

AI-Supported Foreign Language Writing in Higher Education: A Systematic Review of Writing Development and Pedagogical Challenges

Nur Hafizah Abd Aziz^{1 2}, Harun Baharudin^{3*}, Hazrati Husnin⁴

¹ Faculty of Education, Universiti Kebangsaan Malaysia, Selangor, Malaysia

² Fakulti Pengajian Bahasa Utama, Universiti Sains Islam Malaysia, Negeri Sembilan, Malaysia

^{3*} Pusat Kajian Kepelbagaian Pendidikan, Faculty of Education, Universiti Kebangsaan Malaysia, Selangor, Malaysia.
(Corresponding Author)

⁴ Pusat Kajian Inovasi Pembelajaran & Pengajaran, Faculty of Education, Universiti Kebangsaan Malaysia, Selangor, Malaysia

Abstract

This systematic literature review examines the use of artificial intelligence (AI) in foreign language writing instruction in higher education, with particular attention to writing development and pedagogical challenges. Guided by PRISMA 2020 and the PICO framework, the review synthesizes 30 empirical studies published between 2021 and 2025 and retrieved from Scopus, Web of Science, and Google Scholar. The findings show that AI-supported tools, including ChatGPT, Grammarly, QuillBot, Wordtune, and neural machine translation, are most consistently associated with improvements in grammatical accuracy, lexical choice, sentence-level clarity, drafting efficiency, and selected aspects of text quality. Some studies also report reduced writing anxiety, stronger motivation, and increased learner confidence. These benefits, however, are uneven. The evidence is stronger for immediate textual improvement than for rhetorical development, independent authorship, or sustained composing ability. The review argues that AI-supported writing becomes pedagogically meaningful only when embedded in lecturer-mediated instruction, critical AI literacy, and explicit attention to ethical authorship. Without such mediation, AI risks producing more polished texts while leaving learners' writing judgement underdeveloped. The review contributes by distinguishing AI-assisted text improvement from writer development and by positioning lecturer guidance as central to responsible AI integration in foreign language writing pedagogy.

Keywords: *Artificial Intelligence; Foreign Language Writing; Higher Education; Pedagogical integration; Systematic Literature Review*

1. Introduction

Foreign language writing places a heavy burden on university learners because it requires linguistic accuracy, textual organization, coherence, and rhetorical appropriateness to operate at the same time within academically acceptable texts (Hwang et al., 2023; Wale & Kassahun, 2024). At university level, this difficulty becomes more visible. Students are expected to write analytically and with disciplinary sensitivity, although many receive limited individualized guidance during the composing process (Malik et al., 2023; Mohammad, 2024). Empirical studies continue to report weak lexical resources, difficulty paraphrasing academic content, and recurring problems in task achievement, coherence, and grammatical accuracy among tertiary learners (Hwang et al., 2023; Marzuki et al., 2023; Werdiningsih et al., 2024). The problem is not simply that students make errors. They are often required to produce academic writing before they have developed sufficient control over the linguistic, rhetorical, and strategic demands of composing.

AI now occupies a much larger place in foreign language writing instruction than traditional editing software once did. Recent studies show that AI-supported tools are used across idea generation, outlining, drafting, paraphrasing, revision, grammar correction, and feedback uptake (Alharbi, 2023; Guo & Li, 2024; Marzuki et al., 2023). The tools themselves vary widely: automated writing evaluation systems such as Grammarly and Pigai, paraphrasing applications such as QuillBot and Wordtune, generative platforms such as ChatGPT, and integrated learning environments such as Smart RoamLingo, Writerly, and Cami AI (Marzuki et al., 2023; Muslimin et al., 2024; Tsai et al., 2024; Xu & Jumaat, 2024). This range matters pedagogically. AI no longer enters writing only at the point of correction; it can shape how learners plan, draft, revise, and interpret feedback during routine composing practice (Marzuki et al., 2023; Tseng & Lin, 2024). That same expansion also sharpens the risks. Studies have linked AI use to efficiency and text improvement, but also to overreliance, superficial revision, plagiarism, and uncertainty about whether AI-assisted texts represent learners' actual writing competence (Guo & Wang, 2024; Sawangwan, 2024; Tsai et al., 2024; Woo et al., 2024).

The rapid growth of research has not produced a settled pedagogical picture. Existing studies differ in the tools examined, instructional designs adopted, learner populations involved, and outcomes reported. As a result, it remains difficult to separate short-term performance gains from broader developmental effects or to identify the

conditions under which AI meaningfully supports writing instruction (Alharbi, 2023). This review is therefore not designed merely to catalogue recent studies. It synthesizes what current evidence shows about pedagogical outcomes, instructional risks, and the mediating role of teachers in AI-supported foreign language writing in higher education. The review addresses two questions: How does the use of AI influence students' foreign language writing performance and development in higher education? What pedagogical insights and challenges emerge from the integration of AI into foreign language writing instruction in higher education? These questions allow the review to move beyond adoption narratives and examine how AI supports, complicates, and reshapes writing pedagogy.

Methods

Review Protocol

This review followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines, which provide a structured framework for transparent study identification, screening, eligibility assessment, and reporting. PRISMA was selected because it is widely used across education and social science research and offers clear procedural guidance for evidence synthesis. In addition, the review design was informed by the systematic review recommendations of Mohamed Shaffril et al. (2021), particularly in relation to search management, screening procedures, and evidence organization.

Before conducting the search, the reviewers established a protocol covering the review questions, database selection, search terms, eligibility criteria, screening procedures, quality appraisal, and thematic synthesis. The review was limited to English-language peer-reviewed journal articles published between 2021 and 2025 that examined AI-supported foreign language writing in higher education. Two reviewers screened the records independently to reduce selection bias. Disagreements were resolved through discussion until consensus was reached.

Formulation of research questions

The review questions were structured using the PICO framework, because the focus was not only on intervention effects but also on educational practice and context. In this review, the Population comprised higher education students engaged in foreign language writing, the Interest was the use of AI tools to support writing performance and development, and the Context referred to higher education settings in which AI was integrated into foreign language teaching and learning. This framing produced two review questions: i) How does the use of AI influence students' foreign language writing performance and development in higher education?; and ii) What pedagogical insights and challenges emerge from the integration of AI into foreign language writing instruction in higher education?

Systematic search strategies

The search process was organized into three stages: identification, screening, and eligibility. Each stage was linked directly to the review questions and the inclusion criteria.

Identification

The identification stage began with four primary keywords: writing skill, artificial intelligence, foreign language, and higher education. These were expanded into related terms such as writing performance, academic writing, composition, EFL writing, AI chatbot, ChatGPT, machine learning, and AI-assisted writing. The purpose was not to capture every publication mentioning AI, but to identify studies in which AI was used directly to support foreign language writing in university settings. Searches were conducted in Scopus, Web of Science, and Google Scholar. Scopus and Web of Science provided controlled database coverage, while Google Scholar was used only as a supplementary source. To limit noise, only the first six pages of Google Scholar results were screened. This strategy produced 493 initial records for screening.

Table 1: *The search string*

Database	Search String
Scopus	TITLE-ABS-KEY (("writing skill*" OR "writing strateg*" OR "Writing Performance" OR "Academic writing" OR "composition" OR "EFL writing") AND ("artificial intelligence" OR "AI" OR "AI chatbot" OR "ChatGPT" OR "machine learn*" OR "deep learn*" OR "artificial intelligence use*" OR "artificial intelligence assist*") AND ("foreign language" OR "non-native language" OR "english as a foreign language" OR "Arabic as a foreign language" OR "AFL" OR "FLL" OR "EFL") AND ("higher education" OR "college" OR "university" OR "tertiary education"))

WOS ALL=((("writing skill*" OR "writing strateg*" OR "Writing Performance" OR "Academic writing" OR "composition" OR "EFL writing") AND ("artificial intelligence" OR "AI" OR "AI chatbot" OR "ChatGPT" OR "machine learn*" OR "deep learn*" OR "artificial intelligence use*" OR "artificial intelligence assist*") AND ("foreign language" OR "non-native language" OR "english as a foreign language" OR "Arabic as a foreign language" OR "AFL" OR "FLL" OR "EFL")) AND ("higher education" OR "college" OR "university" OR "tertiary education"))

Screening

Screening applied the criteria specified in Table 2. The search was limited to articles published from 2021 to 2025 because the search began in February 2026, before the 2026 publication year was completed. Only English-language empirical studies published as peer-reviewed journal articles were retained. Database filtering and duplicate removal excluded 277 records that were outside the publication period, non-English, non-journal articles, or duplicates. A total of 216 articles then proceeded to eligibility assessment. The total number of records entering screening must be corrected consistently across the text, table, and PRISMA diagram.

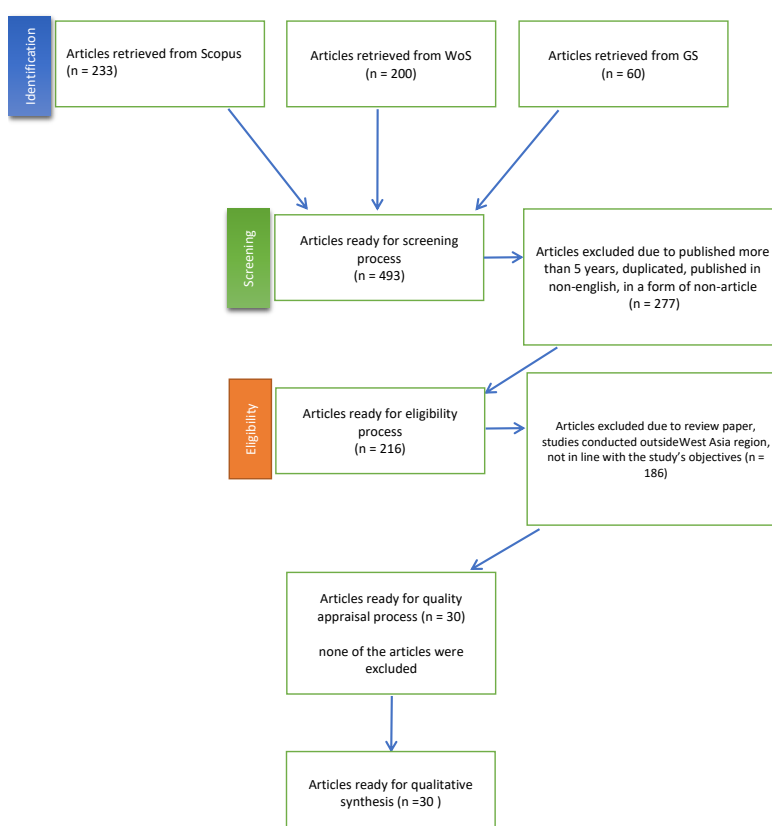


Figure 1: The flow diagram

Eligibility

At the eligibility stage the reviewers assessed the titles and abstracts of the 216 remaining articles against the review objectives and inclusion criteria. A further 186 articles were excluded because they focused on East or Southeast Asia, school education, second language learning only, review-based evidence, or topics outside the scope of the study. This process produced 30 empirical articles for quality appraisal and synthesis.

Table 2: The selection criterion is searching

CRITERION	INCLUSION	EXCLUSION
Language	English	Non-English
Timeline	2021-2025	< 2020 or > 2025
Literature Type	Journal (Article)	Conference, Book, Review

Publication Stage	Final	In Press
Geographic Region	Middle East (West Asia)	East Asia and Southeast Asia
Relevance to RQ	AI in foreign language writing	AI in second language writing

Quality Appraisal

The 30 selected articles were appraised using six criteria adapted from established systematic review practices in educational research (Mohamed Shaffril et al., 2021). The criteria assessed whether each study stated its purpose clearly, established the usefulness of the work, described its methodology, defined key concepts, compared its findings with related work, and acknowledged limitations. Each criterion was scored as Yes (1), Partial (0.5), or No (0). Studies scoring at least three out of six were retained. All 30 articles met this threshold and were therefore included in the final synthesis.

Table 3: The Quality Assessment Criteria

Assessment criteria	
QA1	Is the purpose of the study clearly stated?
QA2	Is the interest and the usefulness of the work clearly presented?
QA3	Is the study methodology clearly established?
QA4	Are the concepts of the approach clearly defined?
QA5	Is the work compared and measured with other similar work?
QA6	Are the limitations of the work clearly mentioned?

Data Extraction and Analysis

Data were extracted from the 30 included articles, with particular attention to the abstract, results, and discussion sections. The extraction focused on evidence relevant to the two review questions, including writing outcomes, forms of AI use, pedagogical conditions, learner responses, lecturer roles, and reported challenges. The extracted information was first organized in a comparison table to support systematic cross-study analysis. Following Braun & Clarke (2019) approach to thematic analysis, the analysis involved identifying recurring patterns, clustering related findings, examining thematic frequency, and interpreting relationships across the extracted evidence.

The analysis produced two broad thematic categories aligned with review questions. The first concerned the influence of AI use on students' foreign language writing performance and development in higher education. The second concerned the pedagogical insights and challenges associated with integrating AI into foreign language writing instruction. Six subthemes were developed within these two categories. Theme naming and refinement were conducted collaboratively by the corresponding author and co-authors through iterative discussions. These discussions focused on resolving inconsistencies, clarifying interpretive boundaries, and ensuring alignment between the themes, review questions, and study objectives.

Results

Table 4: The developed themes and sub-themes

Studies	Study Design	Country	LAWQ	MSEA	AFPI	CAIGS	LMHF	WIVEI
1 (Al-Shaboul et al., 2024)	QL	Jordan	/				/	
2 (Yılmaz Virkan & Tomak, 2025a)	MX	Turkey		/				/
3 (Al Obaydi & Pikhart, 2025)	MX	Iraq & Czech Republic			/		/	
4 (Yılmaz Virkan & Tomak, 2025b)	MX	Turkey		/			/	

5	(Almashour et al., 2025)	QL	Jordan	/	/
6	(Alghannam, 2025)	QL	Saudi Arabia	/	/
7	(Mohammad et al., 2023)	QN	Saudi Arabia	/	/
8	(Alwasidi & Al-Khalifah, 2025)	QN	Saudi Arabia	/	/
9	(Taskiran & Goksel, 2022)	QN	Turkey	/	/
10	(S. E. Al-Alami, 2024)	MX	UAE	/	/
11	(Mahdi et al., 2025)	QN	Iraq	/	/
12	(Mahmud, 2025)	MX	Saudi Arabia	/	/
13	(AbdAlgane et al., 2025)	MX	Saudi Arabia	/	/
14	(S. Al-Alami, 2024)	MX	UAE	/	/
15	(Almahboob, 2025)	QN	Saudi Arabia	/	/
16	(Alshehri, 2025)	MX	Saudi Arabia	/	/
17	(Al Mahmud, 2023)	QN	Saudi Arabia	/	/
18	(Abdullah, 2025)	QL	Oman	/	/
19	(Alyami et al., 2025)	MX	Saudi Arabia	/	
20	(El-Esery, 2025)	QN	Saudi Arabia	/	
21	(Almusharraf et al., 2025)	QN	Saudi Arabia	/	/
22	(Alsofyani & Barzanji, 2025)	MX	Saudi Arabia	/	/
23	(Elmotri et al., 2025)	MX	Saudi Arabia	/	/
24	(Sanosi, 2022)	QN	Saudi Arabia	/	/
25	(Jmaiel et al., 2025)	QN	Saudia Arabia	/	/
26	(Mohammed & Khalid, 2025)	MX	Iraq	/	
27	(Altamimi, 2025)	QL	Saudi Arabia	/	/
28	(Fathi & Rahimi, 2024)	QL	Iran	/	/
29	(Alkamel & Alwagieh, 2024)	QN	Yemen	/	/
30	(Al-Raimi et al., 2024)	QN	Oman	/	/

QN = Quantitative Studies; QL = Qualitative Studies; MX = Mixed-method studies; LAWQ = Linguistic Accuracy and Writing Quality; MSEA = Motivation, Self Efficacy, and Anxiety; AFPI = AI Feedback and Pedagogical Integration; CAIGS = Critical AI Literacy and Guided Scaffolding; LMHF = Lecturer Mediation and Holistic Feedback; WIVEI = Writer Identity, Voice, and Ethical Integrity

Background of the selected article

This review synthesized 30 empirical studies published between 2021 and 2025, reflecting the rapid growth of research on artificial intelligence in foreign language writing instruction in higher education. The largest concentration of studies was found in Saudi Arabia (n = 16), followed by Turkey (n = 3). Smaller clusters were identified in Jordan with two studies each, while Iran and one joint Iraq–Czech Republic study were represented by one study each. This distribution shows that the evidence base is strongly shaped by Arab higher education contexts, with Saudi Arabia functioning as the dominant research site.

The selection of Middle East as the geographical focus of the review was deliberate. Many contexts in this region share relatively centralized curricular structures, assessment-oriented instructional practices, and limited everyday use of the target language outside the classroom. In addition, a substantial proportion of the included studies involve Arabic-speaking learners, whose foreign language writing development may be shaped by distinctive linguistic and rhetorical conditions, including diglossia, cross-linguistic transfer, and sensitivity to formal academic register. These conditions make the region a meaningful site for examining AI-supported writing

pedagogy. Still, the regional frame should not be read as implying uniformity. Middle Eastern higher education systems differ in language policy, institutional capacity, AI regulation, and writing assessment practices.

Methodologically, the studies were distributed across mixed method ($n=12$), quantitative method ($n=12$), and qualitative studies ($n=6$). This distribution gives the review both outcome-oriented and process-oriented lenses. Quantitative studies tended to focus on measurable aspects of writing performance, whereas mixed-method and qualitative studies offered broader insight into learner and teacher experience, feedback processes, instructional design, and ethical concerns surrounding AI use. The publication trend further indicates the recency of the field, with the majority of studies published in 2025, suggesting that research on AI-supported foreign language writing remains in an emergent and rapidly developing stage.

The thematic distribution shows a field moving beyond the question of whether AI improves writing scores. Lecturer roles appeared most frequently ($n=12$), followed by psychological factors ($n=11$) and critical AI literacy and student guidance ($n=11$). AI-generated feedback, human evaluation, and pedagogical integration were also well represented ($n=10$), while linguistic accuracy, productivity, and writing quality appeared in nine studies. By contrast, linguistic identity, writer's voice, and ethical integrity were the least explored issues, appearing in only four studies. This pattern is telling that literature has begun to move beyond a narrow concern with whether AI improves writing performance and is increasingly engaging with broader pedagogical questions such as teacher mediation, learner guidance, and responsible integration. At the same time, the limited attention to writer identity and linguistic integrity indicates a significant gap, particularly in multilingual and culturally situated writing contexts.

AI and Students' Writing Performance and Development

The first major theme concerns how AI affects students' foreign language writing performance and their longer-term development as writers. Three subthemes emerged from the reviewed studies: linguistic accuracy and writing quality; motivation, self-efficacy, and anxiety; and AI feedback and pedagogical integration.

Subtheme 1: Linguistic Accuracy and Writing Quality

The strongest evidence for AI-supported writing appears in visible textual performance. Grammarly, ChatGPT, QuillBot, Wordtune, and neural machine translation were used to correct errors, support drafting, assist paraphrasing, and guide revision. The most consistent gains were reported in grammar, spelling, sentence construction, and lexical choice. AI appears most effective when the writing problem is local, explicit, and technically correctable. Evidence for coherence, organization, and rhetorical control was less stable, and these outcomes seemed more dependent on task design, instructional support, and learner guidance (Alkamel & Alwagieh, 2024; Al-Shaboul et al., 2024; El-Esery, 2025; Sanosi, 2022).

AI also helped students generate ideas, reformulate sentences, elaborate content, and revise drafts, which enabled them to produce clearer and more fluent texts with less immediate difficulty (Jmaiel et al., 2025; Mahdi et al., 2025; Mohammad et al., 2023). However, better written products should not be taken as direct evidence of deeper writing development. In many cases, AI seems to improve the text faster than it develops the writer. The current evidence is stronger for gains in text production than to independent rhetorical control, originality, autonomy, or sustained authorial decision-making. This limitation is especially important because much of the available research is short-term and still centred on English as a foreign language, with limited longitudinal evidence on whether learners can transfer AI-supported improvements into independent writing ability (AbdAlgane et al., 2025; Alkamel & Alwagieh, 2024; Al-Shaboul et al., 2024). The evidence therefore points to a narrower conclusion. AI-supported writing tools improve linguistic accuracy, fluency, and selected aspects of text quality, especially grammar, vocabulary, and revision efficiency. What remains unclear is whether these immediate gains develop into sustained learner autonomy, stronger rhetorical judgement, or the ability to produce stronger texts independently without continued AI support.

Subtheme 2: Motivation, Self-Efficacy, and Anxiety

AI-supported writing also influenced learners' psychological engagement with foreign language writing, particularly in relation to motivation, self-efficacy, and anxiety. Across the reviewed studies, learners often described AI tools as making writing tasks more manageable because they offered immediate support for idea generation, paraphrasing, grammar, and revision. This form of assistance appeared to increase motivation and positive engagement when learners viewed the tools as useful, accessible, and relevant to their immediate writing needs (Almusharraf et al., 2025; Alyami et al., 2025; Mohammad et al., 2024; Mohammed & Khalid, 2025). In this sense, AI functioned not only as a technical writing aid but also as affective support, especially for learners who struggled to begin, extend, or revise written texts.

A similar caution is needed when interpreting gains confidence in writing. Learners frequently reported that AI tools helped them manage vocabulary, grammar, and text development more effectively, while self-efficacy and

satisfaction appeared to influence their willingness to adopt and continue using AI technologies (Almusharraf et al., 2025; Al-Raimi et al., 2024; Altamimi, 2025). Yet confidence is not the same as competence. AI may make students feel more supported during writing, but this does not necessarily mean that they have developed stronger independent control over higher-order thinking, rhetorical judgement, or authorship (S. E. Al-Alami, 2024). The evidence therefore supports a cautious interpretation: AI can strengthen learners' perceived ability to write before it strengthens their independent writing ability.

AI was also associated with reduced writing anxiety and a more comfortable learning environment, but this pattern was conditional rather than universal. The same tools that make writing less intimidating can also create dependence, encourage passive acceptance of feedback, or weaken students' ownership of the text. Concerns about overdependence, plagiarism, inaccurate feedback, weak personalization, and limited cultural appropriateness recur across the studies (*Almahboob, 2025; Altamimi, 2025; Yilmaz Virlan & Tomak, 2025b*). These concerns are significant because much of the evidence remains perception-based and short-term. As a result, current findings justify claims about emotional support, motivation, and perceived self-efficacy more strongly than claims about sustained affective development or independent writing growth.

Subtheme 3: AI Feedback and Pedagogical Integration

AI-generated feedback has become a common source of support in foreign language writing because it offers speed, accessibility, and individualized responses that lecturers may struggle to provide at the same scale. Its strongest contribution lies in lower-level corrective feedback, particularly grammar, spelling, sentence-level clarity, and procedurally clear revision. In some studies, AI feedback performed comparably to teacher feedback, but the evidence does not establish AI as superior to human evaluation. Its value is narrower. AI is useful for immediate and scalable feedback, while human evaluators remain stronger where writing requires interpretation, rhetorical sensitivity, contextual judgement, and response to learner meaning (*Al Mahmud, 2023; Al Obaydi & Pikhart, 2025; Alsofyani & Barzanji, 2025; Taskiran & Goksel, 2022*).

The pedagogical value of AI feedback depends heavily on how it is mediated. Although AI-generated responses were often perceived as efficient and wide-ranging, they could also be inconsistent, overly focused on correctness, and limited in relation to meaning, voice, affective support, and cultural nuance. Students, teachers, and supervisors therefore did not treat AI feedback as automatically authoritative. They evaluated, adapted, accepted, or rejected AI suggestions according to rhetorical purpose, disciplinary expectations, and the intended meaning of the text (*Al Obaydi & Pikhart, 2025; Alghannam, 2025; Almashour et al., 2025; Bani Amer et al., 2025*). This pattern shows that AI feedback is most useful when it becomes something learners judge, not something that replaces judgement.

The literature therefore points toward a hybrid model of pedagogical integration. Stronger outcomes were generally reported when AI feedback was embedded in guided instruction, used to support reflection, and combined with teacher mediation rather than applied in isolation. AI literacy, instructional design, assessment knowledge, and teacher professional development are central to this process because they determine whether feedback is merely received or critically used (*Alshehri, 2025; Elmotri et al., 2025; Fathi & Rahimi, 2024*). AI should therefore be understood as a feedback-extending tool, not a teacher-replacing technology. Its educational value depends on whether it expands feedback opportunities while preserving human interpretation, learner authorship, and sustained developmental support.

Pedagogical Insights and Challenges in AI-Supported Writing

The performance evidence discussed above shows what AI can improve in students' written products. However, it does not explain when AI use becomes pedagogically meaningful. Therefore, the second theme shifts attention from writing outcomes to the instructional conditions that shape AI-supported writing. Three issues were especially prominent across the reviewed studies: critical AI literacy, lecturer mediation, and the protection of writer voice and ethical authorship.

Subtheme 1: Critical AI Literacy and Guided Scaffolding

Across the included studies, AI was most defensible pedagogically when it functioned as a cognitive scaffold rather than as a producer of finished texts. It supported planning, paraphrasing, drafting, and revision within structured learning environments, instead of replacing learners' own composing processes (*Al-Shaboul et al., 2024; Alwasidi & Al-Khalifah, 2025; Mohammad et al., 2023; Yilmaz Virlan & Tomak, 2025a*). Tools such as QuillBot, ChatGPT, and AI-supported translation systems were associated with improvements in paraphrasing, sentence reformulation, idea expansion, and writing performance, particularly among learners who needed support to initiate, extend, or refine written texts (*Al-Raimi et al., 2024; El-Esery, 2025; Mahdi et al., 2025; Mohammad et al., 2023*). The pedagogical value of AI therefore lies not simply in making writing easier, but in reducing

selected cognitive burdens while keeping learners responsible for meaning-making, textual ownership, and final writing decisions.

This scaffolding depends on critical AI literacy and explicit student guidance. Several studies warn against treating surface improvement as evidence of independent higher-order writing development, especially when learners rely on AI without reflective control (S. E. Al-Alami, 2024; Alwasidi & Al-Khalifah, 2025; Jmaiel et al., 2025). AI use therefore requires safeguards against overreliance, plagiarism, and passive acceptance of machine-generated output (Alkamel & Alwagieh, 2024; Almusharraf et al., 2025; Alyami et al., 2025). Guided use was also associated with motivation, self-efficacy, satisfaction, and readiness to write, particularly when students were explicitly taught how and when to use AI productively (Almusharraf et al., 2025; Fathi & Rahimi, 2024; Mohammed & Khalid, 2025). The central issue is not whether AI can support writing. It can. The more important question is whether learners are required to question, revise, and justify its output rather than surrendering judgement, authorship, and higher-order writing decisions to the tool.

Subtheme 2: Lecturer Mediation and Holistic Feedback

A second insight from the reviewed studies is that AI does not diminish the pedagogical importance of lecturers. It changes the nature of their role. AI systems were consistently valued for providing rapid, detailed, and often accurate feedback on grammar, spelling, and lower-order revision. Yet this efficiency does not replace human academic judgement. Comparative studies show that AI feedback may, in some cases, perform similarly to teacher feedback and may even produce stronger short-term gains. However, human feedback remains more reliable when writing requires interpretation, dialogue, rhetorical sensitivity, disciplinary awareness, and purposeful response to learner meaning (Al Mahmud, 2023; Al Obaydi & Pikhart, 2025; Alghannam, 2025; Alsofyani & Barzanji, 2025; Taskiran & Goksel, 2022). Lecturer mediation is therefore most important at the point where writing moves beyond correction into meaning-making.

The reviewed studies also reposition lecturers as validators, monitors, and designers of assessment practice. Instructors valued AI for grammar checking, paraphrasing, coherence support, and idea development, but they remained concerned about originality, creativity, academic integrity, and student overreliance (Almahboob, 2025; Al-Shaboul et al., 2024; Yilmaz Virlan & Tomak, 2025b). In advanced writing contexts, these concerns extended to organization, argument development, learner autonomy, and supervisor-student interaction (Bani Amer et al., 2025). Even when automated writing evaluation and automated written corrective feedback were viewed positively, their educational value depended on teacher oversight, pedagogically informed design, and user-centred refinement (Alshehri, 2025; Elmotri et al., 2025). Lecturers are therefore needed less as routine language correctors and more as academic gatekeepers who evaluate meaning, coherence, argument, authorship, and integrity.

Subtheme 3: Writer Voice, Linguistic Identity, and Ethical Integrity

The preservation of linguistic identity, writer voice, and ethical integrity remains one of the least resolved issues in AI-supported foreign language writing. AI tools were widely reported to support brainstorming, paraphrasing, grammar checking, and coherence improvement, but these benefits were repeatedly accompanied by concerns about originality, authenticity, and excessive dependence. AI can make student writing more technically polished while also flattening personal voice, cultural nuance, and rhetorical distinctiveness, especially when automated systems privilege standardized or monolingual norms (AbdAlgane et al., 2025; Altamimi, 2025; Yilmaz Virlan & Tomak, 2025a). Pedagogically meaningful AI use should therefore be judged not only by textual polish, but also by whether learners retain authorship and interpretive control over their writing.

However, the literature does not portray students simply as passive users of AI. Some studies show that multilingual learners actively evaluated, adapted, or rejected AI suggestions when those suggestions threatened intended meaning, personal voice, or culturally grounded expression. In this sense, translanguaging and critical mediation may be understood as forms of writer agency rather than deficiency (Almashour et al., 2025; Altamimi, 2025). Issues of voice and identity are also closely tied to plagiarism, overreliance, and weak academic ownership, leading several authors to call for explicit AI literacy, institutional guidance, and instructor-mediated use (AbdAlgane et al., 2025; Altamimi, 2025; Yilmaz Virlan & Tomak, 2025b). AI adoption therefore cannot remain tool-centred. It requires pedagogy-centred regulation that protects originality, cultural situatedness, and critical judgement.

Discussion

The main contribution of this review lies in the distinction between textual improvement and writing development. Across the reviewed studies, AI tools such as ChatGPT, Grammarly, QuillBot, Wordtune, and neural machine translation were most consistently associated with grammar, vocabulary, sentence clarity, drafting efficiency, and error correction. These gains matter, especially for foreign language learners who struggle to begin, extend, and

revise academic texts. Yet they are also the areas where AI appears to provide its most reliable support, whereas evidence for its contribution to organization, argumentation, rhetorical control, and independent composing ability remains less consistent. This uneven pattern cautions against treating polished output as proof of learning. AI can improve the document before it improves the writer. Reported gains in motivation, confidence, and reduced anxiety should be interpreted with the same caution. They show that AI can make writing feel more manageable, but they do not, by themselves, demonstrate stronger independent writing competence.

This distinction explains why instructional design and critical AI literacy are central to AI-supported writing. When AI supports planning, paraphrasing, sentence reformulation, idea expansion, and revision within structured learning contexts, it can reduce immediate cognitive burden and make feedback more accessible. The same support, however, can bypass learning when students accept machine-generated suggestions without evaluating meaning, voice, or rhetorical fit. The key pedagogical issue is not whether AI is available. It is whether learners are required to make decisions about AI output. Students need to question, adapt, reject, and justify suggestions. Without that evaluative work, AI may produce better texts while leaving the learner's writing judgement underdeveloped.

The findings also recast the nature of lecturer expertise in AI-supported writing. AI can reduce part of the corrective workload, particularly in grammar, sentence-level revision, and surface accuracy. It cannot, however, reliably judge whether a text presents a defensible argument, preserves the writer's intended meaning, or responds appropriately to academic and cultural expectations. Lecturer guidance, critical AI literacy, and explicit attention to authorship are therefore needed to ensure that AI feedback becomes material for evaluation rather than a shortcut around evaluation. Without such mediation, AI may increase fluency and confidence while leaving the deeper work of writing development largely unresolved.

Conclusion

AI has become a significant resource in foreign language writing instruction in higher education, but its contribution remains uneven. The reviewed studies show that AI-supported tools are most consistently associated with linguistic accuracy, writing efficiency, feedback access, and selected aspects of text quality. Some studies also linked AI use to greater learner confidence and lower writing anxiety (Almusharraf et al., 2025; Mahdi et al., 2025; Sanosi, 2022). These benefits, however, were strongest in lower-order correction, drafting support, and immediate performance. They were weaker in relation to rhetorical development, independent authorship, and sustained composing ability. AI should therefore be understood as targeted support within the writing process, not as automatic writing improvement.

The educational significance of AI depends on how it is integrated into teaching and assessment. The most meaningful outcomes were reported when AI functioned as a scaffold for planning, paraphrasing, revision, and feedback while leaving judgement with learners and lecturers (Alwasidi & Al-Khalifah, 2025; Fathi & Rahimi, 2024; Yilmaz Virlan & Tomak, 2025b). Lecturer expertise remains central in evaluating authorship, coherence, argument quality, and ethical AI use (Alghannam, 2025; Taskiran & Goksel, 2022). Critical AI literacy, pedagogical guidance, and human oversight are therefore not supplementary concerns. They are the conditions under which AI-supported writing becomes educationally defensible.

The current evidence base is still limited by short-term designs, context-specific samples, and a strong focus on English. Research on other foreign languages, longitudinal writing development, and pedagogical models that combine AI scaffolding with sustained human guidance remains comparatively thin. Future studies should examine not only whether AI improves written products, but how learners develop judgement, originality, critical thinking, and ethical responsibility over time. The field now needs evidence on writing development, not only evidence on AI-assisted text production.

Author Contributions

Conceptualization, Nur Hafizah Abd Aziz and Harun Baharudin; methodology, Nur Hafizah Abd Aziz; validation, Harun Baharudin and Hazrati Husnin; formal analysis, Nur Hafizah Abd Aziz; writing (original draft preparation), Nur Hafizah Abd Aziz; writing (review and editing), Harun Baharudin and Hazrati Husnin; supervision, Harun Baharudin. All authors have read and agreed to the published version of the manuscript.

References

1. AbdAlgane, M., Ali, R., Othman, K., Ibrahim, I. Z. A., Alhaj, M. K. M., Ali, E. M. T., & Allehyani, F. S. (2025). Exploring AI Generated Texts vs. Human Written Texts in EFL Academic Writing: A Case Study of Qassim University in Saudi Arabia. *World Journal of English Language*, 16(2), 114–131. <https://doi.org/10.5430/wjel.v16n2p114>

2. Abdullah, M. Y. (2025). Probing into EFL students' perceptions about the impact of utilizing AI-powered tools on their academic writing practices. *Education and Information Technologies*, 30(15), 21189–21220. <https://doi.org/10.1007/s10639-025-13601-w>
3. Al Mahmud, F. (2023). *Investigating EFL Students' Writing Skills Through Artificial Intelligence: Wordtune Application as a Tool*. <https://doi.org/10.17507/jltr.1405.2>
4. Al Obaydi, L. H., & Pikhart, M. (2025). AI partner versus human partner: comparing AI-based peer assessment with human-generated peer assessment in examining writing skills. *Language Testing in Asia*, 15(1). <https://doi.org/10.1186/s40468-025-00375-8>
5. Al-Alami, S. (2024). Exploring the attitudes of EFL university instructors and students toward utilizing ChatGPT for acquiring writing fluency and accuracy skills. *International Journal of English Language and Literature Studies*, 13(3), 421–432. <https://doi.org/10.55493/5019.v13i3.5182>
6. Al-Alami, S. E. (2024). EFL Learners' Attitudes Towards Utilizing ChatGPT for Acquiring Writing Skills in Higher Education: A Case Study of Computing Students. *Journal of Language Teaching and Research*, 15(4), 1029–1038. <https://doi.org/10.17507/JLTR.1504.01>
7. Alghannam, M. S. M. (2025). Artificial Intelligence as a Provider of Feedback on EFL Student Compositions. *World Journal of English Language*, 15(2), 161–173. <https://doi.org/10.5430/wjel.v15n2p161>
8. Alharbi, W. (2023). The use and abuse of artificial intelligence-enabled machine translation in the EFL classroom: An exploratory study. *Journal of Education and E-Learning Research*, 10(4), 689–701. <https://doi.org/10.20448/jeelr.v10i4.5091>
9. Alkamel, M. A. A., & Alwagieh, N. A. S. (2024). Utilizing an adaptable artificial intelligence writing tool (ChatGPT) to enhance academic writing skills among Yemeni university EFL students. *Social Sciences and Humanities Open*, 10. <https://doi.org/10.1016/j.ssaho.2024.101095>
10. Almahboob, T. (2025). Exploring the Impact of AI Technologies on EFL Writing Proficiency. *Forum for Linguistic Studies*, 7(7), 270–282. <https://doi.org/10.30564/fls.v7i7.9952>
11. Almashour, M., Aldamen, H. A. K., & Jarrah, M. (2025). Algorithmic feedback and multilingual identity: Translanguaging practices in Jordanian EFL academic writing. *Social Sciences and Humanities Open*, 12. <https://doi.org/10.1016/j.ssaho.2025.102016>
12. Almusharraf, A., Bailey, D., Almusharraf, N., & Alotaibi, T. (2025). Students' perceptions of generative AI in EFL writing: Strategies, self-efficacy, satisfaction and behavioural intention. In *Australasian Journal of Educational Technology* (Vol. 2025, Number 5).
13. Al-Raimi, M., Mudhsh, B. A., Al-Yafaei, Y., & Al-Maashani, S. (2024). Utilizing artificial intelligence tools for improving writing skills: Exploring Omani EFL learners' perspectives. *Forum for Linguistic Studies*, 6(2). <https://doi.org/10.59400/fls.v6i2.1177>
14. Al-Shaboul, I. A., Al Rousan, R. M., Kalsoom, T., & A-Awawdeh, N. (2024). A critical examination of how AI-driven writing tools have impacted the content, style, and organization of foreign language undergraduates' writing: A survey of lecturers. *Research Journal in Advanced Humanities*, 5(4), 156–172. <https://doi.org/10.58256/pjrf6en83>
15. Alshehri, M. (2025). Integrating Automated Writing Evaluation into Saudi EFL Students' Writing Practice. *Arab World English Journal*, 16(1), 70–87. <https://doi.org/10.24093/awej/vol16no1.5>
16. Alsofyani, A. H., & Barzanji, A. M. (2025). The Effects of ChatGPT-Generated Feedback on Saudi EFL Learners' Writing Skills and Perception at the Tertiary Level: A Mixed-Methods Study. *Journal of Educational Computing Research*, 63(2), 431–463. <https://doi.org/10.1177/07356331241307297>
17. Altamimi, D. H. F. (2025). Unlocking Potential: Saudi EFL Male Students' Perspectives on AI Tools for Enhancing English Writing Proficiency. *Arab World English Journal*, 2025(Special Issue), 40–58. <https://doi.org/10.24093/awej/AI.3>
18. Alwasidi, M. A., & Al-Khalifah, K. S. (2025). Assessing the Impact of ChatGPT on EFL Students' Writing Productivity and Proficiency. *Journal of Language Teaching and Research*, 16(3), 986–995. <https://doi.org/10.17507/jltr.1603.29>
19. Alyami, A., Alotaibi, S., & Khan, W. (2025). Saudi EFL Learners' Perceptions of Using Artificial Intelligence and its Impact on their Writing Skills. *Arab World English Journal*, 16(1), 349–365. <https://doi.org/10.24093/awej/vol16no1.22>

20. Bani Amer, M. I., Rababah, L. M., & Rababah, M. A. (2025). Exploring ChatGPT as a Tool for Thesis Writing: Perspectives of EFL Supervisors in Jordanian Universities. *Journal of Language Teaching and Research*, 16(1), 33–42. <https://doi.org/10.17507/jltr.1601.04>
21. Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
22. El-Esery, A. M. (2025). Shifting Roles: Employing AI-driven Translation Engines to Enhance the Writing Proficiency of EFL Learners. *World Journal of English Language*, 15(6), 11–18. <https://doi.org/10.5430/wjel.v15n6p11>
23. Elmotri, B., Harizi, R., Boujlida, A., Elyasa, Y. M., Garrouri, S., Amri, F., Malik, F. H., & Al-Humari, M. A. (2025). The Impact of AI-Generated Feedback Explicitness (Generic vs. Specific) on EFL Students' Use of Automated Written Corrective Feedback. *Arab World English Journal*, 16(1), 384–402. <https://doi.org/10.24093/awej/vol16no1.24>
24. Fathi, J., & Rahimi, M. (2024). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in EFL learners: a qualitative study. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2024.2374772>
25. Guo, K., & Li, D. (2024). Understanding EFL students' use of self-made AI chatbots as personalized writing assistance tools: A mixed methods study. *System*, 124. <https://doi.org/10.1016/j.system.2024.103362>
26. Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>
27. Hwang, W. Y., Nurtantyana, R., Purba, S. W. D., Hariyanti, U., Indrihapsari, Y., & Surjono, H. D. (2023). AI and Recognition Technologies to Facilitate English as Foreign Language Writing for Supporting Personalization and Contextualization in Authentic Contexts. *Journal of Educational Computing Research*, 61(5), 1008–1035. <https://doi.org/10.1177/07356331221137253>
28. Jmaiel, H. A., Abukhait, R. O., Mohamed, A. M., Shaaban, T. S., Al-khresheh, M. H., & AL-Qadri, A. H. (2025). The role of ChatGPT in enhancing EFL students' ESP writing skills: an experimental study of gender and major differences. *Discover Education*, 4(1). <https://doi.org/10.1007/s44217-025-00700-6>
29. Mahdi, H. S., Milad, M., & Saeed, A. T. (2025). Evaluating the Effectiveness of AI Tools Across the Essay Writing Process: A Comparative Study of Pre-Writing, Drafting, and Post-Writing Stages. *International Journal of Computer-Assisted Language Learning and Teaching*, 15(1). <https://doi.org/10.4018/IJCALLT.391354>
30. Mahmud, F. Al. (2025). Examining the Efficacy of ChatGPT and Human-Derived Corrective Feedback in Addressing Grammatical Errors in Saudi EFL Students' Compositions. *World Journal of English Language*, 16(2), 258–270. <https://doi.org/10.5430/wjel.v16n2p258>
31. Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki. (2023). Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective. *International Journal of Educational Research Open*, 5. <https://doi.org/10.1016/j.ijedro.2023.100296>
32. Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2). <https://doi.org/10.1080/2331186X.2023.2236469>
33. Mohamed Shaffril, H. A., Samsuddin, S. F., & Abu Samah, A. (2021). The ABC of systematic literature review: the basic methodological guidance for beginners. *Quality & Quantity*, 55(4), 1319–1346. <https://doi.org/10.1007/s11135-020-01059-6>
34. Mohammad, T. (2024). Challenging Traditional EFL Writing Classroom Using AI Mediated Tool: A Paradigm Shift. *World Journal of English Language*, 14(2), 211–219. <https://doi.org/10.5430/wjel.v14n2p211>
35. Mohammad, T., Alzubi, A. A. F., & Idris Khan, S. (2023). Assessing The Effectiveness of Quillbot-Mediated Instruction in Enhancing EFL Students' Paraphrasing Skills. *Journal of Theoretical and Applied Information Technology*, 31, 24. <https://www.researchgate.net/publication/377363834>
36. Mohammad, T., Nazim, M., Alzubi, A. A. F., & Khan, S. I. (2024). Examining EFL Students' Motivation Level in Using QuillBot to Improve Paraphrasing Skills. *World Journal of English Language*, 14(1), 501–510. <https://doi.org/10.5430/wjel.v14n1p501>

37. Mohammed, S. J., & Khalid, M. W. (2025). Under the world of AI-generated feedback on writing: mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia*, 15(1). <https://doi.org/10.1186/s40468-025-00343-2>
38. Muslimin, A. I., Mukminatien, N., & Ivone, F. M. (2024). Evaluating Cami AI Across SAMR Stages: Students' Achievement and Perceptions in EFL Writing Instruction. *Online Learning Journal*, 28(2). <https://doi.org/10.24059/olj.v28i2.4246>
39. Sanosi, A. B. (2022). The Impact of Automated Written Corrective Feedback on EFL Learners' Academic Writing Accuracy. *Journal of Teaching English for Specific and Academic Purposes*, 10(2), 301–317. <https://doi.org/10.22190/JTESAP2202301S>
40. Sawangwan, S. (2024). ChatGPT vs Teacher Roles in Developing EFL Writing. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1). <https://doi.org/10.4018/IJCALLT.361235>
41. Taskiran, A., & Goksel, N. (2022). *Automated Feedback and Teacher Feedback: Writing Achievement in Learning English as a Foreign Language at a Distance*.
42. Tsai, C. Y., Lin, Y. T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12722-y>
43. 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>
44. Wale, B. D., & Kassahun, Y. F. (2024). The Transformative Power of AI Writing Technologies: Enhancing EFL Writing Instruction through the Integrative Use of Writerly and Google Docs. *Human Behavior and Emerging Technologies*, 2024. <https://doi.org/10.1155/2024/9221377>
45. Werdiningsih, I., Marzuki, Indrawati, I., Rusdin, D., Ivone, F. M., Basthomi, Y., & Zulfahreza. (2024). Revolutionizing EFL writing: unveiling the strategic use of ChatGPT by Indonesian master's students. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186x.2024.2399431>
46. Woo, D. J., Wang, D., & Guo, K. (2024). *Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process Everwrite: Human-AI Education*. <https://doi.org/10.13140/RG.2.2.22950.70722>
47. Xu, T., & Jumaat, N. F. (2024). ChatGPT-Empowered Writing Strategies in EFL Students' Academic Writing: Calibre, Challenges and Chances. *International Journal of Interactive Mobile Technologies*, 18(15), 95–114. <https://doi.org/10.3991/ijim.v18i15.49219>
48. Yılmaz Virlan, A., & Tomak, B. (2025a). A Q method study on Turkish EFL learners' perspectives on the use of AI tools for writing: Benefits, concerns, and ethics. *Language Teaching Research*, 29(7), 3190–3213. <https://doi.org/10.1177/13621688241308836>
49. Yılmaz Virlan, A., & Tomak, B. (2025b). AI tools for writing: a Q-method study with Turkish instructors of English. *Education and Information Technologies*, 30(12), 16997–17021. <https://doi.org/10.1007/s10639-025-13455-2>