process AI suggestions differently from advanced students-the field is prescribing a tool to learners who have never been properly studied.

- *The Mechanism Gap*. Perhaps the most consequential gap, because it sits at the heart of the autonomy question, is the lack of research on what happens between receiving AI feedback and producing a revision. We know that students accept or reject suggestions; we know far less about the cognitive and behavioral steps in between. *Do they read the explanations, or do they skip to the corrected version? Do they retain anything from the interaction, or is each feedback event functionally isolated?* The included studies provide snapshots of outcomes-accuracy scores, self-efficacy ratings, self-regulation questionnaires-but they rarely illuminate the process.

What is missing are studies that elicit the student's internal reasoning in real time: think-aloud protocols during revision, stimulated recall interviews immediately after, or eye-tracking that reveals which parts of the AI explanation the student actually reads. Without that micro-level evidence, claims about autonomy remain speculative. We can say that a student revised; we cannot say with confidence that they revised autonomously.

- *Comparative Tool Analysis*. Most of the included studies use ChatGPT, often without specifying which version or how it was prompted. That is understandable-ChatGPT was the most accessible generative AI tool during the 2023-2025 research boom-but it creates a conceptual problem. We do not actually know whether the findings transfer to other large language models. Claude, Gemini, and the various education-specific platforms now entering the market may handle error explanation differently, may be more or less prone to generating inaccurate explanations, and may present feedback in interfaces that encourage or discourage critical engagement. A study finding that "AI feedback improves accuracy:" is really a study finding that a particular AI, accessed in a particular way, at a particular moment in its development, improves accuracy. The generalization is assumed, not tested.

What is missing is comparative research: studies that pit different AI tools against each

other within the same population and task, holding constant the pedagogical conditions. Such studies would reveal whether the effects are robust across platforms or whether they are artifacts of a specific interface design.

- *Conceptual Fragmentation.* Beyond empirical design, there is a theoretical gap. The included studies draw on a scattered set of frameworks: self-regulated learning (Zhai, 2025; Nguyen & Doan Thi Hue, 2025), classroom observation and engagement (Han & Li, 2024), error coding and corpus analysis (Mizumoto et al., 2024), and student perspectives (Özdere, 2025). None of the studies attempt to integrate these perspectives into a coherent model of how AI feedback operates in EFL writing development. The result is a literature that produces isolated findings-accuracy here, self-efficacy there, engagement somewhere else-without a shared vocabulary for describing how these outcomes connect.

What is needed is a unified framework that can account for the full sequence: the student produces a draft, receives AI feedback, engages with that feedback cognitively and emotionally, revises the text, and carries the resulting knowledge forward into future writing. Such a framework would need to specify the mediating variables that determine whether the sequence leads to autonomous competence or to dependent performance. It would need to distinguish between behavioral engagement (time spent with feedback), cognitive engagement (depth of processing), and agentic engagement (initiative taken beyond the feedback provided). And it would need to place the AI tool within a sociocultural context-the classroom norms, the assessment pressures, the teacher's role-rather than treating it as a standalone intervention. The conceptual model proposed in Section 6.3 of this review is a preliminary step in that direction, but it remains theoretically derived rather than empirically validated.

- *Contextual and Cultural Variables.* Finally, there is a geographical and institutional bias. The majority of included empirical studies come from East Asian and Middle Eastern EFL Contexts-China, Iran, Jordan, Malaysia-with a smaller number from Turkey and Vietnam. That is not a flaw in those studies; it reflects where the research is being conducted. But it means that the evidence base may not generalize to other EFL environments where cultural attitudes toward technology, teacher authority, and error correction differ. In some educational cultures, questioning a teacher's feedback is considered disrespectful; in others, it is expected. In some contexts, AI is embraced as modern and efficient; in others, it is viewed with suspicion as a threat to human employment or academic tradition. These cultural variables shape how students interact with AI feedback, yet they are rarely treated as variables at all. The field needs more research from underrepresented EFL contexts-Central Asia, sub-Saharan Africa, South Asia, Eastern Europe-to test whether the patterns observed in China or Iran hold elsewhere.
- *Synthesis: Gaps as a Research Agenda.* These gaps are not a reason to dismiss the existing literature. The included studies have established that AI feedback can improve surface-level accuracy under certain conditions, that it can support engagement and self-regulation when pedagogically scaffolded, and that it carries real risks of dependency when students engage uncritically. Those are genuine contributions. But they are contributions to a preliminary map, not a finished one. The gaps identify where the next generation of research needs to go: toward longer timeframes, lower proficiency levels, micro-level process studies, comparative tool analysis, integrated theoretical models, and

broader geographical representation. Until that research is done, the most responsible stance is cautious optimism-recognizing the potential while refusing to overstate the evidence.

Pedagogical Implications

A literature review that ends with "more research is needed" has failed its most useful function. Teachers cannot wait for the perfect study; they have classes next week. This section translates the evidence reviewed so far into concrete, defensible practices-not as prescriptions from on high, but as working principles that can be adapted to different contexts, proficiency levels, and institutional constraints. Every recommendation is grounded in the included studies and theoretical frameworks discussed above.

1. *For Teachers: Positioning AI as a Resource, Not an Authority*. The single most important decision a teacher makes about AI feedback is not whether to allow it, but how to frame its role. If students perceive the AI as an infallible corrector-a digital teacher who never makes mistakes-they will treat it as an authority to be obeyed rather than a source to be interrogated. That perception needs to be challenged explicitly, early, and repeatedly.

One practical way to do this is to make error identification a two-stage process. Before students submit a draft to any AI tool, they mark their own text for what they suspect might be wrong. They write a brief note in the margin: "Not sure about this article" or "I think this transition is weak." Only then do they consult the AI. When the AI feedback arrives, the student has a prior hypothesis to test against. If the AI confirms their suspicion, the learning is reinforced because the external feedback validates an internal judgment. If the AI contradicts their suspicion, they have a specific question to investigate rather than a generic correction to accept. That simple sequence-predict, then verify-transforms the student from a passive recipient into an active evaluator. It is derived directly from the self-assessment principle in Nicol and Macfarlane-Dick's (2006) framework, but it uses the AI as the occasion for self-assessment rather than replacing it.

Teachers should also be transparent about the AI's limitations. Mizumoto, Eguchi, and Oyama (2024) demonstrated that ChatGPT can generate plausible-sounding explanations that do not always hold up under scrutiny. Showing students, a real example of such an inconsistency-a confident but questionable suggestion about an article rule, for instance-does not necessarily undermine the tool's credibility. Instead, it may build the student's critical immunity. A student who knows the AI can be wrong is a student who might pause before accepting every suggestion. That pause is where autonomy begins.

The hybrid approach described by Han and Li (2024) offers a broader structural model. In their classrooms, the AI generated initial comments on surface-level issues, which freed the teacher to focus on higher-order concerns: argumentation, organization, voice, rhetorical awareness. The teacher did not ignore the AI's output; they curated it. In class, they projected representative AI comments and asked the group to evaluate them: "This suggestion changes the meaning. How would you rewrite it yourself?" That public discussion turned individual feedback into a collective learning event. It also signaled to students that the AI is one voice in a larger conversation, not the final word.

2. *For Curriculum Designers: Building AI Literacy into Writing Instruction.* Curriculum designers face a different challenge. They are not managing a single classroom but a program, and programs tend to standardize before they think. The risk is that AI feedback gets bolted onto an existing syllabus as a mandatory tool without any reconsideration of what the syllabus is trying to teach. A writing course that previously aimed to develop

independent writers cannot simply add an AI checker and assume the aim remains unchanged.

What appears to be needed is an explicit scaffold that gradually transfers responsibility from the teacher to the student to the student's own judgment. In the first semester, the teacher might provide most of the feedback while introducing the AI as a supplementary resource for proofreading. In the second semester, the balance shifts: the AI handles surface errors, the teacher conferences on argument and structure, and the student begins writing reflective memos explaining their revision choices. By the third semester, the student is expected to use AI feedback independently, but with a documented protocol-the sequence this review has emphasized: Generate, Evaluate, Explain, Revise, Reflect. The "Reflect' stage is particularly important at this level. Students should be asked, at regular intervals, to identify their own recurring error patterns based on the AI feedback they have received over the term. That metacognitive task forces them to step back from individual corrections and look for systematic gaps in their knowledge, which is precisely the kind of self-directed analysis that Little (2007) associates with genuine autonomy.

Assessment needs to change alongside instruction. If students are graded solely on the quality of their final draft, they have every incentive to let the AI polish the surface as much as possible. If they are graded partly on the process-on the reflective memo, on the quality of their evaluation of AI suggestions, on their ability to explain why they rejected a particular correction-the incentive shifts. The grade rewards the writer, not just the text. That alignment between assessment and learning objectives is basic educational design, but it is easily forgotten when a new tool creates the illusion that the old objectives have already been met.

3. *For Learners: Developing a Personal Feedback Protocol.* Ultimately, the student is the one who must manage the relationship with the AI. Teachers and curricula can create the conditions, but the daily decision to accept, reject, or investigate a suggestion belongs to the learner. That is why students need a personal protocol-a set of questions they ask themselves every time they receive AI feedback.

   The protocol might look like this:

   - **Does this suggestion change my meaning?** If yes, I should reject it and try my own revision, even if it is less fluent.
   - **Can I explain why the AI thinks this is wrong?** If I do not understand the rule behind the correction, I need to look it up or ask, not just accept.
   - **Have I seen this error before?** If this is the third time the AI has flagged my article usage, I should study articles systematically rather than fixing them one by one.
   - **Would I be able to do this without the AI?** If the answer is consistently no, I am not learning; I am outsourcing.

   That last question is the hardest because it requires honesty that the tool itself does not enforce. But it is the question that separates the two pathways described in Section 6. A student who asks it regularly is on the path toward autonomy. A student who never asks it is on the path toward dependence, regardless of how accurate their final drafts become.

4. *Synthesis: The Human-Centred Classroom.* UNESCO's guidance (2023, 2024) insists that AI in education should be human-centred-supporting human judgment rather than displacing it. The pedagogical implications reviewed here align with that principle at every level. For teachers, it means maintaining the relational and interpretive roles that algorithms cannot perform. For curriculum designers, it means structuring progression so that AI use develops rather than replaces student capacity. For learners, it means cultivating the critical stance that turns a tool into a scaffold rather than a crutch.

The goal is not to produce writers who can generate flawless text with AI assistance. It is to produce writers who can eventually generate accurate, meaningful text with or without it-writers who have internalized the standards that the AI once supplied externally. That is a slower, more demanding goal than simple error correction. But it is the only one consistent with what EFL writing instruction has always claimed to value: not just correct English, but competent writers.

Conclusion

This review set out to examine what existing peer-reviewed research suggests about AI-assisted feedback in EFL writing, with particular attention to two outcomes that are often discussed separately but are in fact deeply intertwined: writing accuracy and learner autonomy. The evidence, considered as a whole, supports a nuanced picture rather than a simple verdict.

On accuracy, the included literature is cautiously positive within clear boundaries. AI-assisted feedback-whether from early automated systems or more recent large language models-appears to identify and help correct surface-level errors with a consistency that can complement teacher feedback, particularly in contexts where teacher workload limits the timeliness or detail of response. Students who receive such feedback typically produce cleaner drafts, at least in the short term. But the improvement is uneven. Teacher feedback may still outperform AI on certain accuracy measures, particularly for higher-order concerns like coherence and voice. AI tools struggle with context-dependent judgments, occasionally generate inaccurate explanations, and vary in reliability depending on how they are prompted. They are competent proofreaders and incomplete writing coaches.

On autonomy, the picture is more ambivalent and more interesting. AI feedback can support self-regulated learning. It can reduce anxiety, increase engagement, and create opportunities for students to monitor their own output. Studies by Nguyen and Doan Thi Hue (2025), Yang, Hussin, and Hashim (2025), and Zhai (2025) have documented these positive effects. But they are not automatic. The same tool that scaffolds independent revision can also produce passive dependence when students accept every suggestion without evaluation. The difference lies not in the algorithm but in the pedagogy surrounding it. When AI feedback is embedded in tasks that require critical judgment-when students must explain, justify, and sometimes reject the advice they receive-autonomy tends to grow alongside accuracy. When the feedback flows directly into the final text without that mediating cognitive work, accuracy may rise on paper while the writer's underlying capacity stagnates.

That interdependence is the central argument of this review. Accuracy without autonomy is fragile; it depends on the continued presence of the tool. Autonomy without accuracy is unfocused; it risks self-direction without the linguistic resources to make that direction effective. The goal of EFL writing instruction should be to cultivate both together, through pedagogical designs that treat AI feedback as a resource to be engaged with critically rather than as an authority to be obeyed. The hybrid models described by Han and Li (2024) point toward what such designs might look like: AI and teacher working in complementary roles, with the algorithm handling scale and immediacy while the human maintains judgment, relationship, and the higher-order dimensions of writing that machines do not yet perceive.

For practice, this means resisting the temptation to adopt AI feedback as a cost-saving convenience or as a substitute for teacher attention. It means teaching students to evaluate AI suggestions explicitly, to recognize that the tool can be wrong, and to build personal protocols for revision. It means aligning assessment with process rather than product alone, so that the grade rewards the development of the writer, not just the polish

of the text. And it means attending to the ethical dimensions that UNESCO (2023, 2024) has emphasized: protecting learner agency, ensuring equitable access, and maintaining human oversight in educational decisions that affect young people's relationship with language.

For research, the agenda is clear. The field needs longitudinal studies that track writers across semesters, not weeks. It needs more work with lower-proficiency learners, who may be the population most affected by AI feedback but the least studied. It needs comparative analyses of different AI tools, integrated theoretical frameworks that connect accuracy and autonomy, and process-oriented methods that reveal what students actually think and do when they encounter algorithmic advice. Until that research is available, the most defensible stance is provisional: AI-assisted feedback is a powerful tool with significant potential and significant risks, and its value in any given classroom will depend less on the sophistication of the model than on the wisdom of the teacher who deploys it.

The question is no longer whether AI will shape EFL writing instruction. It already has. The question is whether we can shape its use so that it serves the development of writers who are accurate in their language and autonomous in their learning-writers who can eventually put the tool aside not because they have outgrown the need for feedback, but because they have grown into the capacity to give it to themselves.

References:

1. Alnemrat, A., Aldamen, H., Almashour, M., Al-Deaibes, M., & AlSharefeen, R. (2025). AI vs. teacher feedback on EFL argumentative writing: A quantitative study. *Frontiers in Education, 10*, Article 1614673. https://doi.org/10.3389/feduc.2025.1614673
2. Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
3. Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge University Press.
4. Han, Y., & Li, X. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System, 123*, Article 103354. https://doi.org/10.1016/j.system.2024.103354
5. Khodaie Alvar, F., & Sahebkheir, F. (2025). The effect of AI-assisted feedback vs. teacher corrective feedback on Iranian EFL learners' grammatical accuracy. *Journal of Applied Linguistics Studies, 5*(2), 239–253. https://doi.org/10.57647/jals.2025.0502.08
6. Little, D. (2007). Language learner autonomy: Some fundamental considerations revisited. *Innovation in Language Learning and Teaching, 1*(1), 14–29. https://doi.org/10.2167/illt039.0
7. Little, D., Dam, L., & Legenhausen, L. (2017). *Language learner autonomy: Theory, practice and research*. Multilingual Matters.
8. Mizumoto, A., Eguchi, M., & Oyama, A. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics, 3*(2), Article 100131. https://doi.org/10.1016/j.rmal.2024.100131
9. Nguyen, Q., & Doan Thi Hue, D. (2025). ChatGPT's effects on EFL learners' autonomy and self-efficacy in self-regulated learning contexts. *The JALT CALL Journal, 21*(2), Article 102968. https://doi.org/10.29140/jaltcall.v21n2.102968

10. Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090
11. Özdere, M. (2025). AI in academic writing: Assessing the effectiveness, grading consistency, and student perspectives of ChatGPT and You.com for EFL students. *International Journal of Technology in Education, 8*(1), 123–154. https://doi.org/10.46328/ijte.1001
12. UNESCO. (2023). *Guidance for generative AI in education and research*. https://doi.org/10.54675/QTMR8092
13. UNESCO. (2024). *AI competency frameworks for students and teachers*.
14. Wang, D. (2024). Teacher- versus AI-generated (Poe application) corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *International Review of Research in Open and Distributed Learning, 25*(3). https://doi.org/10.19173/irrodl.v25i3.7646
15. Yang, Y., Hussin, S., & Hashim, H. (2025). Revising with intelligence: ChatGPT feedback and its impact on EFL students' revision and self-efficacy. *Forum for Linguistic Studies, 7*(7), 352–367. https://doi.org/10.30564/fls.v7i7.9845
16. Zhai, Y. (2025). Automated written corrective feedback: Fostering writing accuracy, self-regulation, and motivational dynamics in EFL learners. *BMC Psychology, 13*, Article 402. https://doi.org/10.1186/s40359-025-03765-y