
| RESEARCH ARTICLE

Generative AI-Assisted Academic Writing in Higher Education EFL Contexts: A Critical Review of Learning Outcomes, Self-Regulated Learning, and Ethical Governance

Mohan Li

Undergraduate Student, Foreign Studies College, Hunan Normal University, Changsha, China

Corresponding Author: Mohan Li, **E-mail:** 1311089843@qq.com

| ABSTRACT

GenAI's rapid proliferation has reshaped EFL academic writing, yet research remains fragmented. This review synthesizes 24 publications (2022–2026) on learning outcomes, self-regulated learning (SRL), and ethical governance. Three findings emerge: (1) positive effects on surface-level writing are confirmed but moderated by intervention duration, task complexity, learner proficiency, and evaluation methods; (2) GenAI's impact on SRL depends on its role as a scaffold versus a substitute for learner agency, as illuminated by TAM, SCT, and SRL; (3) ethical governance is shifting from plagiarism policing to proactive AI literacy. Gaps persist in longitudinal and process-oriented designs, representation of underserved groups, and theory-driven pedagogy. The review argues that GenAI's value cannot be judged by writing products alone but requires understanding the interplay among cognition, autonomy, and regulation. An ecologically-informed research agenda is proposed.

| KEYWORDS

Generative AI; academic writing; EFL; self-regulated learning; academic integrity

| ARTICLE INFORMATION

ACCEPTED: 30 July 2026

PUBLISHED: 05 August 2026

DOI: 10.32996/jeltal.2026.8.9.1

1. Introduction

The proliferation of generative artificial intelligence (GenAI), particularly large language models such as ChatGPT, has precipitated a paradigm shift in higher education, with academic writing emerging as one of the most profoundly affected domains (Bozkurt, 2024). Since the public release of ChatGPT in late 2022, an accelerating body of scholarship has documented the rapid infiltration of these tools into university students' writing practices, signalling an urgent need for pedagogical and institutional responses (Dong, 2026). The transformative potential of GenAI—capable of generating coherent prose, revising syntax, and even structuring arguments—has positioned it as both a promising pedagogical asset and a formidable challenge to established academic norms. This convergence of technological accessibility and educational necessity establishes the timeliness and scholarly relevance of a critical examination into how GenAI is reshaping EFL (English as a Foreign Language) academic writing at the tertiary level.

Despite the burgeoning interest, the extant literature is characterised by profound fragmentation and polarised stances. While a growing number of empirical studies report immediate improvements in surface-level linguistic features such as grammatical accuracy and lexical diversity following AI intervention (Fan & Gao, 2026; Wang & Ren, 2024). Conversely, a number of researchers caution that the observed improvements might be offset by diminished higher-order reasoning and thoughtful analytical processing (Zhang & Crosthwaite, 2026; Zizka, 2025). Although some researchers champion GenAI as a revolutionary scaffold capable of democratising access to academic discourse (Cheng et al., 2025), others contend that its unregulated use fosters dependency, erodes writer autonomy, and compromises academic integrity (Alsadoon, 2026). Crucially, these debates are frequently conducted on disparate planes—some focused on linguistic products, others on pedagogical processes, and still others on ethical boundaries—resulting in a scholarly landscape where divergent conclusions are difficult to reconcile. This fragmentation suggests that existing research has predominantly conceived of GenAI as a discrete technological intervention,

Copyright: © 2026 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (<https://creativecommons.org/licenses/by/4.0/>). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

thereby failing to adequately theorise it as a socially and cognitively mediated tool embedded within authentic educational ecologies. Hence, a pressing scholarly priority now is to transcend oversimplified affirmative-or-negative positions and instead develop a refined understanding of the underlying mechanisms that account for heterogeneous effects. At a deeper level, the persistent controversies within the literature suggest that the fundamental inquiry might have been misformulated. Instead of focusing on the inherent virtue or vice of generative AI, a more fruitful direction is to examine the pedagogical and institutional circumstances under which varying academic performances arise.

Recent review studies, while valuable, have not adequately bridged these divides. Existing literature reviews have largely treated GenAI either as a broad educational tool (Dong, 2026) or else have restricted their attention to concerns about academic dishonesty, frequently framing ethics as a standalone regulatory concern that remains disconnected from actual teaching contexts (Yan et al., 2025). Conspicuously absent from the current review landscape is an integrative framework that synthesises learning outcomes (the product), self-regulatory processes (the mechanism), and ethical governance (the contextual boundary) within a unified critical lens. Moreover, the majority of existing reviews draw almost exclusively from empirical journal articles, with scant attention to influential policy documents issued by supranational bodies such as UNESCO (2025), the OECD (2026), and the European Commission (2026). These documents, however, fundamentally shape the institutional and ideological contexts within which AI ethics are debated and implemented; their omission renders much of the ethical discourse in the literature disconnected from macro-level policy realities. This dual gap—theoretically fragmented and empirically insular—constitutes the precise niche that the present review seeks to address. It is important to recognise that this fragmentation is not merely an organisational issue. Instead, it reflects a deeper tendency to examine technological, pedagogical, and ethical questions in isolation, thereby overlooking their dynamic interrelationships.

Accordingly, this review is guided by three research questions:

- (1) What are the established consensuses and salient disagreements in existing research concerning the effects of GenAI on learning outcomes in EFL higher education academic writing?
- (2) In what ways does GenAI transform the self-regulated writing behaviors of learners, and under which circumstances does it act as a supportive structure rather than a source of excessive reliance?
- (3) Which methodological shortcomings and ethical oversight issues should future research prioritize as the most urgent?

This review maintains that the instructional benefit of generative AI for L2 English academic composition cannot be captured solely by examining final written products. Rather, it should be understood as the result of dynamic, multi-level interactions among learning processes (cognitive), learner agency (behavioral), and institutional governance (contextual).

To theorise these multi-level dynamics, the present review adopts an integrated framework that combines the Technology Acceptance Model (TAM; Davis, 1986) for understanding initial AI adoption, Sociocultural Theory (SCT; Vygotsky, 1978) for conceptualising AI as a mediational cultural tool, Self-Regulated Learning (SRL; Zimmerman, 2000) for analysing whether AI use fosters or displaces learner agency, and an ecological perspective (Bronfenbrenner, 1979) for situating individual behaviour within institutional and policy environments. Rather than treating these lenses as separate, the framework posits that their interaction—from technology acceptance through mediated practice and self-regulation, to governance constraints—determines educational outcomes. By incorporating these theoretical perspectives from the very beginning, the present review seeks to resolve the disjointedness that earlier studies have exhibited. In particular, the ecological lens anticipates the later discussion of ethical governance (Section 5), where we will argue that sustainable AI integration requires coordinated action across classroom, institutional, and supranational levels. Likewise, the framework foreshadows the proposal for AI literacy as a core educational outcome, which emerges as a necessary complement to punitive integrity enforcement.

By systematically addressing the three research questions through this integrative theoretical lens, this review aims to provide a coherent roadmap for future empirical inquiry and pedagogical innovation, moving the field beyond isolated debates towards a more holistic and theoretically grounded understanding of GenAI's role in tertiary EFL writing contexts.

2. Review Method

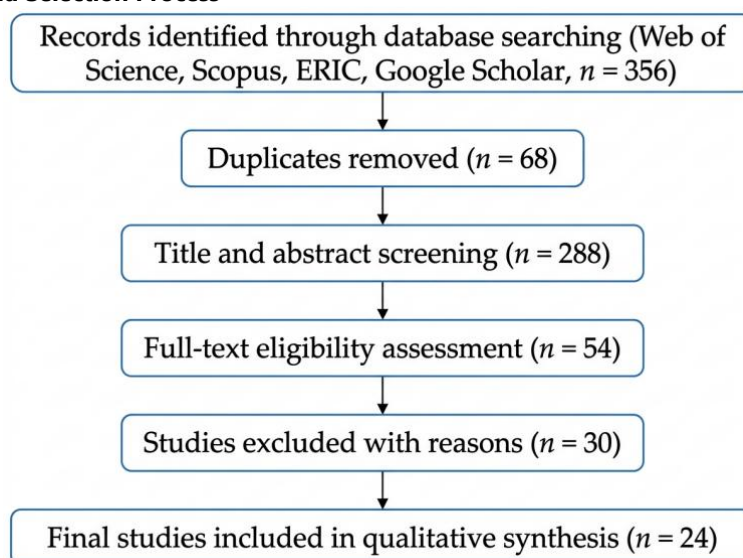
To ensure transparency and replicability, this review adopted a systematic literature search strategy. Four databases were searched: Web of Science Core Collection, Scopus, ERIC (ProQuest), and Google Scholar (the latter employed for supplementary searches and grey literature). The search period was delimited from November 2022 (the public release of ChatGPT) to June 2026, capturing both the initial surge of scholarship following the technology's launch and more recent empirical developments (Bozkurt, 2024; Dong, 2026). The Boolean search string combined terms across four conceptual dimensions: technology ("generative AI" OR "ChatGPT" OR "large language model"), writing task ("academic writing" OR "essay writing" OR "research writing"), learner population ("EFL" OR "ESL" OR "L2 writing" OR "second language"), and educational context ("higher education" OR "university" OR "college").

The following prerequisites governed study inclusion: (a) the work must have appeared as an English-language, peer-reviewed journal paper or as a scholarly book chapter published by an authoritative academic outlet; (b) focused on higher education EFL/ESL learners as participants (or included such learners in the sample); (c) specifically addressed the use of generative AI tools for academic writing tasks; and (d) constituted empirical research, systematic reviews, or meta-analyses. To ensure breadth and depth, the review also incorporated conceptual papers and policy framework documents from authoritative supranational bodies, including the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2025), the Organisation for Economic Co-operation and Development (OECD, 2026), and the European Commission (2026), given their constitutive role in shaping the institutional and ideological contexts of AI ethics discourse (Yan et al., 2025). Studies were excluded if they were: (a) conference proceedings, editorials, book reviews, news articles, or non-academic opinion pieces; (b) confined to K-12 settings, first-language English writing, or non-writing language learning domains; or (c) purely technical system descriptions lacking pedagogical or learning analysis.

Following the application of these criteria, 24 core publications were identified for in-depth review and integration. The study selection was carried out in successive stages: screening by title, evaluation of abstracts, and examination of full texts. Additionally, we scanned the reference lists of the reviews that met inclusion criteria to identify any further pertinent studies. In interpreting and synthesising the findings, this review did not treat all included literature as methodologically equivalent; particular attention was accorded to quality indicators including research design, sample size, intervention duration, and analytical approach, with special emphasis on longitudinal studies, mixed-methods designs, and the inclusion of process-oriented data (Cheng et al., 2025; Du & Chen, 2026; Konradt, 2025). This methodological vigilance ensures that the subsequent synthesis reflects not merely thematic prevalence but evidentiary rigour. Rather than treating all studies as equally informative, this review gives greater interpretive weight to research employing longitudinal designs, authentic classroom contexts, and process-oriented methodologies, as these provide stronger evidence for understanding how AI shapes learning over time.

The review further incorporates an ecological perspective (Bronfenbrenner, 1979) to frame the subsequent governance analysis, recognising that learner behaviour is simultaneously shaped by micro-, meso-, and macro-level factors—a lens that will be operationalised in Section 5 when examining institutional policies and supranational frameworks.

Figure 1
Literature Identification and Selection Process



Note. The review followed a transparent literature selection procedure adapted from PRISMA.

Figure 1 illustrates the transparent literature screening procedure adopted in this review. Compared with narrative reviews, this process improves methodological transparency and strengthens the credibility of the synthesized evidence.

3. Learning Outcomes: What Does the Evidence Actually Show?

3.1 Positive Findings: Surface-Level Improvements

A remarkable consensus is observable in research publications with respect to GenAI's potential for augmenting the formal aspects of second-language academic composition. Multiple empirical studies have documented statistically significant improvements in grammatical accuracy, lexical diversity, and syntactic complexity following AI-assisted writing interventions (Fan & Gao, 2026; Wang & Ren, 2024). In a cross-sectional study examining AI-mediated metacognitive interventions, Fan and Gao

(2026) found that EFL learners demonstrated measurable gains in textual accuracy and lexical sophistication when engaging with generative AI tools for writing revision. Beyond formal linguistic aspects, GenAI is recognized for boosting both writing fluency and efficiency. Evidence from Wang and Ren (2024) indicates that undergraduates in a Hong Kong linguistics course enhanced their writing across several language dimensions by using ChatGPT in drafting. Meanwhile, Cheng et al. (2025) reported that ESL graduate students harnessed ChatGPT to combat writer's block, craft content outlines, and hasten draft preparation, which consequently elevated their productivity. Lee et al. (2025) similarly observed that learners valued the immediate availability of AI-generated suggestions, which reduced cognitive load during the initial stages of writing. Yet a critical question persists: whether these gains extend beyond surface characteristics to substantively improve deep-structural features such as argumentation coherence, logical development, and critical analytical depth—a distinction that remains insufficiently addressed in the current evidence base. At first glance, these consistently positive findings appear to support the educational value of generative AI. However, upon more careful scrutiny, the majority of documented gains appear to be confined to superficial language features, as opposed to more profound aspects of scholarly reasoning.

3.2 Contradictory Findings and Moderating Variables

The apparent contradiction across studies should not be interpreted as evidence of inconsistency in the technology itself. Instead, it reflects substantial variation in research contexts, participant characteristics, instructional designs, and assessment criteria. An integrative review of the available research indicates that no fewer than five moderator factors exert systematic effects on both the extent and the polarity of observed outcomes. This holds true even though the results presented in Section 3.1 are broadly favorable; however, upon more careful inspection, substantial heterogeneity among studies becomes apparent, and the conflicting findings cannot be explained by methodological diversity in isolation. First, intervention duration emerges as a critical factor. Studies employing single-session designs tend to report pronounced short-term gains (e.g., Wang & Ren, 2024; Fan & Gao, 2026), whereas longer-term investigations extending across a full academic semester frequently detect a “novelty effect” attenuation, with initial improvements plateauing or even regressing as learners habituate to the technology (Konradt, 2025). Second, task complexity mediates outcomes considerably. GenAI demonstrates substantially stronger efficacy for low-complexity tasks such as sentence-level error correction or short text revision (Du & Chen, 2026) compared to high-complexity tasks requiring argument construction or knowledge transformation (Zizka, 2025). Existing studies often fail to distinguish between knowledge-telling and knowledge-transforming tasks, yet this very difference marks a critical divergence in intellectual requirements that has significant implications for cognitive load. Third, learner proficiency interacts with AI utility in non-linear ways. Lower-proficiency learners appear to benefit most from grammatical scaffolding and lexical suggestions (Fan & Gao, 2026), while advanced writers may find AI-generated revisions stylistically incongruent with their authorial voice, leading to resistance or selective adoption (Zhang & Crosthwaite, 2026). Fourth, evaluation methodology introduces systematic bias: studies relying on automated indices of linguistic complexity are more likely to report positive outcomes, whereas those employing human holistic ratings or qualitative content analysis expose more nuanced and often less uniformly favourable patterns (Dong, 2026). Finally, a largely unexamined confounding variable is prompt engineering proficiency—learners’ ability to craft effective instructions for AI. Among the various moderating factors, the duration of intervention stands out as perhaps the primary source of inconsistency, given that single-session designs have a systematic tendency to amplify measured outcomes via novelty-related biases. Furthermore, the majority of empirical investigations fail to account for participants’ digital proficiency, which means that some of the observed benefits could be attributed to technological familiarity rather than to the authentic instructional merit of AI. In combination, these confounded moderators render any universal claim about GenAI’s pedagogical effectiveness untenable.

Overall, these conditioning factors point to one essential insight: the pedagogical returns of generative AI cannot be dissociated from the specific learning environments in which it operates; they are inherently context-dependent. Consequently, future debates should shift from asking whether AI works to identifying when, for whom, and under what conditions it contributes to meaningful learning.

3.3 Methodological Critique of the “Effectiveness” Literature

Beyond the empirical inconsistencies identified above, the broader “effectiveness” literature is beset by fundamental methodological limitations that constrain the generalisability of its findings. Design constraints constitute the most pervasive weakness: the overwhelming majority of studies employ short-term quasi-experimental designs conducted in controlled, decontextualised settings, lacking the ecological validity and temporal depth necessary to capture authentic classroom writing processes (van Niekerk et al., 2025). Longitudinal designs spanning eight weeks or more remain exceptionally rare, with Konradt (2025) representing one of the few exceptions. Sampling bias further compounds this issue: convenience samples drawn predominantly from East Asian universities—particularly in China and South Korea—dominate the empirical landscape (Dong, 2026; Du & Chen, 2026; Han & Li, 2026), raising legitimate questions about cross-cultural generalisability and the applicability of findings to diverse educational contexts. Measurement limitations are equally concerning: research in this area demonstrates a marked dependence on self-reported measures, such as learner opinions and attitude questionnaires (Alsadoon, 2026; Lin et al.,

2025), whereas only a small number of investigations have employed direct observations of writing behavior, screen-capture protocols, or analyses across multiple drafts, which might reveal the temporal dynamics of how students make use of AI. This constellation of methodological vulnerabilities—short timeframes, unrepresentative samples, and narrow measurement approaches—renders much of the existing evidence base provisional and contextually bounded. Taken together, the above constraints indicate that the equivocal nature of effectiveness-oriented inquiries may be more attributable to the discipline's entrenched habit of treating AI as a standalone “technological fix” rather than as a socio-cognitive instrument grounded in real scholarly practice, than to any intrinsic didactic uncertainty of generative AI itself. Resolving this impasse requires shifting analytical focus from decontextualised product measures to the processes through which learners engage with AI—a move that directs attention inevitably toward self-regulated learning. Perhaps the most overlooked limitation is that many effectiveness studies continue to conceptualise AI as an isolated technological intervention rather than as a socially situated cognitive tool operating within authentic educational environments.

4. Self-Regulated Learning: Scaffolding or Dependency?

Moving beyond writing outcomes, the more fundamental issue concerns the learning processes through which those outcomes are achieved. Understanding these processes is essential because similar writing products may emerge from entirely different patterns of learner engagement. The divergent findings reviewed in the preceding chapter point to a fundamental inadequacy in treating AI as a neutral technology whose effects can be assessed independently of how learners engage with it. This chapter argues that the educational value of GenAI in EFL academic writing is contingent upon the role it assumes within learners' self-regulated learning (SRL) systems—whether it functions as a scaffold that enhances metacognitive control or as a substitute that displaces cognitive responsibility. Drawing on sociocultural theory (SCT) and Zimmerman's (2000) cyclical model of SRL, this chapter unpacks the conditions under which AI-mediated writing support either promotes or undermines sustainable learner development.

4.1 AI as a Mediation Tool: The SCT Perspective

From a Vygotskian perspective, human cognition is fundamentally mediated by cultural tools—language, symbols, and technologies that shape and extend mental processes (Vygotsky, 1978). Within this perspective, generative AI can be regarded as an “intelligent cultural tool” that may promote writing development when it is employed as a mediating resource rather than as a finished product. Complementing this view, Vygotsky's (1978) ZPD construct—the interval separating unassisted performance from guided achievement—provides a particularly productive heuristic for interpreting the pedagogical affordances of AI in educational settings. When learners engage with AI to receive just-in-time feedback on syntactic or lexical choices, the technology functions as a responsive interlocutor that provides support calibrated to their emerging competence (Fan & Gao, 2026; Wang & Ren, 2024). Similarly, when learners employ AI for dialogic brainstorming—posing questions, soliciting alternative formulations, or testing argumentative lines—they are effectively engaging with a more knowledgeable “other” that expands their thinking beyond current limits (Lee et al., 2025). The critical distinction, however, lies in learner agency: when individuals actively mobilise AI to regulate their own cognitive processes—checking logical coherence, generating counterarguments, or exploring organisational options—the tool becomes a genuine mediational means that enhances, rather than supplants, writerly competence (Cheng et al., 2025; Du & Chen, 2026). This agentic deployment of AI aligns with Bandura's (1989, 2001) conception of human agency, wherein individuals exercise intentional control over their actions and environments rather than passively responding to external stimuli. From this perspective, generative AI should not simply be viewed as an external support tool. Its educational value ultimately depends on whether learners actively appropriate the mediation it provides rather than passively relying upon its outputs.

4.2 The Danger of Dependency: When Mediation Becomes Replacement

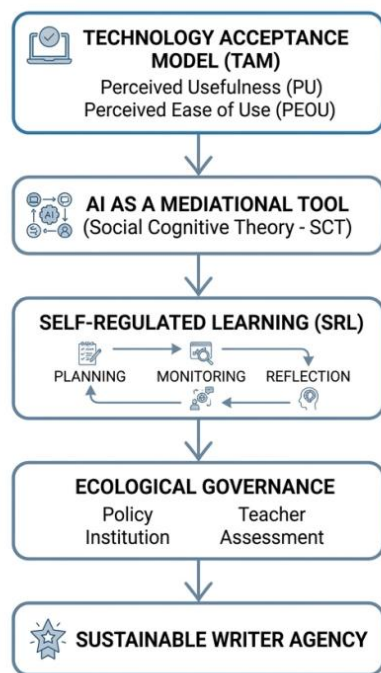
The same technological affordances that enable scaffolding, however, can precipitate dependency when learners cede cognitive responsibility to AI. Zimmerman's (2000) cyclical model of SRL—comprising forethought, performance, and self-reflection phases—provides a diagnostic framework for identifying such displacement. By establishing objectives, drawing on existing knowledge, and organizing their reasoning, self-regulated writers in the planning stage deliberately arrange their composing process. When learners uncritically adopt AI-generated outlines without engaging in their own organisational reasoning, they substitute external guidance for internal strategic planning, thereby short-circuiting a crucial phase of skill development (Konradt, 2025). During the performance phase, where self-regulated writers monitor their execution and deploy task strategies, learners who accept AI's lexical or syntactic revisions without evaluative scrutiny bypass the metacognitive monitoring essential for long-term linguistic development (Zhang & Crosthwaite, 2026). The self-reflection phase, perhaps most critically compromised, involves self-evaluation and adaptive adjustment—processes that are attenuated when learners reduce AI engagement to binary “accept/reject” decisions devoid of reflective analysis (Zizka, 2025). This pattern of uncritical acceptance, identified across multiple studies (Alsadoon, 2026; Lin et al., 2025), exemplifies the problematic transition from “thinking with AI” to “letting AI think for the learner”—a distinction that fundamentally determines whether AI use fosters or undermines sustainable writing competence. The empirical evidence further suggests that such dependency may lead to content

homogenisation and attenuated critical thinking, as learners increasingly rely on AI-generated formulations rather than developing their own argumentative voice (Zhang & Crosthwaite, 2026; Zizka, 2025). It could be argued that the primary pedagogical concern, then, is not to forbid the use of AI itself, but rather to halt the progressive handover of mental accountability from students to automated systems.

4.3 Theoretical Integration: From TAM to Sustainable SRL

The contrasting outcomes delineated above can be systematically understood through a tripartite theoretical framework that traces the trajectory from initial adoption to sustainable learning. First, the Technology Acceptance Model (TAM) explains learners’ initial engagement with GenAI: perceived usefulness and perceived ease of use predict adoption intention, with learners drawn to AI’s capacity to reduce writing anxiety and expedite task completion (Davis, 1986; Lin et al., 2025). Second, sociocultural theory accounts for the mechanism through which AI, as a cultural tool, mediates writing activity: its effects are not intrinsic but emerge through the quality of learner-tool interaction (Vygotsky, 1978; Lee et al., 2025). Third, self-regulated learning theory determines the ultimate outcome: whether this external mediation becomes internalised as durable self-regulatory capability, or results in alienation—a condition where the tool’s affordances paradoxically undermine learner agency and metacognitive development (Zimmerman, 2000; Bandura, 2001). From this overarching perspective, the educational priority is not about choosing to adopt or prohibit AI; rather, it consists of devising instructional strategies that steer students from an initial adoption motivated by the technology acceptance model toward an internalisation process that strengthens self-regulated learning. Nevertheless, this TAM-SCT-SRL integration remains predominantly focused on individual cognitive and behavioural dimensions; it does not yet adequately explain why identical AI-use behaviours yield divergent outcomes across different course policies, instructor beliefs, or assessment cultures. This lacuna—the absence of a robust account connecting macro-level institutional structures with micro-level learning processes—points toward the necessity of an ecological perspective that situates self-regulation within broader environmental systems, a theme to which this review now turns. This theoretical explanation also helps resolve the methodological inconsistencies identified earlier in Section 3. More importantly, this integrated framework also highlights an issue that individual-level theories alone cannot adequately explain: why identical AI practices may produce substantially different outcomes under different institutional policies, assessment cultures, and classroom environments.

Figure 2
Integrated Theoretical Framework for AI-Assisted Academic Writing



Note. The framework is synthesized by the present review based on TAM, SCT, SRL, and ecological perspectives. As illustrated in Figure 2, the educational value of generative AI cannot be understood through a single theoretical perspective. Instead, technology acceptance initiates AI use, social mediation shapes interaction, self-regulated learning determines internalization, and ecological governance provides contextual constraints. Together, these dimensions constitute an integrated explanation of sustainable writer agency. Unlike earlier work that treated TAM, SCT, or SRL as standalone constructs, the current

framework illuminates how technology acceptance, socially mediated learning, student initiative, and policy contexts are dynamically intertwined.

5. Ethical Governance: Beyond Policing Plagiarism

This observation suggests that ethical governance should not be understood as a peripheral concern introduced after technology adoption. Instead, governance constitutes an integral component of the educational ecosystem within which AI-mediated learning takes place. The preceding analysis reveals a fundamental paradox: whether AI facilitates learning depends not only on learners' self-regulatory capacity but also on the multi-layered environmental systems that shape educational activity. This chapter therefore adopts an ecological perspective (Bronfenbrenner, 1979) to critically examine ethical governance, reframing ethics not as an isolated "anti-cheating" concern but as a continuum spanning classroom assessment practices to national policy frameworks. Such an ecological orientation recognises that learner behaviour is simultaneously shaped by microsystem-level institutional policies, mesosystem-level assessment cultures, and macrosystem-level national and supranational policy regimes (UNESCO, 2025; OECD, 2026).

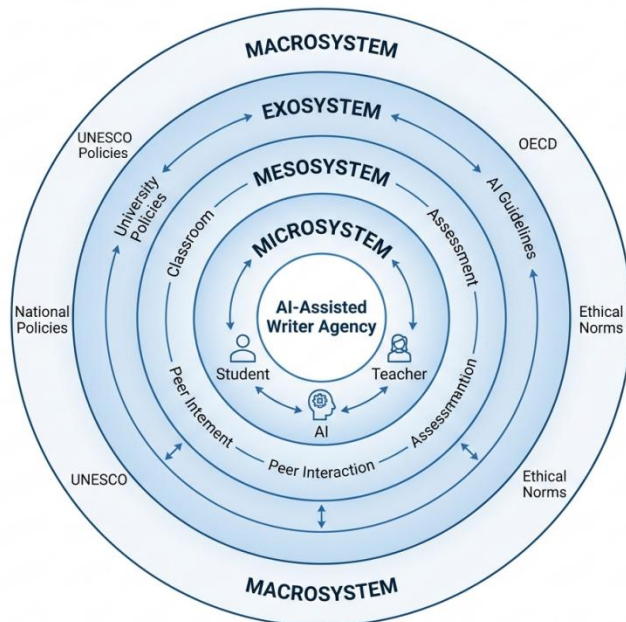
5.1 The Blurred Lines of Academic Integrity

Traditional conceptions of academic integrity, predicated on clear distinctions between original and borrowed work, are fundamentally unsettled by generative AI. The empirical literature documents pervasive ambiguity in how both students and educators delineate acceptable from unacceptable AI use. Alsadoon (2026) found a marked discrepancy between EFL students' perceptions of academic integrity and their actual ChatGPT usage practices, suggesting that normative frameworks have not kept pace with technological affordances. Three particularly vexing boundary cases recur across studies: (a) whether AI-mediated spelling and grammar checking constitutes editing or unauthorised revision (van Niekerk et al., 2025); (b) whether AI-generated outlines represent legitimate scaffolding or unacceptable substitution of independent structuring (Cheng et al., 2025); and (c) whether using AI to reformulate one's own prose for formality constitutes stylistic enhancement or substantive alteration (Zizka, 2025). These ambiguities are compounded by profound institutional divergence: comparative analyses of university policies across North America, Europe, and East Asia reveal dramatically different stances, ranging from permissive integration to categorical prohibition, reflecting the absence of international consensus (Yan et al., 2025). At the institutional level, university policies constitute the microsystem immediately shaping student behaviour; at the supranational level, documents such as UNESCO's (2025) guidance for policymakers and the OECD's (2026) Digital Education Outlook reveal fundamental disagreements about whether AI should be accommodated or resisted—a macrosystem divergence that trickles down to institutional practice. This lack of global alignment in policy means that practices deemed a breach of academic integrity in one educational system might be considered acceptable in another, which creates uncertainty for both students and teachers and simultaneously highlights the critical necessity of establishing coherent regulatory approaches.

5.2 The Crisis of Assessment Validity

Beyond definitional ambiguities, AI poses a more fundamental challenge to the validity of academic writing assessment. When students submit writing products substantially shaped by generative AI, the question arises: are we assessing students' writing competence or their ability to collaborate with AI? This question strikes at the construct validity of academic writing evaluation, destabilising the traditional mesosystem constituted by teachers, students, and assessment instruments (Zizka, 2025; Zhang & Crosthwaite, 2026). The assessment crisis thus represents a systemic conflict within the intermediate educational system, where the introduction of AI disrupts previously stable relationships among pedagogical actors. van Niekerk et al. (2025) argued compellingly that existing assessment paradigms, predicated on the assumption of unaided performance, are rendered obsolete when learners can delegate substantial cognitive labour to AI. In response, divergent proposals have emerged: some advocate a return to invigilated, in-class writing examinations as the only reliable means of assessing individual competence (Han & Li, 2026), while others contend that assessment must evolve to incorporate AI literacy as a legitimate outcome (Bozkurt, 2024; Konradt, 2025). Emerging assessment models—including process-oriented portfolios, reflective statements documenting AI use, and multi-draft submissions with revision justifications—have been proposed as alternatives (Du & Chen, 2026), yet their feasibility, reliability, and scalability remain empirically untested. This assessment impasse reflects a deeper crisis: the educational system has not yet determined what academic writing should measure in an AI-mediated era, a question with profound implications for curriculum design, teacher professional development, and student learning orientations. It may be more constructive to treat this situation not as a danger to academic evaluation, but as a stimulus for rethinking what universities genuinely ought to measure in the era of AI, rather than merely viewing it as a problem.

Figure 3
An Ecological Model of AI-Assisted Academic Writing



Note. Adapted and synthesized by the present review based on ecological perspectives on educational AI governance. Figure 3 conceptualizes AI-assisted academic writing as an ecological system in which learner agency emerges through continuous interactions among individual, classroom, institutional, and policy-level factors. This ecological perspective shifts the discussion beyond plagiarism prevention toward sustainable educational governance. This integrated framework extends previous conceptualisations by incorporating ecological governance into the existing TAM-SCT-SRL sequence.

5.3 From Academic Integrity to AI Literacy

Based on the preceding discussion, any penalizing and regulation-driven strategy for managing AI is essentially insufficient when it comes to resolving issues that are inherently instructional, rather than merely administrative. This review therefore argues that the path forward lies not in stricter enforcement but in more systematic education—specifically, the cultivation of AI literacy as a core educational outcome. AI literacy may be conceptualised as an integrated competence encompassing knowledge of AI capabilities and limitations, technical skills for effective human-AI collaboration, and ethical awareness for responsible usage (Bozkurt, 2024; European Commission, 2026). Within the ecological framework advanced here, AI literacy education represents a systemic intervention designed to restore equilibrium across micro-, meso-, and macrosystem levels: it equips learners at the individual level with self-regulatory strategies (Zimmerman, 2000), prepares teachers to redesign assessment and pedagogy at the institutional level (Han & Li, 2026), and aligns with emerging policy directives at the supranational level (UNESCO, 2025; European Commission, 2026). Crucially, this integrated approach recognises that sustainable ethical governance cannot be achieved through policy alone, nor through individual self-regulation alone, but requires coordinated action across all levels of the educational ecosystem. Yet empirical research investigating how AI literacy can be systematically developed within EFL writing classrooms remains conspicuously scarce, with existing studies offering conceptual proposals rather than tested pedagogical models. This gap directly shapes the scholarly agenda that the present review now pursues, and in doing so it firmly situates ethical governance within the sphere of instructional practice. The governance challenges discussed here further reinforce the theoretical argument proposed in Section 4.

6. Research Gaps and Future Directions: A Research Agenda

Building on the integrated ecological framework advanced throughout this review, this section does not merely enumerate isolated research suggestions but rather proposes a systematic research agenda designed to move the field beyond fragmented debates over technological effectiveness towards a comprehensive understanding of AI-assisted writing ecologies. The field now stands at a threshold where, collectively, the documented shortcomings suggest that additional outcome-oriented studies alone are unlikely to yield fresh insights. This situation calls for a more systematic and integrated inquiry agenda. Specifically, the four categories proposed below emerge from the earlier critiques presented in prior sections, with each category tackling a separate yet mutually connected component of the current research environment.

Table 1
Synthesis of Research Consensus, Contradictions, and Future Research Agenda

Research Theme	Major Consensus	Contradictory Findings	Methodological Limitations	Future Research Agenda
Learning Outcomes	Studies consistently show AI tools improve L2 writing proficiency, vocabulary acquisition, and grammatical accuracy. Students report increased confidence in producing complex academic texts.	Debate exists on whether writing quality gains are sustained without continuous AI use. Some research indicates a decline in critical thinking skills, while others find no significant negative impact.	Dominance of short-term, experimental studies with small sample sizes, limiting the generalizability of long-term learning effects. Lack of robust control groups in many studies.	Conduct longitudinal studies tracking L2 learners' writing development over degrees. Investigate the impact of various AI interaction types (e.g., feedback vs. text generation) on specific language skills.
Self-Regulated Learning (SRL)	AI functions as a scaffold, assisting students in goal-setting, planning, and revising their writing. Learners increasingly adopt AI for self-monitoring and feedback-seeking behaviors.	Research is mixed on whether AI promotes or hinders metacognitive strategies. Some argue it encourages dependency, while others suggest it facilitates critical evaluation when used reflectively.	Heavily reliant on self-reported survey data, which may not accurately reflect actual SRL behaviors or cognitive processes. A lack of fine-grained, process-oriented studies.	Employ mixed-methods approaches combining think-aloud protocols, screen recordings, and surveys to understand the nuance of human-AI collaboration. Investigate effective pedagogical interventions that enhance SRL with AI.
Ethical Governance	Plagiarism and academic integrity are primary concerns, requiring clear institutional policies and ethical guidelines. Transparency about AI use is essential for responsible scholarship.	Disagreement persists regarding the definition of authorship and the acceptable boundaries of AI assistance in high-stakes academic contexts. Perspectives on AI attribution vary widely across disciplines and cultures.	Limited empirical research on institutional policy development and enforcement. Studies often focus on student perceptions rather than actual policy effectiveness.	Examine the effectiveness of diverse academic integrity education and policy models. Develop technological and pedagogical solutions for identifying unethical AI use and fostering ethical decision-making.
Theoretical Development	Theories like Sociocultural Theory, Cognitive Load Theory, and Connectivism are commonly used to understand human-AI interaction in writing. Researchers call for a synthesis of relevant theoretical frameworks.	Current theories may not fully account for the dynamic, adaptive nature of generative AI. Ongoing debate exists concerning the most appropriate theoretical lenses for interpreting the changing landscape of AI-assisted writing.	Theoretical frameworks are often applied post hoc rather than guiding research design from the outset. Studies sometimes lack clear operationalization of theoretical concepts.	Develop and refine theories specifically tailored to generative AI-assisted academic writing. Interdisciplinary collaboration is needed to explore the cognitive, social, and ethical dimensions of this technology.

Note. The table synthesizes the principal findings of the reviewed literature.

Table 1 synthesizes the major patterns identified throughout this review. Rather than listing isolated research gaps, it integrates accumulated evidence into a coherent research agenda spanning methodological innovation, theoretical refinement, pedagogical intervention, and educational governance.

6.1 Methodological Priorities

The methodological critique developed in Section 3.3 foregrounds the most urgent priority: the field requires longitudinal, process-oriented, mixed-methods investigations that move beyond cross-sectional, decontextualised designs. In order to track the progression of AI-facilitated writing growth and to observe whether early improvements either diminish or strengthen over time, longitudinal designs should last for a minimum of eight weeks (Konradt, 2025). Critically, researchers must supplement product-oriented outcome measures with fine-grained process data, including screen-recording protocols, stimulated recall interviews, and multi-draft textual analysis that illuminate how learners interact with AI over successive writing episodes (Du & Chen, 2026; Lee et al., 2025). Such methodological enrichment would enable the field to move beyond the question of whether AI improves writing outcomes to the more theoretically significant question of how such improvement occurs—or fails to occur—through specific interactional patterns.

6.2 Population and Contextual Expansion

The current evidence base remains skewed toward intermediate-to-advanced proficiency learners in East Asian university settings (Dong, 2026; Han & Li, 2026). This narrow demographic focus constitutes a significant gap, given that the moderating effect of proficiency level documented in Section 3.2 suggests that lower-proficiency learners may experience qualitatively different AI interactions than their more advanced counterparts (Fan & Gao, 2026; Zhang & Crosthwaite, 2026). Future research must therefore extend to under-represented populations, including beginner EFL writers, discipline-specific academic writing contexts (e.g., STEM versus humanities), and diverse geographical and institutional settings beyond East Asia (van Niekerk et al., 2025). Equally conspicuous is the scarcity of studies examining teacher perspectives and pedagogical beliefs regarding AI integration—a gap that must be addressed if ecologically valid implementation strategies are to be developed (Han & Li, 2026; Yan et al., 2025).

6.3 Pedagogical Intervention and Design

To address this, design-based research is urgently needed to devise and test instructional strategies that can steer students beyond an early, technology-acceptance-driven phase and toward lasting integration of self-regulated learning, as called for by the theoretical synthesis presented in Sections 4.3 and 5.3. Specifically, future research should investigate how AI literacy—conceptualised as an integrated competence encompassing technical proficiency, critical evaluation, and ethical awareness (Bozkurt, 2024; European Commission, 2026)—can be systematically cultivated within EFL writing curricula. This agenda encompasses the design and empirical validation of instructional models that scaffold learners’ “prompt literacy” (the ability to formulate effective queries), metacognitive reflection on AI use, and the development of evaluative judgment regarding AI-generated content (Cheng et al., 2025; Konradt, 2025). Critically, such interventions must be situated within authentic classroom ecologies and evaluated through ecologically valid, longitudinal designs.

6.4 Theoretical Refinement and Validation

Finally, the integrated TAM-SCT-SRL-ecological framework proposed in this review requires empirical validation. Current scholarship remains characterised by theoretical fragmentation, with studies invoking isolated theoretical lenses that fail to capture the multi-level dynamics of AI-mediated writing (Bandura, 1989, 2001; Davis, 1986; Vygotsky, 1978; Zimmerman, 2000). Future research should employ the integrative framework advanced here to examine how individual cognitive processes (TAM), sociocultural mediation (SCT), self-regulatory development (SRL), and environmental systems (ecological perspective) interact in shaping educational outcomes. Particular attention should be directed toward the macrosystem-level influences articulated in Section 5.1—national and supranational policy frameworks that fundamentally condition institutional practice and learner behaviour (UNESCO, 2025; OECD, 2026; European Commission, 2026)—and their interaction with mesosystem-level assessment cultures and microsystem-level pedagogical practices. This agenda thus calls for multi-level, multi-method research designs that systematically examine the dynamic interplay among tools, learners, and environments.

In sum, the research agenda proposed here charts a coherent path forward: the field must move from cross-sectional outcome studies to longitudinal process investigations, from convenience samples to diverse populations, from atheoretical experimentation to theoretically grounded pedagogical design, and from fragmented conceptualisations to integrative, ecologically informed frameworks that can guide both empirical inquiry and evidence-based practice. Given that it ties conceptual frameworks directly to instructional realities, the foremost task within these research priorities is the empirical testing of instructional designs situated in naturalistic educational settings.

7. Conclusion

This review set out to critically examine the multifaceted role of generative AI in EFL academic writing within higher education, and the synthesis yields three principal insights. First, regarding learning outcomes (RQ1), the evidence reveals no unitary answer: positive effects on surface-level linguistic features are broadly corroborated, yet such gains are systematically moderated by intervention duration, task complexity, learner proficiency, and evaluation methodology—factors frequently overlooked in the existing literature (Dong, 2026; Fan & Gao, 2026; Wang & Ren, 2024; Zhang & Crosthwaite, 2026). Second, concerning self-regulated learning (RQ2), the analysis demonstrates that AI’s educational value is contingent upon whether it functions as a mediational scaffold that promotes cognitive internalisation or as a proxy that displaces learner agency—a distinction theoretically illuminated through the integrated TAM-SCT-SRL framework (Bandura, 2001; Davis, 1986; Vygotsky, 1978; Zimmerman, 2000). Third, with respect to methodological and ethical gaps (RQ3), the field is characterised by a pronounced scarcity of longitudinal, process-oriented investigations and is undergoing a fundamental normative shift from punitive plagiarism policing toward the proactive cultivation of AI literacy as an educational imperative (Bozkurt, 2024; European Commission, 2026; UNESCO, 2025).

By drawing together disparate threads of evidence from the three interrelated spheres of cognition, autonomy, and governance, this review accomplishes two objectives: it consolidates current understanding of GenAI in English-as-a-foreign-language writing, and it offers a conceptually anchored platform for later empirical work. The accumulated evidence confirms that the pedagogical worth of these tools cannot be captured by textual outputs alone; rather, it is a function of shifting interactions among composing practices, learner choice, and administrative oversight. Artificial intelligence in this domain is neither uniformly beneficial nor categorically harmful—its real-world effects are mediated by assignment formats, students’ regulative skills, and the wider evaluation systems that frame instruction. For future research, the chief priority lies in redirecting focus from transient score improvements toward lasting development of authorial authority in AI-infused academic settings. This demands abandoning segmented analyses that treat outputs, operations, or moral concerns in quarantine. In the final analysis, this review argues for a middle ground: GenAI is neither a revolutionary fix nor an existential menace, but its educational relevance is perpetually conditioned by the nexus of student agency, pedagogical decisions, and institutional rules. The integrated, reality-grounded research scheme set forth here provides not only a blueprint for enhanced text production but also a pathway to sustainable authorial self-determination in a fundamentally transformed scholarly landscape.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Alsadoon, R. (2026). Perceptions vs. practices: Academic integrity and actual ChatGPT use among EFL students. *Frontiers in Psychology*, 17, Article 1796737. <https://doi.org/10.3389/fpsyg.2026.1796737>
- [2] Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, 44(9), 1175–1184. <https://doi.org/10.1037/0003-066X.44.9.1175>
- [3] Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26. <https://doi.org/10.1146/annurev.psych.52.1.1>
- [4] Bozkurt, A. (2024). GenAI et al.: Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI. *Open Praxis*, 16(1), 1–10. <https://doi.org/10.55982/openpraxis.16.1.654>
- [5] Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.
- [6] Cheng, D., Li, M., & Lee, T. (2025). Leveraging ChatGPT for research writing: An exploration of ESL graduate students' practices. *Computers and Composition*, 76, Article 102934. <https://doi.org/10.1016/j.compcom.2025.102934>
- [7] Davis, F. D. (1986). A technology acceptance model for empirically testing new end-user information systems: Theory and results [Doctoral dissertation, Massachusetts Institute of Technology]. ProQuest Dissertations & Theses Global.
- [8] Dong, Y. (2026). Generative AI technologies and educational outcomes: A comprehensive meta-analysis comparing traditional and AI-driven approaches. *Humanities & Social Sciences Communications*, 13(1), Article 559. <https://doi.org/10.1057/s41599-026-06903-y>
- [9] Du, J., & Chen, Y. (2026). Exploring English for academic purposes learners' engagement with ChatGPT-4 in academic writing revision: A case study from a private language institute in China. *Frontiers in Artificial Intelligence*, 9, Article 1802007. <https://doi.org/10.3389/frai.2026.1802007>
- [10] European Commission. (2026, June 9). Guidelines for the ethical use of AI and data in teaching. Digital Education Action Plan. <https://education.ec.europa.eu/focus-topics/digital-education/actions/plan/ethical-guidelines-for-educators-on-using-artificial-intelligence>
- [11] Fan, K., & Gao, T. (2026). The role of AI-mediated metacognitive interventions in EFL writing development: A cross-sectional study of textual accuracy, lexical sophistication, and cohesion. *Acta Psychologica*, 267, Article 107069. <https://doi.org/10.1016/j.actpsy.2026.107069>
- [12] Han, J., & Li, M. (2026). ChatGPT-supported teacher feedback on L2 students' writing: EFL teachers' perspectives. *The Asia-Pacific Education Researcher*, 35, 727–744. <https://doi.org/10.1007/s40299-025-01068-9>
- [13] Konradt, L. (2025). Limitations support learning: A design-based study of ethically scaffolded generative AI tools for long-form argumentation and metacognition [Doctoral dissertation, University of Central Florida]. ProQuest Dissertations & Theses Global.
- [14] Lee, M., Yi, T. R., Der-Thang, C., Song, H. J., & David, H. W. L. (2025). Exploring interactions between learners and ChatGPT from a learner agency perspective: A multiple case study on historical inquiry. *Education and Information Technologies*, 30(6), 7167–7189. <https://doi.org/10.1007/s10639-024-13314-y>
- [15] Lin, M. Y., Hong, J. C., IH, T. K., & Jiang, P. H. (2025). ChatGPT addiction predicted by openness to using new technology and self-efficacy, and reflected in continuous intention. *Interactive Learning Environments*, 33(8), 4805–4819. <https://doi.org/10.1080/10494820.2025.2471910>
- [16] Organisation for Economic Co-operation and Development [OECD]. (2026, January 19). OECD digital education outlook 2026. <https://www.oecd.org/education/digital-education-outlook-2026.htm>
- [17] United Nations Educational, Scientific and Cultural Organization [UNESCO]. (2025, April 14). Artificial intelligence and education: Guidance for policymakers. <https://en.unesco.org/ai-education-guidance-policymakers>
- [18] van Niekerk, J., Delport, P. M. J., & Sutherland, I. (2025). Addressing the use of generative AI in academic writing. *Computers and Education: Artificial Intelligence*, 8, Article 100342. <https://doi.org/10.1016/j.caeai.2024.100342>
- [19] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
- [20] Wang, L. X., & Ren, B. (2024). Enhancing academic writing in a linguistics course with generative AI: An empirical study in a higher education institution in Hong Kong. *Education Sciences*, 14(12), Article 1329. <https://doi.org/10.3390/educsci14121329>
- [21] Yan, Y., Liu, H., & Chau, T. (2025). A systematic review of AI ethics in education: Challenges, policy gaps, and future directions. *Journal of Global Information Management*, 33(1), 1–50. <https://doi.org/10.4018/JGIM.386381>
- [22] Zhang, M., & Crosthwaite, P. (2026). More human than human? Differences in lexis and collocation within academic essays produced by ChatGPT-3.5 and human L2 writers. *IRAL: International Review of Applied Linguistics in Language Teaching*, 64(2), 733–760. <https://doi.org/10.1515/iral-2024-0196>
- [23] Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82–91. <https://doi.org/10.1006/ceps.1999.1016>
- [24] Zizka, L. (2025). "It looks good enough": Recognizing the quality of generative AI output in academic writing tasks in higher education. *Journal of Hospitality & Tourism Education*. Advance online publication. <https://doi.org/10.1080/10963758.2025.2496663> 生成 txt 文本