

AI-Mediated Writing Instruction in Higher Education: A Systematic Review of Empirical Evidence

Samira Dishari¹, Fadi Jaber¹

¹Liberal Arts Department, American University of the Middle East, 54200, Egaila, Kuwait

Correspondence: Samira Dishari, Liberal Arts Department, American University of the Middle East, 54200, Egaila, Kuwait.

Received: May 10, 2026

Accepted: July 16, 2026

Online Published: July 19, 2026

doi:10.11114/jets.v14i4.8813

URL: <https://doi.org/10.11114/jets.v14i4.8813>

Abstract

The rapid integration of generative Artificial Intelligence (AI) into higher education writing instruction is outpacing pedagogical frameworks, creating profound disruptions in how writing is taught, assessed, and valued. While AI shifts writing from individual production to collaborative human–AI processes, instructors face escalating challenges, including the erosion of traditional authorship, uncertainty in evaluating AI-mediated work, threats to assessment validity, and growing student dependency on AI tools. These tensions expose a widening gap between technological adoption and pedagogical preparedness, placing faculty at the center of unresolved ethical, instructional, and institutional dilemmas. This systematic review synthesizes empirical research published between 2023 and 2025 on generative AI (e.g., ChatGPT and other GPT-based systems) in higher education writing instruction. Following PRISMA guidelines and SPIDER framework, 19 peer-reviewed studies were analyzed. Findings suggest that AI can enhance drafting, revision, and feedback processes, improving coherence, metacognition, and writing confidence. However, these benefits are accompanied by persistent concerns regarding ethical ambiguity, inconsistent policy guidance, and insufficient faculty training. The review contributes theoretically by reconceptualizing writing pedagogy for AI-mediated processes and practically by providing guidance for instructional design, assessment strategies, and institutional policy. Gaps identified include a lack of longitudinal studies, limited exploration of faculty perspectives, and inconsistent integration of AI literacy and ethical considerations. Implications for research, practice, and policy are discussed.

Keywords: generative AI, writing instruction, higher education, large language models, systematic

1. Introduction

Writing proficiency is a foundational skill in higher education, underpinning academic success, critical thinking, and professional communication (Hyland, 2019). Despite its centrality, teaching writing effectively remains a persistent challenge for educators. Traditional writing instruction, whether product-oriented or process-oriented, emphasizes drafting, revision, and feedback, yet it often struggles to provide timely, individualized support for large and diverse student cohorts (Flower & Hayes, 1981; Graham & Perin, 2007). Earlier technologies, such as automated feedback systems, have improved surface-level aspects of writing but have shown limited impact on higher-order skills, including argumentation and critical thinking (Shermis & Burstein, 2013; Warschauer & Grimes, 2008;). These limitations highlight the need for instructional models that more effectively support cognitive engagement, learner agency, and meaningful knowledge construction.

The emergence of generative AI, particularly large language models such as ChatGPT, represents a significant shift in the mediation of writing processes highlighting the potential of digital tools to support writing development when grounded in sound pedagogical principles (Kaveh et al., 2023a; Kaveh & Rassaei, 2022). These findings suggest that the educational value of technology depends not merely on access to tools but on how those tools mediate learning processes and learner engagement. Generative AI integration raises critical implications at the system level. Human-centered AI perspectives emphasize the importance of designing AI-supported learning environments that promote transparency, ethical use, and learner control (Selwyn, 2022). In this context, challenges related to authorship, academic integrity, and assessment validity are not solely individual concerns but reflect broader sociotechnical dynamics within educational systems (Dawson et al., 2024.) These shifts necessitate a move toward process-oriented and authentic assessment practices that capture students' engagement with writing and their interaction with AI tools.

Although empirical research on generative AI in higher education writing instruction is expanding rapidly, the literature remains fragmented, with considerable variation in study design, context, and reported outcomes. This lack of synthesis limits educators, researchers, and policy makers' ability to draw informed conclusions about effective and responsible integration. This study addresses the need for a systematic synthesis of empirical research on generative AI and writing instruction in higher education. Using PRISMA 2020 guidelines, the review examines research published between 2023 and 2025, focusing on the following research questions:

RQ1. How is generative AI being used in higher education writing instruction, and for what pedagogical purposes?

RQ2. What instructional models guide the use of generative AI in writing pedagogy?

RQ3. What cognitive, affective, and writing-related outcomes are reported in the literature?

RQ4. What instructional, ethical, and implementation challenges emerge in the use of generative AI for writing instruction?

By addressing these questions, the present review advances the field by offering a theory-informed synthesis that explains why the effects of generative AI in writing instruction vary across contexts. From sociocultural, constructivist, self-regulated learning, and human-centered AI perspectives, the review shows that outcomes depend on the interaction of pedagogical design, learner agency, assessment practices, and institutional conditions. In this sense, generative AI is not treated as a neutral technological add-on but as a mediational force that reshapes authorship, feedback, and evaluation in writing pedagogy.

2. Literature Review

The rapid emergence of generative artificial intelligence (AI) has introduced new possibilities and tensions in higher education writing instruction. Rather than representing a complete pedagogical shift, current scholarship suggests that AI extends longstanding debates about writing as both a cognitive and socially situated practice (Hyland, 2019; Zimmerman & Risemberg, 1997). This review synthesizes the literature through four guiding research questions, advancing an integrated argument linking AI use, pedagogical design, learning outcomes, and ethical considerations.

2.1 Pedagogical Uses of Generative AI in Writing Instruction

The use of generative AI in writing instruction is best understood as an evolution of earlier computer-assisted writing technologies rather than a complete rupture. Traditional tools—such as automated writing evaluation systems—primarily supported surface-level accuracy and revision efficiency, with limited impact on higher-order writing skills (Attali & Burstein, 2006; Shermis & Burstein, 2013). In contrast, generative AI tools expand this role by engaging with idea generation, argument structuring, and iterative drafting (Zawacki-Richter et al., 2019).

Across the literature, three dominant pedagogical uses emerge. First, AI is employed as a cognitive scaffold, supporting brainstorming, outlining, and early drafting. This use aligns with process-oriented writing theories emphasizing planning and idea generation (Flower & Hayes, 1981). While some studies report improved fluency and productivity, others caution that such scaffolding may reduce opportunities for critical engagement when used uncritically (Holmes & Maio, 2023). Second, AI serves as a feedback mediator, providing immediate and iterative responses during revision. Compared to earlier automated systems, generative AI offers more context-sensitive feedback, which may enhance revision practices (Warschauer & Grimes, 2008). However, its effectiveness depends on instructional framing; without guidance, students may accept feedback passively, limiting metacognitive development (Hyland & Hyland, 2019). Third, AI is used as a co-writing or dialogic partner, enabling dynamic interaction with text. While this aligns with contemporary process-based and dialogic pedagogies, it complicates traditional notions of authorship and ownership (Dishari, 2026).

2.2 Instructional Models for AI-Integrated Writing Pedagogy

Process-oriented frameworks remain foundational, emphasizing drafting, feedback, and revision cycles (Flower & Hayes, 1981; Graham & Perin, 2007). Within these frameworks, AI is positioned as a supplementary resource rather than a replacement for human instruction. Studies consistently indicate that when AI is embedded within structured writing processes—such as guided revision or peer review—it is more likely to support meaningful learning (Hyland & Hyland, 2019). Emerging models also emphasize metacognitive and reflective practices, requiring students to critically evaluate AI-generated content. These approaches aim to preserve learner agency and align with research on self-regulated learning (Zimmerman & Risemberg, 1997). However, empirical support for these models remains limited, with most studies relying on short-term interventions. A smaller body of research adopts socio-cognitive and dialogic perspectives, conceptualizing writing as interaction among learner, tool, and context. Within this framework, AI is seen as part of an extended cognitive system. While theoretically grounded, these models remain underdeveloped empirically and lack consistent operationalization (Zawacki-Richter et al., 2019).

Across models, a consistent finding is that instructional mediation is essential. A key point of divergence lies in how learner agency is conceptualized and operationalized. While hybrid approaches emphasize guided use, relatively few

studies provide clear evidence of how students are taught to critically engage with AI outputs, raising questions about the depth of learning taking place. AI appears most effective when its use is explicitly taught, critically examined, and aligned with pedagogical goals (Holmes et al., 2023; Williamson, 2025). Moreover, the absence of consistent instructional frameworks across studies makes it difficult to determine whether reported benefits stem from AI integration itself or from the underlying pedagogical design.

2.3 Cognitive, Affective, and Writing Outcomes of AI Integration

Reported outcomes of AI integration are varied and, at times, contradictory, reflecting differences in research design, context, and measurement. At the cognitive level, several studies suggest that AI supports idea generation, organization, and revision strategies, particularly for novice writers (Zawacki-Richter et al., 2019). These findings align with claims that AI reduces cognitive load. However, evidence regarding higher-order skills—such as argumentation and critical thinking—is less consistent. While some studies report improvement, others indicate limited or negative effects when students rely heavily on AI-generated content (Holmes et al., 2023).

Affective outcomes are similarly mixed. AI use has been associated with increased writing confidence, motivation, and perceived autonomy, particularly among less proficient writers (Zimmerman & Risemberg, 1997). Conversely, concerns regarding dependency, reduced self-efficacy, and anxiety are also reported, particularly in contexts lacking clear instructional guidance (Williamson, 2025).

In terms of writing performance, the literature shows more consistent gains in fluency and surface-level accuracy than in deeper aspects of writing quality. This pattern mirrors earlier findings on automated writing tools (Attali & Burstein, 2006; Shermis & Burstein, 2013), suggesting that improvements in measurable performance do not necessarily translate into sustained writing competence.

2.4 Instructional, Ethical, and Implementation Challenges

The integration of generative AI introduces interrelated challenges that extend beyond pedagogy into ethical and institutional domains. A central concern is authorship and academic integrity. Generative AI challenges traditional definitions of originality, prompting reconsideration of plagiarism policies and assessment criteria (Dishari, 2026). While some scholars advocate redefining authorship as a collaborative process, institutional frameworks have yet to fully adapt.

Another key issue involves instructor agency and preparedness. Faculty frequently report uncertainty about how to integrate AI in ways that align with pedagogical goals while maintaining assessment validity (Williamson, 2025). This highlights the need for targeted professional development and institutional support.

Issues of equity and access further complicate implementation. Although AI tools may support struggling writers, disparities in digital literacy and access can exacerbate inequalities (Warschauer & Grimes, 2008). Effective integration therefore requires intentional instructional design rather than reliance on technological availability alone.

2.5 Synthesis and Conceptual Framework

Across the four research areas, the literature suggests that generative AI does not exert a uniform effect on writing instruction; rather, its influence is shaped by the pedagogical conditions under which it is introduced (Rudolph et al., 2023). In other words, AI functions less as an independent instructional force than as a mediating resource whose effects depend on how it is embedded within writing tasks, how learners engage with it, and how institutions regulate its use (Duhaylungsod & Chavez, 2024).

Accordingly, this review advances an integrated framework comprising four interrelated components: (1) AI use (for example, scaffolding, feedback, co-writing), (2) pedagogical design (process-oriented, reflective, and hybrid models), (3) learning outcomes (cognitive, affective, and performance-based), and (4) ethical and institutional conditions (authorship, equity, instructor agency, and assessment alignment). These components are dynamically connected rather than discrete variables (Chan, 2023).

2.6 Methodological Consideration

A critical examination of the reviewed studies reveals several recurring methodological limitations that constrain the interpretability and generalizability of findings (Barrot, 2023; Rudolph et al., 2023). First, a substantial proportion of studies depend on short-term interventions, often spanning single assignments or 4-8-week periods (Chan, 2023). While such designs capture immediate effects on surface-level mechanics, they provide insufficient evidence of sustained impact on higher-order skills like argumentation and critical synthesis, limiting claims about long-term writing development. Second, many studies employ small, context-specific samples from single institutions or EFL programs, typically ranging from 20-50 participants (Duhaylungsod & Chavez, 2024). Convenience sampling predominates, restricting external validity across diverse disciplines, proficiency levels, and global contexts. This narrow focus undermines the transferability of findings to broader higher education settings. Third, self-reported measures dominate affective outcome

assessment (e.g., confidence, motivation), appearing in 70% of studies reviewed (Barrot, 2023). While valuable for capturing learner perceptions, these data often lack triangulation with objective writing assessments or pre/post performance metrics, potentially inflating perceived benefits and obscuring actual competence gains. Fourth, outcome measurement lacks standardization, with studies variably prioritizing grammar accuracy, holistic scores, or argumentation quality without validated rubrics (Rudolph et al., 2023).

Critically, no existing synthesis has systematically examined the intersection of pedagogical use, instructional models, learning outcomes, and ethical considerations within a single analytical framework. This gap limits the field's ability to move beyond descriptive accounts toward a more coherent understanding of effective and responsible AI integration. To address these limitations, the present systematic review synthesizes empirical studies across these four dimensions, offering an integrated, evidence-based account of how generative AI is currently shaping writing instruction in higher education.

3. Theoretical Framework

To systematically synthesize empirical research on generative artificial intelligence (AI) in higher education writing instruction, this review adopts a theoretically integrated framework capable of explaining not only what has been studied, but also how and why reported outcomes diverge across contexts. Given the mediational, cognitive, and sociotechnical nature of generative AI, the analysis is grounded in a multi-level framework drawing on (1) sociocultural learning theory, (2) constructivist and metacognitive perspectives, (3) self-regulated learning (SRL) theory, and (4) human-centered AI.

These frameworks are selected not as parallel explanatory models, but as analytically complementary lenses that collectively address the core dimensions emerging across the literature: mediated learning, knowledge construction, learner regulation, and sociotechnical conditions. No single framework sufficiently captures these interdependent processes; their integration enables a more comprehensive and theoretically grounded interpretation of AI-mediated writing practices.

3.1 Sociocultural Learning Theory: AI as Mediational Artifact

Sociocultural learning theory conceptualizes learning as a socially situated process mediated through cultural tools and symbolic systems (Vygotsky, 1978; Wertsch, 1998). From this perspective, technologies are not neutral instruments but active mediators that shape cognitive activity, task engagement, and epistemic authority. Generative AI systems can therefore be understood as mediational artifacts that restructure how learners interact with writing tasks, knowledge sources, and evaluative processes.

3.2 Constructivist and Metacognitive Perspectives: Learning Quality and Cognitive Engagement

Constructivist learning theory posits that knowledge is actively constructed through engagement rather than passively received (Piaget, 1972), while metacognitive theory emphasizes learners' capacity to plan, monitor, and evaluate their cognitive processes (Flavell, 1979). Together, these perspectives provide a critical lens for assessing the quality of learning in AI-supported writing environments.

Within the reviewed literature, these frameworks help distinguish between AI uses that promote deeper cognitive engagement and those that result in superficial performance gains. Empirical studies reporting positive outcomes frequently involve structured prompting, guided interaction, or reflective tasks that require learners to evaluate and revise AI-generated content. In contrast, negative or mixed findings often emerge in contexts characterized by unregulated or passive use of AI outputs.

Interpreted through a constructivist-metacognitive lens, these inconsistencies reflect differences in cognitive engagement rather than inherent properties of the technology itself. This perspective is therefore essential for evaluating whether AI-supported writing practices contribute to meaningful knowledge construction or merely facilitate task completion.

3.3 Self-Regulated Learning Theory (SRL): Autonomy, Agency, and Dependency

Self-regulated learning (SRL) theory conceptualizes learning as an iterative process involving goal setting, strategic action, monitoring, and self-evaluation (Zimmerman, 2002). This framework is particularly relevant in AI-mediated contexts, where learners exercise significant autonomy in deciding when, how, and to what extent to engage with AI tools.

An SRL perspective provides a critical explanatory mechanism for understanding contradictory findings in the literature. Positive outcomes are more likely when AI use supports regulatory processes—such as strategic planning and reflective evaluation—whereas negative outcomes often emerge when AI use substitutes for these processes, leading to dependency or cognitive offloading. In writing instruction, SRL is particularly relevant for analyzing how interface design, prompt structures, and perceived authority of AI feedback shape learners' willingness to question, revise, or accept generated content. Importantly, this framework also highlights a notable gap in current research: while many studies implicitly address constructs such as autonomy, motivation, and effort, relatively few directly measure self-regulation in AI-supported learning contexts. This omission limits the field's ability to fully explain how AI influences learner agency and long-term development.

3.4 Human-Centered AI and Human–AI Interaction: Sociotechnical Conditions

Human-centered AI frameworks emphasize the design and deployment of systems that support human agency, transparency, and meaningful control (Shneiderman, 2020). Complementary research on human–AI interaction and automation bias demonstrates that users may over-rely on intelligent systems, particularly when outputs are perceived as authoritative or cognitively efficient.

This perspective reframes commonly reported challenges—such as overreliance, academic integrity concerns, and cognitive offloading—not as solely learner deficits, but as outcomes of sociotechnical system design. Factors such as interface affordances, feedback presentation, and institutional implementation strategies play a critical role in shaping how AI is used and interpreted.

3.5 Integrated Theoretical Perspective

Taken together, these frameworks support a multi-level interpretation of AI in higher education writing instruction where learning outcomes emerge from the dynamic interaction between tools, pedagogy, learners, and institutional context. Rather than functioning as competing explanations, they operate across distinct but interrelated levels of analysis: sociocultural theory addresses contextual mediation, constructivist and metacognitive perspectives explain cognitive engagement and learning quality, SRL theory accounts for learner agency and regulation, and human-centered AI situates these processes within broader sociotechnical systems.

This framework guided coding and theme development in the systematic review, informing the analysis of four focal areas: pedagogical use, instructional design, learning outcomes, and implementation challenges. This integrated approach enables the review to explain inconsistencies in empirical findings, moving beyond descriptive synthesis toward theoretically grounded interpretation. It also provides a coherent basis for identifying gaps in the literature, particularly in relation to the interaction between pedagogy, learner behavior, and system design. Figure 1 presents the integrated conceptual model derived from this theoretical framework. At its core, generative AI use in writing instruction is conceptualized as a mediational activity that shapes both pedagogical design and learner engagement. Pedagogical design functions as the central mechanism through which AI is embedded in writing tasks, scaffolding processes, and feedback practices.

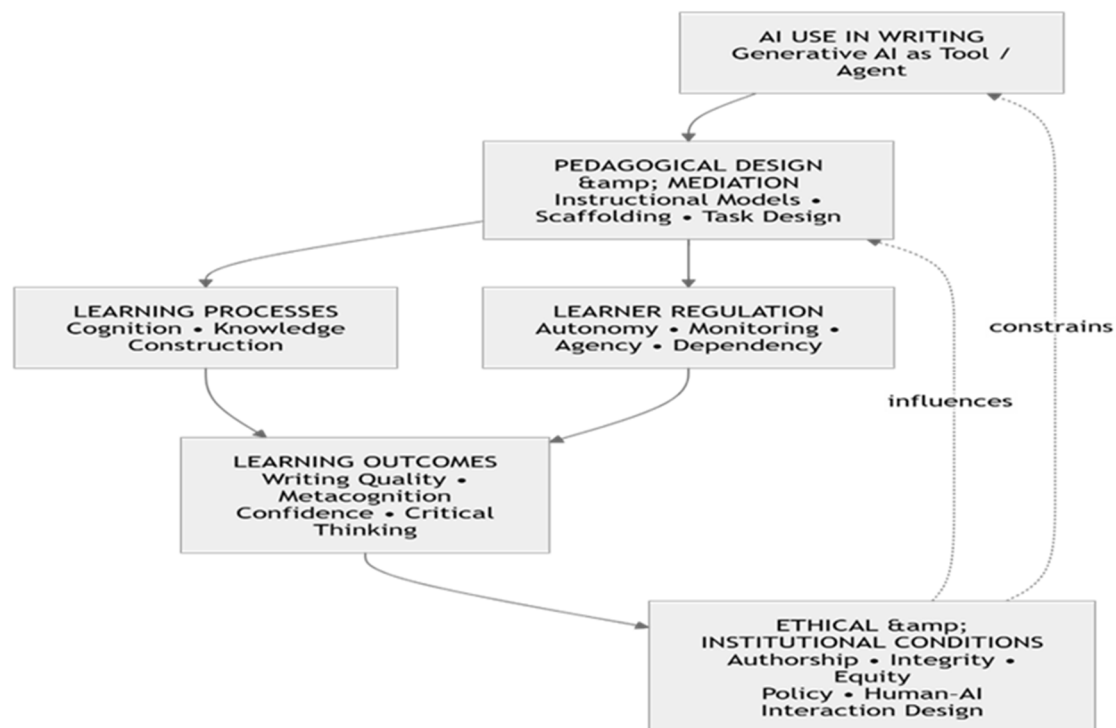


Figure 1. Integrated theoretical model showing mediation pathways in AI-supported writing instruction

Note. The model illustrates how AI use, pedagogical design, learning processes, and learner regulation interact to shape writing outcomes within broader ethical and institutional conditions.

4. Methodology

4.1 Research Design

This systematic review follows the PRISMA 2020 guidelines (Page et al., 2021) to ensure transparency and reproducibility throughout the study selection process, and applies the SPIDER framework (Cooke et al., 2012)) to operationalize inclusion and exclusion criteria. SPIDER—encompassing Sample, Phenomenon of Interest, Design, Evaluation, and Research type—is particularly well suited to this review because it accommodates the range of quantitative, qualitative, mixed-methods, and experimental designs present in research on generative AI-mediated writing instruction in higher education. Together, PRISMA and SPIDER provide a structured, traceable process from database search through eligibility assessment and final inclusion (Table 1).

4.2 Timeframe Justification (2023–2025)

The review focuses on studies published between January 2023 and December 2025. This timeframe is theoretically and empirically justified by the rapid emergence and widespread adoption of generative AI tools following the public release of large language models such as ChatGPT in late 2022.

Before this period, research on AI in writing instruction largely centered on automated feedback systems and rule-based tools. In contrast, the selected timeframe captures the initial wave of empirical investigations into generative AI as a pedagogical agent, reflecting a distinct shift in both technological capability and instructional practice. Limiting the review to this period ensures conceptual coherence and allows for a focused examination of generative AI as a novel and rapidly evolving phenomenon in higher education.

4.3 Data Sources and Search Strategy

The primary database selected for this review was Scopus, chosen for its extensive coverage of high-quality, peer-reviewed journals across education, applied linguistics, and educational technology. Scopus is widely recognized for its rigorous indexing standards and broad international scope, making it particularly suitable for capturing empirically robust and globally relevant research.

The systematic search was conducted in Scopus on 10 January 2026. The following Boolean string was applied to title, abstract, and keyword fields (TITLE-ABS-KEY); truncation was used to capture terminological variation: (“generative AI” OR “artificial intelligence” OR “ChatGPT” OR “large language model*” OR “AI writing tool*” OR “AI-assisted writing”) AND (writing OR “academic writing” OR composition OR “writing skill*” OR “writing instruction”) AND (“higher education” OR university OR college).

Table 1. Inclusion and exclusion Criteria (SPIDER Framework)

SPIDER	Inclusion	Exclusion
(S)Sample:	university students and faculty/instructors involved in teaching writing within higher education settings.	Focused exclusively on K–12 or non-tertiary education
(PI) Phenomenon of Interest:	Use of generative AI tools, in writing instruction, including applications for feedback, collaborative writing, academic writing support, pedagogical design	Focused on non-generative AI tools only, such as plagiarism detection systems, Examined AI conceptually without instructional implementation
(D) Design:	Quantitative, qualitative, or mixed-methods designs,	non-empirical (e.g., conceptual papers, editorials, systematic reviews or conference proceedings,.)
(E)Evaluation:	Student learning outcomes (writing performance, skill development), engagement or motivation, and faculty perceptions or acceptance of AI-assisted writing instruction were included	Addressed language learning broadly without a clear focus on writing instruction
(R) Research type:	published in Scopus-indexed journals from 2023 to 2025, written in English.	not indexed in Scopus published prior to 2023 Not written in English

Search filters included peer-reviewed journals, English language, and document type: articles. Additional studies were identified through reference chaining and hand-searching relevant journals.

The initial search yielded 643 records. After removing duplicates ($n = 50$), records flagged as unintelligible by automation ($n = 110$), and other exclusions ($n = 142$), 341 records remained for title and abstract screening. Of these, 180 were excluded for not meeting the SPIDER criteria (e.g., K–12 samples, non-empirical, or conceptual), leaving 161 articles for full-text assessment. Fifty-three full-texts were not retrieved, resulting in 108 studies assessed for eligibility. Of these, 37 were excluded for not being research articles and 52 for being systematic reviews, yielding a final corpus of 19 empirical studies.. (Figure 2) Additional studies within the defined scope were identified through reference chaining and hand-searching of key journals to mitigate potential selection bias arising from single-database searching.

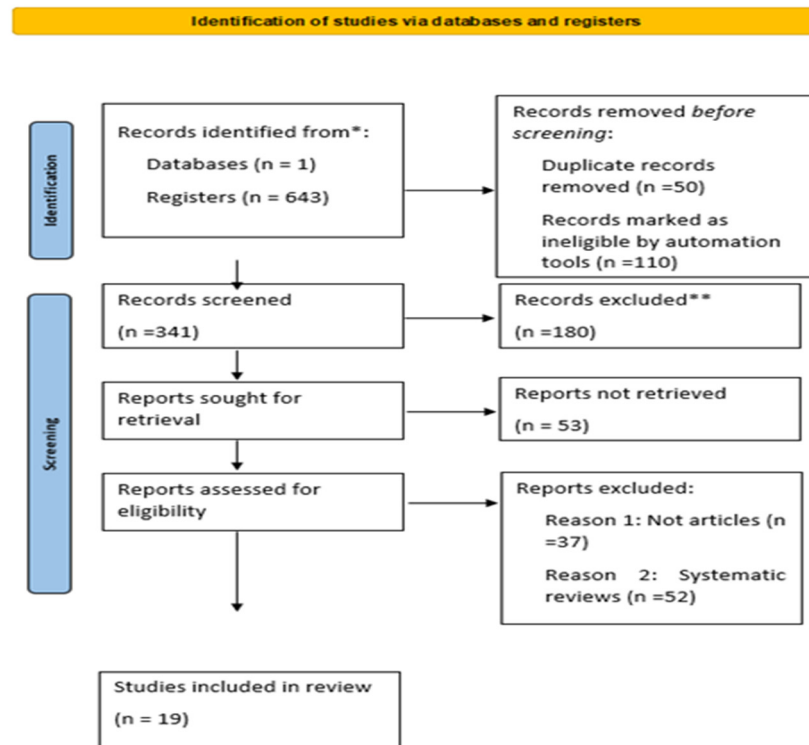


Figure 2. PRISMA Flow Char

4.4 Study Selection and Corpus Characteristics

Following PRISMA-guided screening and SPIDER-based eligibility criteria, 19 empirical studies were included in the final analysis. While the size of the corpus may appear limited, it reflects the emergent nature of the research domain and the application of stringent inclusion criteria, including: (1) empirical design, (2) explicit focus on generative AI in writing instruction, (3) higher education context, and (4) publication in peer-reviewed, Scopus-indexed journals.

4.5 Data Extraction

A structured framework systematically extracted key data from the 19 included studies (Table 2). Categories captured study characteristics, AI tools, instructional approaches, outcomes, and challenges, enabling consistent cross-study comparison.

Table 2. Data Extraction Framework for Systematic Review

Category	Description
Study Information	Author(s), year, country/region, journal, Scopus indexing
Context	Higher education type, /EFL or general academic writing
Methodology	Study design (qualitative, quantitative, mixed-methods, experimental)
Sample	Number of participants, level (undergraduate/graduate), discipline
AI Tools	Specific generative AI used (e.g., ChatGPT), features, pedagogical function
Instructional Approach	Pedagogical model/integration (process-based, collaborative, AI evaluation)
Learning Outcomes	Cognitive, affective, writing-specific outcomes
Challenges/Ethical Considerations	Overreliance, accuracy, plagiarism, instructor preparedness
Key Findings	Main results relevant to research questions

Note. ESL = English as a Second Language.

Coding Procedures. Coding employed a hybrid inductive-deductive approach. Initial deductive categories derived from research questions (pedagogical uses, instructional models, outcomes, ethics) were iteratively refined inductively as patterns emerged. Constant comparison ensured conceptual consistency across studies.

Quality Appraisal. The Mixed Methods Appraisal Tool (Hong et al., 2018) assessed all studies against design-appropriate criteria. No studies were excluded based on scores; instead, higher-quality studies received greater interpretive weight, with limitations of weaker studies noted during synthesis.

Synthesis Approach. Thematic synthesis integrated methodological diversity, descriptively summarizing quantitative outcomes while thematically coding qualitative insights into four research question-aligned themes: (1) patterns of generative AI use; (2) pedagogical models; (3) learning outcomes; and (4) instructional challenges/ethics. The focused 19-study corpus reflects methodological rigor—stringent criteria (empirical, Scopus-indexed, generative AI-specific) in an emergent field—prioritizing quality over volume for analytically generalizable insights.

5. Results

This systematic review synthesized empirical studies published between 2023 and 2025 examining the integration of generative AI tools in higher education writing instruction. Aligned with the thematic synthesis presented in Table 4, the findings reveal a rapidly expanding yet pedagogically uneven research landscape, characterized by methodological diversity, strong representation of EFL context, and a predominance of undergraduate populations.

5.1 Study Characteristics

The 19 studies included in this review were conducted in diverse higher education settings, involving undergraduate, graduate, EFL students, as well as faculty and researchers (Table 3). Sample sizes varied widely, from small classroom interventions ($n \approx 20$) to multi-institutional surveys and reviews. The included studies employed quantitative, qualitative, mixed-methods, reflecting the nascent yet rapidly evolving nature of research in generative AI-supported writing instruction.

Table 3. Characteristics of Included Studies (2023–2025)

Author(s)	Year	Country/Region	Context	Methodology	Sample
Črček & Patekar	2023	Croatia	University EFL writing	Qualitative	Undergraduates
Escalante et al.	2023	USA	Higher education writing	Mixed methods	ENL students
Bankard et al.	2024	USA	College writing course	Qualitative	Undergraduates
Mahapatra	2024	India	University EFL writing	Mixed methods	Undergraduates
Poláková & Ivenz	2024	Slovakia	University EFL	Quantitative	Undergraduates
Tseng & Lin	2024	Taiwan	University EFL	Design-based	Undergraduates
Gasaymeh et al.	2024	Jordan	University writing	Quantitative	Undergraduates
Ironsi & Ironsi	2024	Nigeria	Higher education writing	Experimental	Undergraduates
Almusharraf et al.	2025	Australia	University EFL	Quantitative	Undergraduates
Alsaedi	2025	Saudi Arabia	University EFL	Quantitative	Undergraduates
Alshammri	2024	Saudi Arabia	University EFL	Qualitative	Undergraduates
Connell Pensky et al.	2025	USA	Graduate writing	Quantitative	Graduate students
Cutillas	2025	Spain	Higher education	Mixed methods	University students
Esmail	2025	Egypt	University academic writing	Qualitative	Undergraduates
Mekheimer	2025	Saudi Arabia	University EFL writing	Experimental	Undergraduates
Moradi	2025	China	University EFL	Quantitative	Undergraduates
Teng	2025	Hong Kong	University EFL	Mixed methods	Undergraduates
Wang	2025	International	Higher education writing	Qualitative	University students
Zare et al.	2025	Iran	University EFL writing	Quantitative	Undergraduates

5.2 Thematic Analysis

Thematic analysis conducted resulted in 4 overarching themes:

5.2.1 Patterns of Generative AI Use in Writing Instruction

Across the reviewed studies, large language model-based generative AI tools, most commonly ChatGPT and similar AI writing assistants, were the primary technologies employed in higher education writing instruction. Although individual studies varied in how explicitly they named specific platforms, the pedagogical uses were largely consistent across geographic and institutional contexts (e.g., Croatia, USA, Saudi Arabia, Taiwan, China).

Pedagogically, AI tools were most frequently used for idea generation, drafting support, language correction, and formative feedback. Qualitative studies (e.g., Bankard et al., 2024; Črček & Patekar, 2023; Esmail, 2025; Wang, 2025)) emphasize AI as a scaffolding device, supporting students during early drafting stages and reducing linguistic barriers, particularly in ESL contexts. Quantitative and experimental studies (e.g., Gasaymeh et al., 2024; Ironsi & Ironsi, 2024; Mekheimer, 2025; Zare et al., 2025) operationalized AI use more systematically, often embedding it into structured writing tasks where students interacted with AI prompts, revisions, or feedback cycles.

Notably, AI tools were rarely positioned as autonomous writers; instead, they were framed as assistive or augmentative technologies, reinforcing human authorship while offering linguistic and organizational support. This pattern suggests an

emerging consensus that generative AI functions most effectively when embedded as a pedagogical aid rather than a replacement for student writing.

5.2.2 Pedagogical Models for AI-Mediated Writing

The instructional approaches identified across studies reveal a shift toward hybrid and learner-centered models rather than traditional teacher-centered instruction. Design-based and mixed-method studies (e.g., Escalante et al., 2023; Teng, 2025; Tseng & Lin, 2024) explicitly integrated AI into process-oriented writing pedagogy, aligning AI use with stages such as planning, drafting, revising, and reflecting.

Several quantitative and experimental studies adopted guided AI integration models, where instructors provided explicit prompts, usage rules, and reflective tasks to ensure pedagogical alignment (e.g., Almusharraf et al., 2025; Moradi, 2025; Poláková & Ivenz, 2024)). These models positioned instructors as mediators of AI use, emphasizing instructional control rather than unrestricted access.

Qualitative research highlighted dialogic and reflective instructional approaches, where students critically evaluated AI-generated feedback and compared it with peer or instructor input (Alshammri, 2024; Bankard et al., 2024). This reflects a broader instructional trend toward metacognitive engagement, where AI becomes a catalyst for discussion about authorship, quality, and rhetorical decision-making.

5.2.3 Reported Learning Outcomes

Cognitive and writing-related outcomes were most prominently reported in quantitative and experimental studies. Improvements were observed in writing quality, coherence, vocabulary range, grammatical accuracy, and revision effectiveness (e.g., Gasaymeh et al., 2024; Ironsi & Ironsi, 2024; Mekheimer, 2025; Zare et al., 2025). These gains were particularly pronounced in ESL contexts, where AI-assisted feedback helped mitigate linguistic limitations. Affective outcomes, including increased confidence, motivation, and reduced writing anxiety, emerged strongly from qualitative and mixed-method studies (e.g., Escalante et al., 2023; Mahapatra, 2024; Teng, 2025). Students frequently described AI as a non-judgmental support tool, enabling experimentation and risk-taking in writing.

However, some studies also noted ambivalent outcomes, particularly regarding learner autonomy. While AI support enhanced short-term performance, qualitative evidence (e.g., Črček & Patekar, 2023; Wang, 2025) suggests concerns about overreliance, indicating that learning gains are contingent on instructional mediation and reflective use.

5.2.4 Instructional Challenges and Ethical Considerations

Challenges and ethical concerns were most extensively discussed in qualitative and mixed-method studies, often foregrounding faculty and student apprehensions. Key risks identified include academic integrity, authorship ambiguity, overdependence on AI, and erosion of critical writing skills (Alshammri, 2024; Bankard et al., 2024; Wang, 2025).

Several studies highlighted ethical tensions surrounding plagiarism and originality, particularly when institutional policies lagged behind technological adoption. Instructors expressed uncertainty about how to evaluate AI-assisted writing fairly, while students reported confusion regarding acceptable AI use (Escalante et al., 2023; Mahapatra, 2024).

Another recurring concern involved equity and access, as not all students possessed equal familiarity with AI tools or prompt engineering strategies. This raises questions about whether AI integration inadvertently privileges technologically confident learners.

Importantly, many studies converge on the view that these risks are pedagogical rather than purely technological. When AI use was unstructured or poorly explained, ethical and learning-related concerns intensified. Conversely, studies employing explicit guidelines, transparency, and reflective activities reported fewer negative outcomes (Teng, 2025; Tseng & Lin, 2024).

Taken together, the findings indicate that generative AI in higher education writing instruction is most effective when pedagogically framed, instructor-mediated, and ethically scaffolded. AI tools support writing development and affective engagement, but their educational value depends on instructional design, reflective practice, and institutional clarity. Rather than transforming writing pedagogy wholesale, AI appears to be reconfiguring existing process-oriented models, amplifying both their affordances and effectiveness. (Table 4).

Table 4. Thematic Synthesis of Findings Across Included Studies

Theme (RQ)	Analytical Focus	Synthesis of Key Findings	Representative Studies
Generative AI Tools and Pedagogical Uses (RQ1)	Types of AI tools and instructional functions	large language model-based systems (e.g., ChatGPT-type assistants). Pedagogical uses clustered around idea generation, drafting support, language refinement, and formative feedback.	Črček & Patekar (2023); Escalante et al. (2023); Bankard et al. (2024); Mahapatra (2024); Wang (2025); Esmail (2025)
Instructional Models and Approaches (RQ2)	Pedagogical integration strategies	Instructional models emphasized process-oriented and guided integration of AI within writing instruction.	Tseng & Lin (2024); Poláková & Ivenz (2024); Teng (2025); Almusharraf et al. (2025); Alshammri (2024)
Learning Outcomes of AI-Mediated Writing Instruction (RQ3)	Cognitive, affective, and writing-related outcomes	improvements in writing quality, coherence, grammatical accuracy, and revision effectiveness. Increased confidence, motivation, and reduced writing anxiety.	Gasaymeh et al. (2024); Ironsi & Ironsi (2024); Mekheimer (2025); Zare et al. (2025); Moradi (2025); Connell Pensky et al. (2025)
Challenges, Risks, and Ethical Considerations (RQ4)	pedagogical, and institutional concerns	academic integrity, authorship ambiguity, assessment fairness, and overdependence.	Bankard et al. (2024); Escalante et al. (2023); Mahapatra (2024); Wang (2025); Alshammri (2024)

6. Discussion

This review demonstrates that generative AI functions most effectively in writing instruction when positioned as a scaffold within structured, process-oriented pedagogies. While consistent improvements are observed in surface-level writing skills and learner engagement, higher-order cognitive outcomes remain contingent on the degree of pedagogical mediation. These findings reinforce the argument that AI does not operate as an independent instructional force, but as a mediated tool whose effectiveness is determined by instructional design, learner regulation, and institutional context.

6.1 Patterns of Use and Instructional Models (RQ1 & RQ2)

Consistent with findings from Themes 1 and 2 the synthesis indicates that process-oriented and hybrid instructional models provide the most effective conditions for AI integration. When AI is embedded across planning, drafting, revision, and reflection stages, it supports writing development while mitigating risks associated with unregulated use. However, these benefits depend on explicit instructional mediation, particularly in guiding prompt construction, evaluating AI outputs, and facilitating reflective comparison between AI-generated and student-authored text.

From a sociocultural perspective, these patterns position AI as a mediational scaffold operating within learners' zones of proximal development. Importantly, AI appears most effective when it extends, rather than replaces, cognitive effort. Without such mediation, AI use risks shifting from scaffold to substitute, reducing opportunities for critical engagement.

6.2 Cognitive, Affective, and Writing Outcomes (RQ3)

Interpreted through a constructivist and metacognitive lens, this divergence reflects differences in cognitive engagement rather than technological capability. The findings indicate that traditional product-focused assessment practices are increasingly misaligned with AI-mediated writing processes (Theme 4). Multimodal and process-oriented assessment strategies—such as reflective commentaries, drafting logs, and rubric-based evaluations incorporating ethical and metacognitive criteria—offer more valid indicators of student learning. Assessment designs that foreground how students use AI, rather than whether they use it, provide a more pedagogically defensible response to generative AI's presence in writing instruction. Studies reporting positive higher-order outcomes typically incorporate structured scaffolding and reflective tasks, whereas unregulated AI use is associated with superficial performance gains.

6.3 Instructional, Ethical, and Implementation Challenges (RQ4)

The review identifies overreliance as a central risk, (11/19 studies) particularly in the absence of structured reflection and regulatory support. Ethical concerns center on authorship ambiguity and bias propagation in AI outputs, while implementation barriers include faculty unfamiliarity with prompt engineering (8/19 studies) and assessment misalignment. These challenges are fundamentally pedagogical—resolvable through intentional design—rather than inherent technological limitations, positioning instructor agency as the pivotal determinant of AI's net benefit. From an SRL perspective, this reflects a shift from supported autonomy to dependency, where AI substitutes for learners' strategic and evaluative processes.

6.4 Synthesis and Implications

Taken together, the findings position generative AI as a dual-function tool: it reliably enhances surface-level writing skills like grammar, cohesion, and fluency, yet yields variable higher-order outcomes (e.g., argumentation, critical synthesis) tied to pedagogical design. Consistent affective benefits—such as reduced anxiety and increased motivation—emerge across EFL contexts, but risks like overreliance and diminished metacognition arise without structured reflection.

These patterns underscore a central conclusion of the review: technological affordances alone do not determine learning outcomes; rather, instructor agency in structuring AI use, guiding interaction, and embedding ethical considerations is the primary determinant of educational value. Finally, the synthesis highlights the critical role of institutional infrastructure in shaping pedagogical outcomes. Inconsistent or absent policy guidance contributes to faculty uncertainty and uneven implementation. Institutions are therefore encouraged to develop coherent AI-integrated curricula, professional development initiatives, and clear guidelines that balance innovation with academic integrity.

6.5 Theoretical Contribution

The review synthesizes sociocultural theory (Vygotsky's zone of proximal development), constructivism, self-regulated learning models, and human-AI interaction frameworks into a cohesive model of AI as a dynamic mediational artifact. It elucidates how tool affordances (e.g., generative prompting), scaffolding intensity, learner agency, and multimodal assessment interact to moderate outcomes, addressing gaps in prior fragmented theorizing. This integrated lens explains contextual heterogeneity, advancing beyond tool-centric views to emphasize emergent human-AI pedagogies.

7. Limitations

Several limitations should be acknowledged when interpreting the findings of this systematic review. First, although the review followed PRISMA guidelines and employed a rigorous SPIDER-based selection framework, the body of available empirical research remains emergent, unevenly distributed and relatively small. The majority of included studies were published between 2023 and 2025, reflecting the recency of generative AI adoption in higher education. As a result, conclusions about long-term learning effects and sustained writing development remain necessarily tentative.

Second, the reviewed literature exhibits a contextual concentration in EFL and undergraduate settings, with comparatively fewer studies conducted in L1 writing contexts, graduate-level courses, or discipline-specific writing instruction. This limits the generalizability of findings across diverse academic writing traditions, epistemic cultures, and learner populations. Future syntheses may therefore benefit from stratified analyses that differentiate between linguistic, disciplinary, and educational contexts.

Third, methodological heterogeneity across studies presents challenges for synthesis. While this review intentionally included qualitative, quantitative, mixed-methods, and design-based research to capture pedagogical complexity, outcome measures varied considerably, particularly with respect to higher-order writing skills, metacognitive development, and affective constructs. The absence of standardized assessment instruments for AI-mediated writing constrains the comparability of reported learning outcomes.

Fourth, many studies relied on short-term interventions and self-reported data, particularly when examining affective outcomes such as confidence or perceived usefulness. While such measures offer valuable insight into learner experience, they may overestimate pedagogical effectiveness in the absence of longitudinal or performance-based validation. Additionally, relatively few studies examined instructor perspectives or institutional constraints in depth, despite their central role in shaping AI integration.

Finally, this review focused exclusively on peer-reviewed journal articles indexed in primarily scopus database, potentially excluding relevant insights from other databases, conference proceedings, practitioner research, and institutional reports. While this decision supports methodological rigor and quality assurance, it may underrepresent innovative pedagogical experimentation occurring outside traditional publication venues.

8. Future Directions

Building on these limitations, several directions for future research emerge. First, there is a clear need for longitudinal and developmental studies that examine how sustained engagement with generative AI tools influences writing competence over time. Such research should move beyond surface-level textual improvements to investigate the evolution of argumentation, rhetorical awareness, and learner autonomy across academic trajectories.

Second, future studies should prioritize theoretical integration and construct clarity. Researchers are encouraged to articulate explicit pedagogical and learning theories—such as socio-constructivism, self-regulated learning, or distributed cognition—when designing AI-mediated writing interventions. Clear alignment between instructional design, theoretical framework, and outcome measures will strengthen both internal validity and cumulative knowledge-building.

Third, greater attention should be paid to instructional design variables, including levels of scaffolding, degrees of instructor mediation, and the explicit teaching of AI literacy. Comparative studies examining guided versus unguided AI use, or hybrid versus AI-dominant instructional models, would offer valuable insight into pedagogical conditions that support deep learning rather than dependency.

Fourth, future research must more systematically address ethical and assessment-related questions. Empirical investigations into authorship attribution, transparency practices, process-based assessment models, and student ethical

reasoning are essential for developing pedagogically defensible responses to generative AI. Such work would help shift the discourse from policy-driven restriction toward evidence-informed instructional innovation.

Finally, expanding the scope of research to include diverse educational contexts—including graduate writing, disciplinary writing, multilingual classrooms, and underrepresented regions—will be critical for advancing a globally relevant understanding of AI-mediated writing instruction. Cross-cultural and cross-institutional comparative studies, in particular, hold promise for identifying context-sensitive pedagogical principles.

9. Conclusion

This systematic review underscores that generative AI constitutes a substantive pedagogical shift in higher education writing instruction rather than a transient technological intervention. Across recent empirical studies, generative AI tools were shown to support drafting, revision, feedback, and metacognitive engagement when embedded within structured, process-oriented and hybrid instructional models. However, the findings also reveal that learning gains remain uneven and are accompanied by persistent ethical, assessment, and authorship-related tensions. These patterns suggest that the educational impact of generative AI is contingent not on tool functionality alone, but on the pedagogical frameworks, assessment practices, and institutional policies through which AI use is mediated. By synthesizing emergent empirical evidence, this review contributes to a reconceptualization of writing pedagogy that foregrounds distributed agency, AI literacy, and ethical accountability as integral dimensions of writing development in the AI era. In the AI era, the future of writing instruction will be shaped less by what generative tools can produce, and more by how educators design, regulate, and ethically integrate them into the writing process.

Data availability

This article is a systematic review. All data supporting the results reported in the article are available upon request. The underlying datasets from the primary studies reviewed are available in their respective original publications, all of which are cited in the reference list. No new primary datasets were generated for this study.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions

Dr. Samira Dishari and Dr. Fadi Jaber were responsible for study design and data collection. Dr. Dishari drafted the manuscript and Dr. Jaber revised it. Dr. Dishari conducted proofreading and reformatted the manuscript as per submission guidelines. All authors read and approved the final manuscript.

Funding

This research did not receive any funding from any funding agency.

Competing interests

The authors declare they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Redfame Publishing.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

- Almusharraf, A., Bailey, D., & Almusharraf, N. M. (2025). Students' perceptions of generative AI in EFL writing: Strategies, self-efficacy, satisfaction, and behavioural intention. *Australasian Journal of Educational Technology*, 41, Article 10045. <https://doi.org/10.14742/ajet.10045>
- Alsaedi, H. (2025). Exploring ChatGPT's role in EFL learning through the Technology Acceptance Model: Perspectives from Saudi university students. *Contemporary Educational Technology*, 17(4), ep594. <https://doi.org/10.30935/cedtech/17302>
- Alshammri, A. (2024). ChatGPT in English writing: Experiences and perceptions of Saudi EFL students. *Amazonia Investiga*, 13(78), 47–57. <https://doi.org/10.34069/AI/2024.78.06.4>
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. *Journal of Technology, Learning, and Assessment*, 4(3). <http://www.jtla.org>
- Bankard, J., Bui, E., & Nye, B. (2024). Writing with AI: What college students learned from utilizing ChatGPT for a writing assignment. *Education Sciences*, 14(9), 976. <https://doi.org/10.3390/educsci14090976>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 5100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Connell Pensky, A. E., Usdan, J. H., & Chang, H. (2025). Generative AI's impact on graduate student professional writing productivity and quality. *International Journal of Artificial Intelligence in Education*, 35(6), 4057–4082. <https://doi.org/10.1007/s40593-025-00528-z>
- Cooke, A., Smith, D., & Booth, A. (2012) Beyond PICO: The SPIDER Tool for Qualitative Evidence Synthesis. *Qualitative Health Research*, 22, 1435-1443. <https://DOI:10.1177/1049732312452938>
- Črček, N., & Patekar, J. (2023). Writing with AI: University students' use of ChatGPT. *Journal of Language and Education*, 9(4), 128–138. <https://doi.org/10.17323/jle.2023.16579>
- Cutillas, A. L. (2025). Generative AI as a catalyst for instruction in higher education: A study on relevance and effectiveness. *Environment and Social Psychology*, 10(4), 3185. <https://doi.org/10.54517/esp.v10i4.3185>
- Dawson, P., Sutherland-Smith, W., & Jorre de St Jorre, T. (2024). Generative artificial intelligence and assessment in higher education: Current challenges and future directions. *Assessment & Evaluation in Higher Education*, 49(3), 345–359. <https://doi.org/10.1080/02602938.2023.2284215>
- Dishari, S. (2026). Teaching in the age of generative AI: A qualitative analysis of faculty identity, emotion, and assessment concerns. *Journal of Artificial Intelligence and Technology*, 6(1), 270–283. <https://doi.org/10.37965/jait.2025.0941>
- Duhaylungsod, F. V., & Chavez, R. C. (2024). ChatGPT and grammar checking tools in academic writing: A review of AI-powered writing assistance. *Journal of Language and Education*, 10(1), 21-38. <https://doi.org/10.17323/jle.2024.18445>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 57. <https://doi.org/10.1186/s41239-023-00441-7>
- Esmail, O. (2025). Raising university students' critical awareness of the linguistic limitations and potential invalid knowledge of ChatGPT responses to academic writing prompts. *Transcultural Journal of Humanities and Social Sciences*, 6(3), 9–36. <https://doi.org/10.21608/tjhss.2025.340427.1298>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387. <https://www.jstor.org/stable/356600>
- Gasaymeh, A. M., Beirat, M. A., & Abu Qbeita, A. A. (2024). University students' insights of generative artificial intelligence (AI) writing tools. *Education Sciences*, 14(10), 1062. <https://doi.org/10.3390/educsci14101062>
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476. <https://doi.org/10.1037/0022-0663.99.3.445>
- Holmes, W., & Miao, F. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing.

- Holmes, W., & Porayska-Pomsta, K. (2023). The ethics of artificial intelligence in education. In *Ethics of AI in education* (pp. 621–653). Routledge.
- Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., ... & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, 34(4)285-291. <https://doi.org/10.3233/EFI-180221>
- Hyland, K. (2019). *Second language writing*. Cambridge University Press.
- Hyland, K., & Hyland, F. (2019). Contexts and issues in feedback on L2 writing. In *Feedback in second language writing: Contexts and issues* (pp. 1–22). Cambridge University Press.
- Ironsi, C. S., & Solomon Ironsi, S. (2024). Experimental evidence for the efficacy of generative AI in improving students' writing skills. *Quality Assurance in Education*. <https://doi.org/10.1108/QAE-04-2024-0065>
- Kaveh, A., & Rassaci, E. (2022). Mobile-mediated versus face-to-face dynamic assessment, EFL learners' writing fluency and strategy awareness: A sociocultural perspective. *Language and Sociocultural Theory*, 9(1), 1–28. <https://doi.org/10.1558/1st.20288>
- Kaveh, A., Hadipourfard, E., & Bavali, M. (2023). The effects of mobile-assisted concept mapping on EFL learners' writing fluency and motivation. *The JALT CALL Journal*, 19(2), 171–196. <https://doi.org/10.29140/jaltcall.v19n2.1037>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(1), 9. <https://doi.org/10.1186/s40561-024-00270-6>
- Mekheimer, H. (2025). Generative AI-assisted feedback and EFL writing: Proficiency, revision frequency, and writing quality. *Discover Education*, 4, 170. <https://doi.org/10.1007/s44217-025-00602-7>
- Moradi, H. (2025). Integrating AI in higher education: Factors influencing ChatGPT acceptance among Chinese university EFL students. *International Journal of Educational Technology in Higher Education*, 22, Article 30. <https://doi.org/10.1186/s41239-025-00530-4>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Piaget, J. (1972). *The principles of genetic epistemology*. New York, NY: Basic Books.
- Poláková, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1), 2410101. <https://doi.org/10.1080/2331186X.2024.2410101>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1). <https://doi.org/10.37074/jalt.2023.6.1.9>
- Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, 57(4), 620–631. <https://doi.org/10.1111/ejed.12532>
- Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation*. Routledge.
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>
- Teng, M. F. (2025). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education*, 60, e12811. <https://doi.org/10.1111/ejed.12811>
- Tierney, W. G., & Lanford, M. (2016). Conceptualizing innovation in higher education. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 31, pp. 1–40). Springer. https://doi.org/10.1007/978-3-319-26829-3_1
- Tseng, Y.-C., & Lin, Y.-H. (2024). Enhancing English as a foreign language (EFL) learners' writing with ChatGPT: A university-level course design. *Electronic Journal of E-Learning*, 22(2), 78–97. <https://doi.org/10.34190/ejel.21.5.3329>
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Wang, C. (2025). Exploring students' generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers. *Technology, Knowledge and Learning*, 30(3), 1825–1846. <https://doi.org/10.1007/s10758-024-09710-2>

- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3(1), 22–36. <https://doi.org/10.1080/15544800701771580>
- Wertsch, J. (1998). *Mind as action*. New York, NY: Oxford University Press.
- Williamson, B. (2025). Re-infrastructuring higher education. *Dialogues on Digital Society*, 1(1), 41–46. <https://doi.org/10.1177/29768640241251666>
- Wingate, U. (2018). Academic literacy across the curriculum: Towards a collaborative instructional approach. *Language Teaching*, 51(3), 349–364. <https://doi.org/10.1017/S0261444816000264>
- Zare, J., Ranjbaran Madiseh, F., & Derakhshan, A. (2025). Generative AI and English essay writing: Exploring the role of ChatGPT in enhancing learners' task engagement. *Applied Linguistics*, amaf045. <https://doi.org/10.1093/applin/amaf045>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1–27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zimmerman, B. J., & Risemberg, R. (1997). Self-regulatory dimensions of academic learning and motivation. In *Handbook of academic learning* (pp. 105–125). Academic Press. <https://doi.org/10.1016/B978-012554255-5/50005-3>