Assessing L2 Student Writing in the AI Era: A Systematic Review on Challenges and Best Practices

Dau Thi Thanh Luy[1*], Lam Thuy Trang[1]

[1]University of Management and Technology Ho Chi Minh City, Vietnam
[*]Corresponding author's email: luy.dauthithanh@umt.edu.vn
https://orcid.org/0009-0004-7618-259X
https://doi.org/10.54855/ijte.26615

Received: 08/03/2025 Revision: 11/02/2026 Accepted: 28/02/2026 Online: 28/02/2026

ABSTRACT

This paper aims to provide a systematic review of the existing literature on the effects of AI on L2 (ESL/EFL) student writing and assessment. A comprehensive search was conducted through ScienceDirect, ERIC and Taylor & Francis Online, providing qualitative analysis of three major themes: the effects of AI on L2 writing, challenges in identifying AI-generated content, and practices for adapting writing assessment. PRISMA 2020 guidelines and the CASP checklist were used to select appropriate articles and assess validity, relevance and ethical issues. The review included 20 studies published between 2023 and 2025. The results show that AI offers numerous advantages, such as giving individualized feedback and support, improving writing quality and helpfully assisting the grading processes. However, issues related to academic integrity, originality and the difficulties in distinguishing between AI-created writing and student-written work have emerged. The paper proposes best practices for integrating AI into assessment frameworks in which academic integrity is still maintained while AI is utilized to improve learning outcomes.

Keywords: AI tools, EFL students, writing assessment, academic integrity

Introduction

Recently, rapid advances in artificial intelligence (AI) have not only created new opportunities but also posed challenges for educators and institutions, especially in balancing student learning outcomes with the benefits of AI tools (Thanh et al., 2023). In L2 (second/foreign language) writing contexts, educators effectively employ AI, such as ChatGPT, to provide students with personalized feedback (Dai et al., 2023; Cotton et al., 2024) and automated essay scoring systems to facilitate teachers' grading processes (Mizumoto & Eguchi, 2023). Additionally, AI technologies enhance the writing process for English learners through feedback and revision (Alharbi, 2023; Thangthong et al., 2024) and improve their writing quality (Marzuki et al., 2023; Mahapatra, 2024).

Nonetheless, the above-mentioned benefits could be undermined when students rely on AI to submit their assignments (Thanh et al., 2023). This presents significant challenges to the field

CITATION | Dau, T. T. L., & Lam, T. T. (2026). Assessing L2 Student Writing in the AI Era: A Systematic Review on Challenges and Best Practices. *International Journal of TESOL & Education*, *6*(1), 72-93. DOI: https://doi.org/10.54855/ijte.26615

of assessment (Luo, 2024). Concerns about academic integrity were raised when researchers reported over 85% of students used ChatGPT to do their homework (Westfall, 2023; Fakir et al., 2024), and one-third of tertiary students were dishonest in doing their assignments when they used AI tools to complete them (Sullivan et al., 2023). Therefore, questions of originality, authorship and fair assessment become a critical issue to be addressed (Luo, 2024). In fact, a few universities implemented bans on students' utilization of AI for classroom tasks and assessment due to their undecided AI policies (Sullivan et al., 2023). Such bans may inadvertently hinder innovation in pedagogy and fail to sustain the positive effects of AI on writing development (Graham, 2023). Instead of prohibiting these technologies, educators should instruct students to use them responsibly (Smerdon, 2024). This also presents a valuable opportunity to renew assessment methods, making them suitable with AI availability (Rudolph et al., 2023). However, the question remains: how can the new methods both ensure the objectives and outcomes of students' learning and maintain the effective support of AI (Crawford et al., 2023)? This paper continues to discuss the effects of AI on L2 writing and how teachers can overcome the challenges of identifying AI-generated content, so that they can propose the best ways to evaluate writing assignments. In this review, L2 writing is used as an umbrella term covering both ESL (English as a Second Language) and EFL (English as a Foreign Language) contexts. While these terms differ in sociolinguistic environments, they share common pedagogical concerns related to writing development and assessment. Notably, the majority of studies included in this review are situated in tertiary EFL settings, reflecting the current concentration of empirical research on AI and writing assessment in higher education. Accordingly, the findings are most directly transferable to university-level L2 writing contexts.

Literature Review

Common AI Writing Tools

Grammarly, a widely recognized Automated Writing Evaluation (AWE) tool, is known for its thorough grammar checks and style recommendations, helping students improve their writing by identifying errors that may otherwise be overlooked (Karyuatry, 2018; Ha & Ho, 2025). It provides a range of features, including grammar and punctuation correction, spell checking, plagiarism detection, and writing style assessment (Fitriana & Nurazni, 2022). This tool helps students save time and effort in editing their writing. This, in turn, contributes to their overall satisfaction with the tool (Ha & Ho, 2025). However, Grammarly may overlook context-specific errors, just focus on grammar, spelling, and clarity while failing to notice broader aspects of content, organization, and coherence of a piece of writing, or inaccurately flag correct elements as mistakes, leading to confusion and frustration among users (Yurika et al., 2023; Ha & Ho, 2025). In addition, users sometimes doubted the accuracy of Grammarly's suggestions, as certain corrections altered the intended meaning of their writing (Fitriana & Nurazni, 2022).

ChatGPT**,** an advanced chatbot released in November 2022, is commonly used in academic writing for grammar correction, paraphrasing, and expanding ideas, thereby enhancing the efficiency of students' written products and creating a less stressful classroom (Pham & Le, 2024). Mahapatra (2024) also notes that ChatGPT helps students generate focused ideas, improve sentence links, and use grammar more accurately. Besides, ChatGPT can deliver personalized feedback that helps to overcome language, time, and location constraints, effectively addressing the challenges in providing formative feedback in large classrooms (Mahapatra, 2024; Pham & Cao, 2025). Ultimately, ChatGPT can give scores that are closely similar to human evaluators when it comes to automated grading processes, especially in

formative assessments (Mizumoto & Eguchi, 2023; Tate et al., 2024), which optimizes teachers' time for other essential tasks such as planning lessons and supporting students (Owan et al., 2023). However, several researchers express their concerns regarding students' over-reliance on ChatGPT. They argue that excessive dependence on this technology may hinder the development of important writing capabilities, such as creativity, reasoning, and problem-solving skills (Marzuki et al., 2023; Pham & Le, 2024; Bui & Tong, 2025). In addition, ChatGPT demonstrates its technical limitations in understanding contexts and cultural affairs, as well as in handling questions that require high cognitive skills (Johinke et al., 2023; Kim et al., 2024). This indicates ChatGPT struggles with advanced writing skills, which presents a significant challenge for ESL learners. Last but not least, ethical concerns regarding the privacy of users' data and the trustworthiness of AI-generated content have also been raised for further investigation (Bui & Tong, 2025).

QuillBot is another popular writing tool well known for its user-friendly features and support for students' writing processes (Kurniati & Fithriani, 2022; Latifah et al., 2024). Several studies acknowledge QuillBot's ability to improve students' paraphrasing skills, increase vocabulary, correct grammar and vocabulary errors, and save students time and effort when writing (Fitria, 2022; Mohammad et al., 2024; Latifah et al., 2024). Nonetheless, Quillbot faces difficulties adapting to different students' learning levels and also blocks students' creativity (Mohamad et al., 2024). Sometimes the created sentences do not meet users' expectations, even failing to follow language structures, punctuation, and correct spelling (Fitria, 2022).

Effects of AI on Academic Integrity

As a result of the widespread use of AI, students easily commit plagiarism and misinterpret authorship of written work (Mizumoto & Eguchi, 2023; Thangthong et al., 2024; Maphoto et al., 2024; Bui & Tong, 2025; Pham & Cao, 2025). In addition, Generative AI tools such as ChatGPT can produce highly sophisticated texts that resemble human writing, posing challenges for teachers to ensure academic integrity and fair assessment (Cotton et al., 2024; Perkins et al., 2023; Herbold et al., 2023). Researchers also raise concerns that ChatGPT may hinder students' abilities in independent writing and original idea development, as they found that many university students use ChatGPT not only for proofreading and paraphrasing but also for generating content (Črček & Patekar, 2023). Therefore, clear guidelines for responsible AI use in academic settings are required to address these challenges (Perkins et al., 2023).

Complexities in Evaluating Students' Writing

Traditional assessments in higher education institutions have been significantly impacted since the public introduction of new AI tools due to concerns about academic integrity, cheating, and plagiarism (Moorhouse et al., 2023; Ibrahim, 2023). As AI-generated content is more accessible to AI (Herbold et al., 2023), and advanced AI technologies can write compelling essays on any topic (Perkins et al., 2023), educators are facing a range of complexities related to fair assessment and the authenticity of students' written work. One of the rising obstacles is teachers' inability to accurately distinguish AI-generated content from student-written texts. According to Waltzer et al. (2024), although instructors demonstrated a moderate success in identifying AI-generated essays, distinguishing AI-generated content remains a complex task. Similarly, Fleckenstein et al. (2024) found that both pre-service and experienced teachers faced challenges in evaluating the sources of texts written by AI and students.

Furthermore, the recent emerging research shows a variation in the effectiveness of AI detectors such as Open AI Detector, Crossplag, GPT-2 Output Detector, iThenticate, and Turnitin, and there remains inconsistency about the detection reliability (Perkins et al., 2023; Ibrahim, 2023;

Khalil & Er, 2023). According to Alexander et al. (2023), a large amount of AI-generated text remains undetected by plagiarism detection software, with even higher rates of undetected content when evaluated by human raters. The limitations of current assessment practices necessitate a reconsideration of writing evaluation criteria that emphasize process-based writing approaches (Fleckenstein et al., 2024; Cong-Lem et al., 2024; Waltzer et al., 2024).

Research Gaps

Although there is a number of research investigating AI-based tools for writing, their role in student writing assessment needs to be better understood. Existing studies demonstrate how AI-driven tools can improve the quality of students' writing, coherence, and feedback (Marzuki et al., 2023; Mahapatra, 2024; Chan & Hu, 2023). However, there is limited systematic analysis of how educators can integrate AI into L2 assessment frameworks without diminishing students' independent writing skills.

A major challenge in AI-assisted writing is its impact on academic integrity. As generative AI tools become more sophisticated, students increasingly rely on them for idea generation, summarization, and even full-text composition (Črček & Patekar, 2023). While studies discuss AI-driven plagiarism and concerns about students' declining originality (Bui & Tong, 2025; Herbold et al., 2023), research has yet to fully address how educators can effectively distinguish between AI-generated content and students' authentic work. Furthermore, AI detection tools are unreliable (Fleckenstein, 2024; Waltzer et al., 2024), making it difficult for teachers to maintain assessment integrity. More research is needed to explore practical strategies for ensuring authenticity in L2 student writing.

Additionally, the best practices for evaluating students' writing in the AI era remain underexplored. Current studies emphasize the importance of balancing AI with traditional writing instruction (Cotton et al., 2024), but there is still little research offering clear guidelines on how educators should adapt writing assessment to encourage critical thinking and originality. Therefore, addressing these gaps through a systematic review will provide educators with proper strategies for integrating AI in teaching, ensuring academic integrity and fair evaluation practices.

Research Questions

To address the research gaps, this review is guided by the questions below:

1. How do AI tools impact L2 student writing?
2. What challenges do teachers have when distinguishing between AI-created content and students' original writing products?
3. What are the best practices for evaluating L2 student writing in the age of AI?

Methods

Design of the Study

The study adheres to a Systematic Literature Review methodology by following PRISMA 2020 guidance. This method is often used to identify and review evidence from previous studies, summarize the current state of knowledge, identify gaps and priorities for future research, and highlight problems in primary studies that need improvement (Page et al., 2021). The current study aims to synthesize existing findings from relevant academic literature to better understand

the impact of AI on L2 student writing assessment and to suggest the most effective methods for this practice.

Search Strategy

A comprehensive literature search was conducted using three core academic databases: ScienceDirect, ERIC, and Taylor & Francis Online. In addition, ResearchGate and Google Scholar were used as a supplementary hand-searching source to locate open-access full texts and to conduct reference chaining for studies already identified through the primary databases.

To ensure comprehensive coverage, two Boolean search strings were developed based on the research questions. The first targeted studies on AI-assisted L2 writing and assessment practices, while the second focused on challenges related to AI-text detection, originality, and academic integrity.

Table 1.

Boolean search strings

String 1 (Themes 1 & 3): Writing & assessment	(“artificial intelligence” OR “AI” OR “generative AI”) AND (“academic writing” OR “essay writing” OR “writing skills”) AND (“assessment” OR “feedback” OR “evaluation”)
String 2 (Theme 2): Detection & originality	(“artificial intelligence” OR “AI” OR “generative AI”) AND (“academic writing” OR “student writing”) AND (“detection” OR “originality” OR “academic integrity” OR “plagiarism”)

Criteria for Selection and Omission

Peer-reviewed journals, conference proceedings, and academic books issued between 2023 and 2025 were selected. This period was chosen because it captures the rapid advances in AI for language education, providing up-to-date findings and relevant insights for the study. Only studies that addressed AI in relation to student writing, particularly in ESL and EFL contexts, were included. Research focused on AI in non-academic settings or technical evaluations of AI technologies outside the context of education was omitted. In addition, to ensure accessibility, the authors just chose the studies available in English.

To evaluate the standards of the included articles, the authors also used the Critical Appraisal Skills Program (CASP) checklist which is popularly employed in systematic reviews (Long et al., 2020). Although CASP was originally developed for qualitative research, selected core criteria (e.g., clarity of aims, methodological rigor, transparency of analysis, and ethical considerations) were adapted pragmatically and applied across qualitative, quantitative, and mixed-methods studies. The selected articles had to meet 8 out of 10 items in the following CASP checklist:

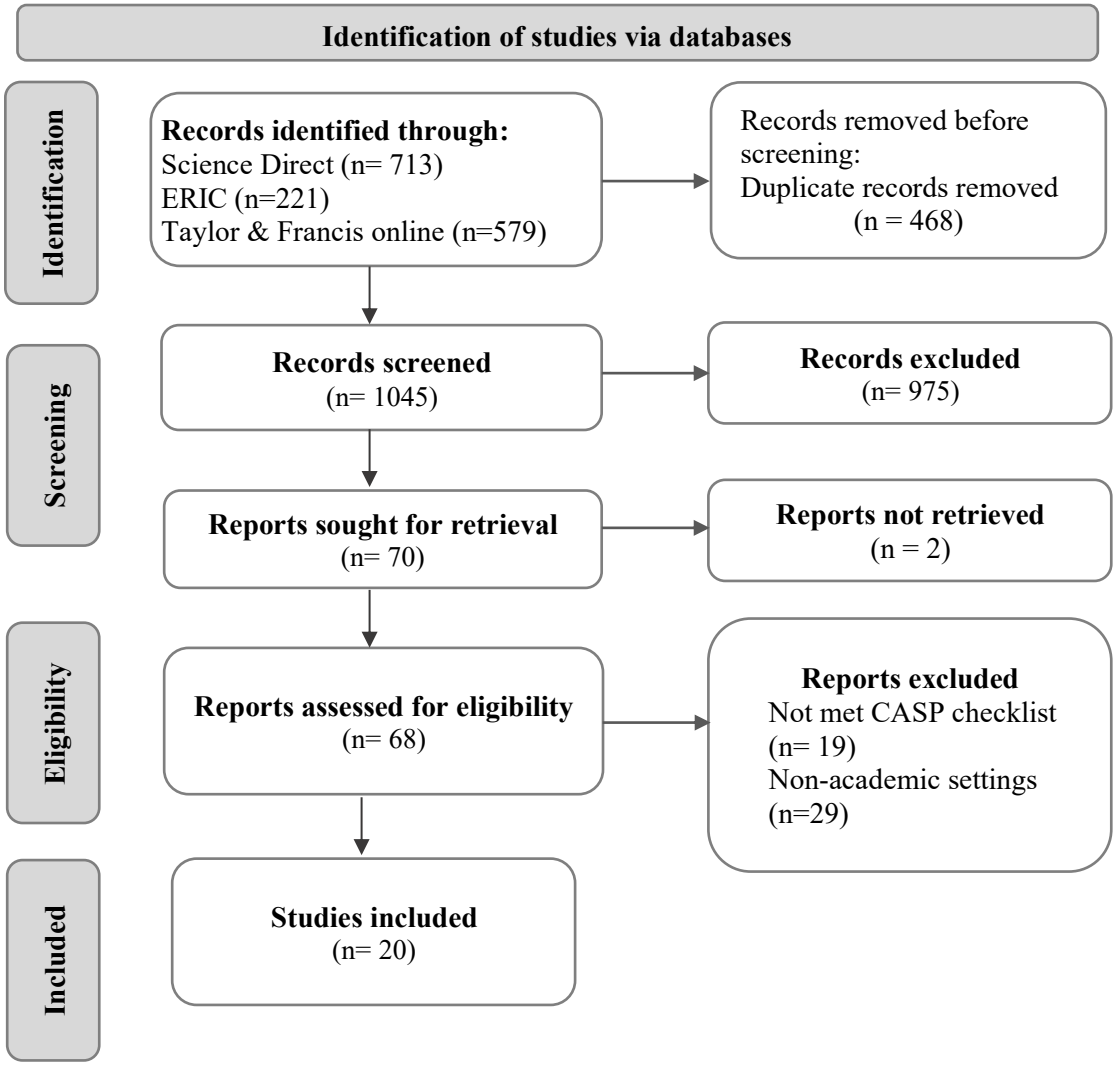
1. Was there a clear statement of the aims of the research?
2. Is a qualitative methodology appropriate?
3. Was the research design appropriate to address the aims of the research?
4. Was the recruitment strategy appropriate to the aims of the research?
5. Was the data collected in a way that addressed the research issue?
6. Has the relationship between researcher and participants been adequately considered?
7. Have ethical issues been taken into consideration?
8. Was the data analysis sufficiently rigorous?

9. Is there a clear statement of findings?
10. How valuable is the research?

A total of 1513 articles were first identified from the databases. After removing duplicate records and records excluded before screening, titles and abstracts were screened for relevance. Full-text articles were then assessed for eligibility based on the inclusion and omission criteria. A total of 20 studies were eligible for review. Figure 1 describes the procedure of the research using the PRISMA 2020 approach.

Figure 1.

PRISMA 2020 flowchart of article selection



Data Analysis

The literature selected was categorized into thematic areas deductively, based on the predefined research questions: 1) AI's effects upon student writing, (2) Challenges in identifying AI-generated texts and students' original writing, (3) Recommendations for assessing students' writing.

This study employed a qualitative thematic synthesis to integrate findings across the included studies. Two reviewers independently coded the extracted data using an initial deductive

codebook aligned with the research questions and prior literature on AI-assisted writing and assessment. During the coding process, inductive codes were added to capture emerging concepts not represented in the initial framework. In addition, any coding discrepancies were resolved through discussion until consensus was reached. Codes were then iteratively compared and grouped into higher-order themes through constant comparison across studies. For example, codes such as *personalized feedback*, *revision support*, and *enhancing writing performance* were synthesized into the theme *Advantages of AI writing tools*.

Each study was further examined to identify recurring themes, recommendations and reports on how AI tools are being implemented in writing assessment, especially in English education contexts. Special attention was given to the pedagogical challenges and ethical considerations highlighted by researchers, as well as the recommendations for adapting assessment practices to incorporate AI responsibly. Table 2 summarizes the 20 articles that meet the demands of inclusion criteria.

Table 2.

Summary of the included studies

Study No.	Author(s), Year	Title	Publication Venue	Publication Type	Indexing
1	Marzuki et al., 2023	The Impact of AI Writing Tools on the Content and Organization of Students' Writing: EFL Teachers' Perspective	*Cogent Education*	Journal article	Scopus (Q2)
2	Hossain et al., 2025	Exploring EFL Students' AI Literacy in Academic Writing: Insights into Familiarity, Knowledge and Ethical Perceptions	*Journal of Theoretical Educational Science*	Journal article	Scopus indexing not identified
3	Marghany, 2023	Using Artificial Intelligence-based Instruction to Develop EFL Higher Education Students' Essay Writing Skills	*CDELT Occasional Papers in the Development of English Education*	Journal article	Scopus indexing not identified
4	Mahapatra, 2024	Impact of ChatGPT on ESL Students' Academic Writing Skills: A Mixed Methods Intervention Study	*Smart Learning Environments*	Journal article	Scopus (Q1)
5	Kim et al., 2024	Exploring Students' Perspectives on Generative AI-assisted Academic Writing	*Education and Information Technologies*	Journal article	Scopus (Q1)
6	Thangthong et al., 2024	Navigating AI Writing Assistance Tools: Unveiling the Insights of Thai EFL Learners	*THAITESOL Journal*	Journal article	Scopus indexing not identified
7	Mizumoto and Eguchi, 2023	Exploring the Potential of Using an AI language Model for Automated	*Research Methods in Applied*	Journal article	Scopus (Q1)

		Essay Scoring	*Linguistics*		
8	Tate et al., 2024	Can AI Provide Useful Holistic Essay Scoring?	*Computers and Education: Artificial Intelligence*	Journal article	Scopus (Q1)
9	Chan and Hu, 2023	Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education	*International Journal of Educational Technology in Higher Education*	Journal article	Scopus (Q1)
10	Khampusaen, 2025	The Impact of ChatGPT on Academic Writing Skills and Knowledge: An Investigation of Its Use in Argumentative Essays	*LEARN Journal: Language Education and Acquisition Research Network*	Journal article	Scopus (Q2)
11	Cong-Lem et al., 2024	Academic Integrity in the Age of Generative AI: Perceptions and Responses of Vietnamese EFL Teachers	*Teaching English with Technology*	Journal article	Scopus (Q1)
12	Bui and Tong, 2025	The Impact of AI Writing Tools on Academic Integrity: Unveiling English Majored Students' Perceptions and Practical Solutions	*AsiaCALL Online Journal*	Journal article	Scopus indexing not identified
13.	Herbold et al., 2023	A Large-scale Comparison of Human-Written Versus ChatGPT-generated Essays	*Scientific Reports*	Journal article	Scopus (Q1)
14	Fleckenstein et al., 2024	Do Teachers Spot AI? Evaluating the Detectability of AI-Generated Texts among Student Essays	*Computers and Education: Artificial Intelligence*	Journal article	Scopus (Q1)
15	Waltzer et al., 2024	Can You Spot the Bot? Identifying AI-generated Writing in College Essays	*International Journal for Educational Integrity*	Journal article	Scopus (Q1)
16	Alexander et al., 2023	Who Wrote This Essay? Detecting AI-generated Writing in Second Language Education in Higher Education	*Teaching English with Technology*	Journal article	Scopus (Q1)
17	Perkins et al., 2023	Detection of GPT-4 Generated Text in Higher Education: Combining Academic Judgment and Software to Identify	*Journal of Academic Ethics*	Journal article	Scopus (Q1)

		Generative AI Tool Misuse			
18	Ibrahim, 2023	Using AI-based Detectors to Control AI-assisted Plagiarism in ESL Writing: "The Terminator versus the Machines."	*Language Testing in Asia*	Journal article	Scopus (Q1)
19	Khalil and Er, 2023	Will ChatGPT Get You Caught? Rethinking Plagiarism Detection	*Springer*	Conference proceeding	Scopus
20	Bordalejo et al., 2025	"Scarlet Cloak and the Forest Adventure": A Preliminary Study of the Impact of AI on Commonly Used Writing Tools	*International Journal of Educational Technology in Higher Education*	Journal article	Scopus (Q1)

Quality appraisal outcomes of included studies based on CASP checklist.

Supplementary Table S1.

Quality appraisal CASP scores

Study No.	Author(s), Year	Research Method	CASP Score
1	Marzuki et al., 2023	Qualitative	9/10
2	Hossain et al., 2025	Quantitative	9/10
3	Marghany, 2023	Mixed methods	8/10
4	Mahapatra, 2024	Mixed methods	9/10
5	Kim et al., 2024	Qualitative	9/10
6	Thangthong et al., 2024	Qualitative	8/10
7	Mizumoto and Eguchi, 2023	Quantitative	9/10
8	Tate et al., 2024	Quantitative	9/10
9	Chan and Hu, 2023	Quantitative	8/10
10	Khampusaen, 2025	Mixed methods	8/10
11	Cong-Lem et al., 2024	Quantitative	9/10
12	Bui and Tong, 2025	Mixed methods	8/10
13.	Herbold et al., 2023	Mixed methods	9/10
14	Fleckenstein et al., 2024	Quantitative	9/10
15	Waltzer et al., 2024	Quantitative	8/10
16	Alexander et al., 2023	Qualitative	8/10
17	Perkins et al., 2023	Quantitative	9/10
18	Ibrahim, 2023	Quantitative	8/10
19	Khalil and Er, 2023	Quantitative	8/10
20	Bordalejo et al., 2025	Quantitative	9/10

Findings and Discussion

Overview of the Included Studies

The literature is predominantly situated in EFL/ESL contexts, with most studies conducted at

universities, while evidence from primary, secondary, or non-tertiary settings remains limited. In addition, most studies concentrated in Asian and European contexts and relatively little represented from other regions. Furthermore, longitudinal investigations are scarce, as most studies adopt cross-sectional or short-term designs. These patterns suggest that, although research on AI and L2 writing assessment is growing, important gaps remain in terms of educational levels, regional diversity, and long-term empirical evidence.

Effects of AI on Student Writing

The findings show that using AI tools in teaching, learning and evaluation of writing has both positive and negative effects. Table 3 provides a summary of studies exploring the effects of AI on students' writing, focusing on their advantages, drawbacks and the effects in academic integrity and assessment.

Table 3.

Summary of the effects of AI on students' writing

Advantages of AI writing tools	**Study No.**
Enhancing students' writing performance	(1), (2), (4), (3), (5), (6), (10), (12)
Providing quick and personalized feedback and support	(5), (4), (9)
Demonstrating a notable degree of accuracy and reliability in automated scoring	(7), (8)
Drawbacks of AI writing tools	**Study No.**
Concerns about over-reliance, preventing critical thinking and problem-solving skills	(1), (5), (6), (11), (12)
Occasional ineffectiveness, making errors	(5), (6), (9)
Producing hallucination	(5), (13), (16)
Lack of contextualization and pedagogical skills	(5)
Effects on academic integrity and assessment	**Study No.**
Ethical issues: academic fraud, misconduct and plagiarism	(2), (6), (9), (10), (11), (12), (13), (16), (19)
Challenges to distinguish human written texts from AI-generated content	(11), (13), (14), (15), (16), (17), (18), (19), (20)
Potential of unfairness in assessment	(11), (13), (15), (17), (20)

Advantages of AI Writing Tools

As can be seen from the results, AI tools provide substantial benefits in enhancing students' writing performance. AI-driven tools like Grammarly, ChatGPT, and Quillbot help improve students' writing quality, structure, and coherence (Marzuki et al., 2023; Marghany, 2023; Mahapatra, 2024; Thangthong et al., 2024; Kim et al., 2024; Khampusaen, 2025; Bui & Tong, 2025). In addition, when comparing learning with Grammarly to traditional methods, Marghany (2023) found that Grammarly significantly improved students' essay writing skills, particularly

in grammar, tenses, punctuation, and revision. This observation aligns with Ha and Ho (2025), who highlighted the effectiveness of Grammarly in saving students' time and effort in the writing and editing processes. While this shows promise for learning to write in general, the same benefits may not apply to different students' learning styles and strategies in writing. Additionally, most studies lack investigation into the long-term effects on students' independent writing ability.

Another widely recognized benefit of AI tools in teaching student writing is their ability to provide quick and personalized feedback and support (Chan & Hu, 2023; Mahapatra, 2024; Kim et al., 2024). As noted by Naz and Robertson (2024), ChatGPT-3 demonstrates a good understanding of paragraph structure and can provide timely and individualized comments that encourage deeper student engagement. This aligns with Krashen's (1982) theories of language acquisition, which emphasize authentic, interactive communication as a key to language development. Mahapatra (2024) added that in large classes, where providing individualized feedback is challenging for teachers, ChatGPT could help, overcoming language, time, and location constraints to deliver personalized evaluations for students. However, the effectiveness of AI-generated feedback remains under debate in the literature (Thangthong et al., 2024). Further research is needed to evaluate whether they can replace or complement teachers' feedback in meaningful ways.

Finally, AI tools demonstrate notable accuracy and reliability in automated scoring by analyzing essays, reports, and other written assignments and providing comments on syntax, grammar, spelling, and organization (Mizumoto & Eguchi, 2023; Tate et al., 2024). For instance, GPT-based AES (Automated Essay Scoring) systems can achieve high accuracy and reliability, significantly supporting teacher evaluations in formative assessment (Mizumoto & Eguchi, 2023). Similarly, Tate et al. (2024) compared multiple versions of ChatGPT's automated scoring of secondary-school essays across three established corpora against expert human ratings and found no statistically significant differences in the examined datasets or scoring criteria. Although it is not recommended to use AI for important summative assessments (e.g., end-of-course tests, standardized tests, achievement tests) or to substitute for instructors' evaluation, this technology is proposed to be a valuable assistance during the first stages of the writing process and in other progress evaluations (Tate et al., 2024). Therefore, the potential of AI in writing assessment and giving feedback can enhance the efficiency and fairness of student work evaluations (Mizumoto & Eguchi, 2023).

Drawbacks of AI Writing Tools

Throughout the literature, a significant area of concern for educators and researchers is students' over-reliance on AI tools (Marzuki et al., 2023; Thangthong et al., 2024; Kim et al., 2024; Cong-Lem et al., 2024; Bui & Tong, 2025). AI over-dependence potentially weakens students' ability to think critically and find solutions for problems, inhibits students' confidence in their writing capabilities, and reduces their capacity to be involved in cognitive activities independently and achieve true knowledge (Marzuki et al., 2023; Thangthong et al., 2024; Cong-Lem et al., 2024). These points are further elaborated by Johinke et al. (2023) and Bui and Tong (2025), who note that teachers' fear of students using AI not just to refine language in their writing but also to generate ideas, which severely affects students' creativity and original thought. However, these claims are largely based on teachers' perceptions rather than longitudinal evidence, so the degree of the problem remains uncertain.

Technical limitations of AI tools are another area of concern. Studies highlight issues such as occasional ineffectiveness, making errors, and hallucination (Chan & Hu, 2023; Thangthong et al., 2024; Kim et al., 2024). For instance, Grammarly may fail to notice context-specific errors

or inaccurately flag correct items as mistakes (Thangthong et al., 2024). In addition, AI tools still lack understanding of culture and pedagogies, which are critical in academic settings (Kim et al., 2024). These observations align with Ha and Ho (2025), who expressed that the shortcomings could lead to confusion among students. Such limitations need to be addressed thoroughly.

Emerging Challenges in Academic Integrity and Writing Assessment

The findings also reveal major concerns that AI tools encourage dishonesty in assignment completion and threaten academic integrity. Ethical issues related to academic fraud, misconduct, and plagiarism are raised by several researchers (Hossain et al., 2025; Thangthong et al., 2024; Cong-Lem et al., 2024; Herbold et al., 2023; Chan & Hu, 2023; Khampusaen, 2025; Bui & Tong, 2025; Alexander et al., 2023; Khalil & Er, 2023). The concerns have been discussed regarding the possibility of AI introducing inaccurate information, copying available materials, creating original work lacking human knowledge, causing bias in AI-created content (Frye, 2023), and misrepresenting the authorship and originality of students' written products (Perkins et al., 2023). Such concerns are increased by recent AI advancements, which can now produce writing comparable to, or in some cases more persuasive than, human essays (Perkins et al., 2023; Herbold et al., 2023). Studies indicate that a significant proportion of students use AI to find ideas, summarize, paraphrase, proofread, and even write full assignments (Črček & Patekar, 2023; Fakir et al., 2024). Some scholars attribute this reliance to students' lack of original ideas, low motivation, and limited linguistic skills (Cong-Lem et al., 2024), while others point to the accessibility of AI as an easy way to higher grades for their assignments (Herbold et al., 2023). Obviously, this cannot be the purpose of education, and teachers are facing difficulty retaining academic honesty and fair evaluation. However, the literature is not entirely in agreement: Fakir et al. (2024) found no difference in the final performance scores of students who used AI or not in their writing, and many students believed that using AI for assignments was not a form of cheating. This viewpoint aligns with Smerdon (2024), who argues that there may be no risk of academic violations given the neutral impact of AI on academic performance. Therefore, ongoing research is needed to clarify the contradiction.

Challenges in Distinguishing AI-generated Content from Students' Original Work

Table 4 summarizes studies on challenges in identifying AI-generated texts.

A key challenge in L2 writing assessment today involves differentiating between AI-generated texts and human-written submissions. Fleckenstein et al. (2024) found that both pre-service and experienced teachers struggled with identifying the original source of the texts. Pre-service teachers had difficulty regardless of text quality. In contrast, experienced teachers showed higher accuracy in identifying polished AI-created texts but struggled with flawed ones. This indicates that experienced teachers may be unaware that AI can generate low-quality texts with grammar and spelling mistakes. In terms of text quality assessment, neither group favored AI-created texts over those produced by students, though experienced teachers tended to give higher ratings to high-quality AI-created texts. Importantly, text quality did not help teachers to better identify the origin of the texts, emphasizing the need to familiarize teachers with AI's capabilities (Fleckenstein et al., 2024).

Waltzer et al. (2024) further showed that recognition ability remains inconsistent across participating groups. University teachers accurately identified only 70% of essays written by ChatGPT; students achieved 60%, and ChatGPT itself got 63% accuracy. The participating teachers also expressed concerns about assigning take-home written tasks and grading students on how good and creative they are. Therefore, this raises questions about the accuracy of the

current assessment and academic standards.

Table 4.
Summary of challenges in identifying AI-generated texts and assessment suggestions

Author(s), Year	Key findings	Recommendation
Fleckenstein et al., 2024	Less experienced teachers could not accurately recognize the origins of the examined texts. Experienced teachers struggled to find the origins of flawed texts but were able to identify polished ones.	Rethinking the assessment strategies and criteria, considering AI presence; training teachers on the awareness of AI content.
Waltzer et al., 2024	University lecturers accurately recognized 70% of essays produced by ChatGPT, students achieved 60%, ChatGPT itself got 63%.	Designing assessments encouraging cognitive engagement; doing more practice on identifying AI created content.
Alexander et al., 2023	AI detectors accurately identify fully AI-created texts but struggle with mixed human-AI content. No fully reliable methods exist for differentiating AI and human produced content, leading to potential misjudgments.	Training digital literacy for teachers, students and administrators; revising assessment policies and procedures; training teachers to effectively identify features of an AI-written work.

Alexander et al. (2023) reported similar challenges among ESL lecturers. Accuracy ranged from 33% to 66%, with lecturers often associating AI-generated texts with advanced vocabulary, complex sentences, and error-free writing, while attributing low language proficiency, repetitive words and grammatical structures, weak connections, or failure to follow the correct essay organization to students. However, such features are not always reliable indicators, since repetition of vocabulary and grammar is believed to be typical in AI-generated writing. Therefore, these findings underscore the variation in teachers' judgements and the lack of suitable criteria for distinguishing AI from human work.

Given the obstacles, researchers widely recommended AI literacy training for teachers, especially on how to identify features of AI-created content (Alexander et al., 2023; Fleckenstein et al., 2024), and teachers themselves need to do more practice to get their grading fair and accurate in students' assignments (Waltzer et al., 2024). AI literacy training sessions not only build teachers' confidence and capabilities in addressing AI-related challenges but also help teachers and students effectively apply this technology in their teaching and learning (Bui & Tong, 2025).

The Efficacy of AI Detection Tools

Several studies have examined the effectiveness of AI detectors. Table 5 summarizes the efficacy of these tools.

Table 5.
Summary of the efficiency of AI detectors

Author(s), Year	Key findings	Recommendation
Perkins et al., 2023	Turnitin effectively recognized 91% of AI-generated submissions but marked just 54.8% of AI-created content. Teachers detected only 54.5% of the identified cases for academic fraud.	Adapting assessment strategies with the presence of AI; providing training for teachers and students to ensure academic integrity; improving detection software.
Ibrahim, 2023	GPT 2 Output and Crossplag detectors could identify AI-generated texts with a relatively high degree of accuracy but inconsistencies in the accuracy levels remain throughout the data records.	Double-checking the flagged results with another detection tool; combining the suggested results of the detector with traditional methods; improving the detectors' accuracy.
Khalil and Er, 2023	Among 50 essays written by ChatGPT that were analyzed, iThenticate and Turnitin classified 40 as highly original.	Improving the detector tools; encouraging critical thinking assignments; training AI literacy, academic integrity and responsible use; giving clear AI policies.
Bordalejo et al., 2025	Turnitin and GPTZero showed significant shortcomings in distinguishing AI-created and human-paraphrased content.	Requiring careful interpretation of the detectors' results; developing clear AI use guidelines and policies.

Across studies, AI detection tools show highly variable performance depending on text type, degree of human-AI collaboration, and evaluation criteria, limiting their validity as standalone evidence of misconduct.

Perkins et al. (2023) found that Turnitin could effectively detect 91% of AI-generated submissions but identified only 54.8% of fully AI-generated content, while faculty themselves recognized only 54.5% of cases of academic fraud. Similarly, Ibrahim (2023) reported that GPT-2 Output Detector and Crossplag achieved high accuracy in flagging AI-created content but still lacked consistency in the detection reliability, suggesting cross-checking flagged texts across multiple detectors and adopting supplementary verification methods such as student interviews.

Other studies identified even greater challenges. Khalil and Er (2023) found that Turnitin and iThenticate rated around 80% of the AI-generated essays (40 out of 50 examined essays) as highly original. Bordalejo et al. (2025) further added significant shortcomings of Turnitin and GPTZero in distinguishing AI-created and human-paraphrased content. Especially when evaluating modified content, Turnitin tended to skip all the interventions while GPTZero remained inconsistent. Obviously, there is variation in the accuracy of AI detectors, raising concerns about the trustworthiness of the tools currently used in several universities. As a result, scholars recommend carefully interpreting the detection results, improving the reliability of detection tools, and clear guidelines on AI use to avoid unfair punishments to students (Ibrahim, 2023; Khalil & Er, 2023; Bordalejo et al., 2025)

Practices for Evaluating Students' Writing

Concerns about the effects of AI and questions about how to evaluate students' writing properly in the age of AI have prompted educators and institutions worldwide to rethink traditional assessment practices. The three main approaches drawn from literature include redesigning assessments, collaborating between humans and AI, and establishing clear institutional policies.

Redesigning Assessments

Perkins et al. (2023) recommended designing assignments that incorporate group-based projects, require distinctive organization, pre-submit topics for approval, or use data specifically collected to minimize opportunities for students relying on AI-completed assignments. Such measures ensure that students engage deeply with the subject matter and foster genuine learning. Testing tasks also integrate AI by requiring students to analyze and evaluate AI outputs, promoting students' responsible use of the tools. The testing tasks should also require students to critically evaluate AI-generated content and prompt engineering to help students learn how to work with AI effectively and ethically. Similarly, Moorhouse et al. (2023) suggested that integrating AI into student assessments is vital to enhancing students' AI literacy.

Holistic and process-oriented assessment approaches that prioritize higher-order thinking skills and student engagement are also emphasized. Assignments that focus on completion procedures, such as submitting notes, drafts, or proposals for feedback, or structuring the assignment into small steps, are recommended (Moorhouse et al., 2023). In addition, portfolio-based teaching and flipped-classroom approaches can enhance student engagement and provide real-world learning experiences (Cong-Lem et al., 2024). Flipped activities, such as assigning multimedia tasks for students to prepare in advance or giving presentations, also help teachers save time for reviewing and providing feedback on students' work (Rudolph et al., 2023).

In addition, since AI technologies often thrive in low-order tasks but struggle with creating, evaluating, and applying complex concepts (Thanh et al., 2023), assessments should therefore focus on tasks that emphasize high-order skills (Moorhouse et al., 2023). Furthermore, it's necessary to revise evaluation criteria, integrating skills-based assessment such as oral examinations, interviews, and presentations to complement written assignments and in-class writing. These practices help teachers to verify if students understand their work and reduce academic dishonesty (Waltzer et al., 2024; Fleckenstein et al., 2024; Bordalejo et al., 2025).

Human and AI Collaboration

Given the limitations of AI detectors and teachers' challenges in distinguishing AI-generated content from students' original writing, it is important to use a balanced assessment method to uphold academic honesty and fair evaluation (Perkins et al., 2023). Ibrahim (2023) highlights the importance of using multiple detectors to cross-check flagged texts and combining these tools with traditional methods, such as interviewing students or comparing submissions with in-class work, to mitigate false positives. Educators should remain aware of the limitations of AI detectors and exercise their own judgment to ensure fair and accurate assessments. Likewise, Pham and Le (2024) suggest that AI tools should act as a powerful complement, not a replacement, to teachers because human evaluators remain critical for contextual judgment, particularly for hybrid texts that blend human and AI content.

Institutional Policies and Guidelines to Promote AI Literacy and Responsible Use

As noted by Moorhouse et al. (2023), just under half of the leading universities currently provide publicly available standards for communicating with students about AI use, planning assessments, and maintaining integrity in education. Therefore, there is an emerging need for

clear guidelines and policies on AI use and assessment across the broader landscape (Khalil & Er, 2023; Alexander et al., 2023; Bordalejo et al., 2025). These practices aim to support teachers in determining what might be considered an inappropriate use of AI, create protocols for students to acknowledge or cite AI use, make assignments more difficult for AI to complete alone, and discuss how AI can be used in assignment tasks (Moorhouse et al., 2023). Calls for global collaboration stress the need for adaptable frameworks and AI literacy initiatives that prepare educators and learners to endgame with these tools responsibly. Another critical issue is how to define originality in students' work in light of the increasing use of AI and student collaboration. Luo (2024) argues that current policies often regard AI-supported work as less authentic, so updating policies to treat AI as a collaborative tool can foster a better understanding of originality. Developing such policies requires broad participation of teachers, students, and administrators (Hong, 2023; Cong-Lem et al., 2024). Teachers and students should be involved in building policies to ensure a culture of trust and support (Luo, 2024).

Implications

Drawing primarily on evidence from tertiary EFL/ESL writing contexts, the following implications are mainly intended for higher education.

For students, given the AI's ability to improve writing proficiency, tertiary L2 students can exploit these technologies as a helpful tool at different stages of writing: brainstorming, drafting, and checking errors in final products, but not to create original thought. Students also need to critically examine AI-suggested content, verify the reliability of AI-provided knowledge, and recognize the ethical implications of AI-created writing. These strategies can help students use AI effectively to improve writing in English while keeping creativity, critical thinking, and academic integrity.

For teachers, providing careful guides on AI usage, critically evaluating AI suggestions, and organizing AI awareness workshops and classroom discussions on responsible AI use are necessary to build students' confidence in using the tools without violating academic rules. In addition, given the limited ability to differentiate between AI-produced content and students' original work, teachers require sustained professional development in AI literacy to familiarize themselves with common AI writing patterns, detector limitations, and ethical assessment practices. This can be done by joining AI literacy training workshops, interacting with various AI tools while evaluating students' assignments, critically examining AI writing patterns and students' writing styles, and sharing observations and experiences with colleagues.

For writing assessment design, several strategies are drawn. First, to prevent students from turning to AI to complete their assignments, process-oriented assessments that include portfolio writing, pre-submitted multiple drafts for teachers' approval (e.g., outlines, drafts, revisions), and reflective commentaries are highly recommended. These stages should be included in the grading criteria. Second, assessment frameworks should include higher-order thinking tasks that require students to think critically, combine ideas, and apply complex contexts because these skills are considered shortcomings of AI tools. Third, to ensure fairness and accuracy, a collaboration between AI and teachers is needed in the assessment processes. To this end, teachers need to cross-check flagged texts using multiple detectors and using supplementary methods such as interviews or in-class assignments. Fourth, teachers can develop AI-resistant assessments by requiring students to defend their written products with oral presentations or reflections. These approaches also help reduce dishonesty and misconduct in academic settings and, importantly, raise students' awareness of integrity in education.

For institutional policies, university administrators should collaborate with multiple stakeholders, e.g., teachers, students, curriculum designers, AI tools developers, and their counterparts to build clear and transparent policies for AI use in education, clarifying acceptable practices, expectations for original works, and the consequences of misconduct. Once developed, the policies should be clearly communicated through student handbooks, university websites, and official communication channels. Both students and teachers should also receive training on these policies and AI literacy to ensure understanding and proper implementation.

Conclusion

This systematic review synthesized evidence from 20 peer-reviewed studies published between 2023 and 2025, following PRISMA 2020 guidelines, to examine the effects of AI on L2 student writing and assessment. The findings suggest that AI tools are perceived as beneficial for enhancing students' writing efficiency, but also raise concerns about excessive dependence on these advanced technologies, including the loss of students' creativity and critical thinking skills, violations of academic integrity, and difficulties in ensuring fairness and correctness under current assessment methods. The results underscore the need to transform assessment approaches given the easy accessibility of AI support. The best way is to embrace AI as a potential supporting tool in the assessment process. Redesigning assessment methods, encouraging AI responsible use, and equipping students and teachers with AI literacy, especially on how to distinguish AI-created content from students' original writing, are essential to avoid academic violations.

Despite the valuable findings, this study has several limitations. First, since this is a secondary research analysis, it uses the findings and perspectives of the previous studies, which may limit the generalizability of the conclusions due to the limitations of geographic diversity and education levels. Second, although database-specific search strings and screening procedures were reported, the review was limited to three major databases, which may have excluded relevant studies indexed elsewhere. Third, the corpus comprised a mix of qualitative, quantitative, and mixed-methods designs; while CASP criteria were applied pragmatically across study types with a defined inclusion threshold, the tool was not originally designed for cross-design appraisal, which may affect comparability of quality judgements. Furthermore, under the emerging challenges of AI on assessment, there might be new solutions and updated AI policies from universities around the world, which could affect the validity of the current insights. Finally, conclusions regarding AI detection tools should be interpreted cautiously, as detector performance is rapidly evolving and highly sensitive to task type, text manipulation, and software version. Therefore, on-going research is needed to keep the results updated. Future research should explore the development of innovative assessment methods, cross-regional studies that extend beyond dominant higher-education contexts, and AI's long-term effects in L2 writing competence.

Acknowledgments

We would like to send many thanks to valued researchers for their great contribution to the review. We are also grateful to editors and reviewers for their valuable feedback on the article.

References

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher education. *Teaching English with Technology, 23*(2), 25–43. https://doi.org/10.56297/BUKA4060/XHLD5365

Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. *Education Research International,* 2023(1), Article 4253331. https://doi.org/10.1155/2023/4253331

Bordalejo, B., Pafumi, D., Onuh, F., Khalid, A. K. M. I., Pearce, M. S., & O'Donnell, D. P. (2025). *Scarlet cloak and the forest adventure: A preliminary study of the impact of AI on commonly used writing tools*. *International Journal of Educational Technology in Higher Education, 22*(6), 1-25. https://doi.org/10.1186/s41239-025-00505-5

Bui, T. T. U., & Tong, T. V. A. (2025). The impact of AI writing tools on academic integrity: Unveiling English-majored students' perceptions and practical solutions. *AsiaCALL Online Journal*, 16(1), 83–110. https://doi.org/10.54855/acoj.251615

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, Article 43. https://doi.org/10.1186/s41239-023-00411-8

Cong-Lem, N., Tran, T. N., & Nguyen, T. T. (2024). Academic integrity in the age of generative AI: perceptions and responses of Vietnamese EFL teachers. *Teaching English with Technology*, 24(1), 28-47. https://doi.org/10.56297/FSYB3031/MXNB7567

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Crawford, J. A., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice, 20*(3), Article 02. https://doi.org/10.53761/1.20.3.02

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.2023.17379

Dai, Y., Liu, A., & Lim, C. P. (2023). Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education. *Procedia CIRP, 33rd CIRP Design Conference*. Elsevier, *119*, 84-90 https://doi.org/10.1016/j.procir.2023.05.002

Fakir, S. A., Marnaoui, S., & Al Anqodi, H. A. (2024). Written assignments and generative artificial intelligence: Challenges and considerations for English education major students at A'Sharqiyah University, Oman. *Arab World English Journal, 15*(4), 22–38. https://doi.org/10.24093/awej/vol15no4.2

Fitria, T. N. (2022). Avoiding plagiarism of students' scientific writing by using the QuillBot paraphraser. *ELSYA: Journal of English Language Studies, 4*(3), 252–262. https://doi.org/10.31849/elsya.v4i3.9917

Fleckenstein, J., Meyer, J., Jansen, T., Keller, S. D., Köller, O., & Möller, J. (2024). Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays. *Computers and Education: Artificial Intelligence, 6*(5), 100209. https://doi.org/10.1016/j.caeai.2024.100209

Frye, B. L. (2023). Should using an AI text generator to produce academic writing be plagiarism? *Fordham Intellectual Property, Media & Entertainment Law Journal, 33*(4), 946. https://ir.lawnet.fordham.edu/iplj/vol33/iss4/5

Graham, S. S. (2023). Post-process but not post-writing: Large language models and a future for composition pedagogy. *Composition Studies, 51*(1), 162–168. https://files.eric.ed.gov/fulltext/EJ1390327.pdf

Ha, Y. N., & Ho, N. P. (2025). EFL postgraduate students' perceptions on the use of Grammarly and peer feedback to improve their academic writing skills. *International Journal of TESOL & Education*, *5*(1), 25-49. https://doi.org/10.54855/ijte.25512

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, *13*, Article 18617. https://doi.org/10.1038/s41598-023-45644-9

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, *5*(1), 37-45. https://doi.org/10.61414/jeti.v5i1.103

Hossain, Z., Çelik, Ö., & Hiniz, G. (2025). Exploring EFL students' AI literacy in academic writing: Insights into familiarity, knowledge, and ethical perceptions. *Journal of Theoretical Educational Science (Kuramsal Eğitimbilim Dergisi), 18*(1), 157–181. https://hdl.handle.net/11630/12556

Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing: "The Terminator versus the machines." *Language Testing in Asia, 13*(1), Article 46. https://doi.org/10.1186/s40468-023-00260-2

Johinke, R., Cummings, R., & Di Lauro, F. (2023). Reclaiming the technology of higher education for teaching digital writing in a post-pandemic world. *Journal of University Teaching & Learning Practice*, 20(2), Article 01. https://doi.org/10.53761/1.20.02.01

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies, 30*(1), 1265-1300. https://doi.org/10.1007/s10639-024-12878-7

Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking plagiarism detection. In P. Zaphiris & A. Ioannou (Eds.), *Learning and collaboration technologies: HCII 2023* (Lecture Notes in Computer Science, *14040*, 442–457). Springer. https://doi.org/10.1007/978-3-031-34411-4_32

Khampusaen, D. (2025). The impact of ChatGPT on academic writing skills and knowledge: An investigation of its use in argumentative essays. *LEARN Journal: Language Education and Acquisition Research Network, 18*(1), 963-988. https://doi.org/10.70730/PGCQ9242

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press.

Kurniati, E. Y., & Fithriani, R. (2022). Post-graduate students' perceptions of Quillbot utilization in English academic writing class. *Journal of English Language Teaching and Linguistics, 7*(3), 437-451. https://doi.org/10.21462/jeltl.v7i3.852

Latifah, S., Muth'im, A., & Nasrullah, N. (2024). The use of QuillBot in academic writing. *Journey: Journal of English Language and Pedagogy, 7*(1), 110–121. https://doi.org/10.33503/journey.v7i1.872

Long, H. A., French, D. P., & Brooks, J. M. (2020). Optimising the value of the Critical Appraisal Skills Programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Research Methods in Medicine & Health Science, 1*(1), 31-42. https://research.manchester.ac.uk/en/publications/optimising-the-value-of-the-critical-appraisal-skills-programme-c/

Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. *Assessment & Evaluation in Higher Education*, *49*(5), 651–664. https://doi.org/10.1080/02602938.2024.2309963

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11,* Article 09. https://doi.org/10.1186/s40561-024-00295-9

Maphoto, K. B., Sevnarayan, K., Mohale, N. E., Suliman, Z., Ntsopi, T. J., & Mokoena, D. (2024). Advancing students' academic excellence in distance education: Exploring the potential of generative AI integration to improve academic writing skills. *Open Praxis*, *16*(2), 142–159. https://doi.org/10.55982/openpraxis.16.2.649

Marghany, M. M. (2023). Using artificial intelligence-based instruction to develop EFL higher education students' essay writing skills. *CDELT Occasional Papers in the Development of English Education*, *82*(1), 219-240. https://www.researchgate.net/publication/373432641_Using_artificial_intelligence-based_instruction_to_develop_EFL_higher_education_students'_essay_writing_skills

Marzuki, U., 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), 2236469. https://doi.org/10.1080/2331186X.2023.2236469

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*, Article 100050. https://doi.org/10.1016/j.rmal.2023.100050

Mohammad, T., Alzubi, A. A. F., Nazim, M., & Khan, S. I. (2024). Evaluating the effectiveness of QuillBot in improving students' paraphrasing skills: Teachers' voices. *Journal of Theoretical and Applied Information Technology*, *102*(6), 2556–2565. https://jatit.org/volumes/Vol102No6/25Vol102No6.pdf

Moorhouse, B. L., Yeo, M., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open, 5*(2), 100151. https://doi.org/10.1016/j.caeo.2023.100151

Naz, I., & Robertson, R. (2024). Exploring the feasibility and efficacy of ChatGPT-3 for personalized feedback in teaching. *European Journal of e-Learning, 22*(2), 98-111. https://doi.org/10.34190/ejel.22.2.3345

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *EURASIA Journal of Mathematics, Science and Technology Education, 19*(8), Article em2307. https://doi.org/10.29333/ejmste/13428

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* Article n71. https://doi.org/10.1136/bmj.n71

Perkins, M., Hickerson, D., Roe, J., Postma, D., & McGaughran, J. (2023). Detection of GPT-4 generated text in higher education: Combining academic judgment and software to identify generative AI tool misuse. *Journal of Academic Ethics, 22,* 89-113. https://doi.org/10.1007/s10805-023-09492-6

Pham, M. T., & Cao, T. X. T. (2025). The Practice of ChatGPT in English Teaching and Learning in Vietnam: A Systematic Review. *International Journal of TESOL & Education*, *5*(1), 50-70. https://doi.org/10.54855/ijte.25513

Pham, V. P. H., & Le, A. Q. (2024). ChatGPT in Language Learning: Perspectives from Vietnamese Students in Vietnam and the USA. *International Journal of Language Instruction*, *3*(2), 59–72. https://doi.org/10.54855/ijli.24325

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Smerdon, D. (2024). AI in essay-based assessment: Student adoption, usage, and performance. *Computers and Education: Artificial Intelligence, 7*, Article 100288. https://doi.org/10.1016/j.caeai.2024.100288

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching, 6*(1), 31-40. https://doi.org/10.37074/jalt.2023.6.1.17

Tate, T. P., Steiss, J., Bailey, D., Graham, S., Moon, Y., Ritchie, D., Tseng, W., & Warschauer, M. (2024). Can AI provide useful holistic essay scoring? *Computers and Education: Artificial Intelligence, 7*, Article 100255. https://doi.org/10.1016/j.caeai.2024.100255

Thanh, B. N., Vo, D. T. H., Nhat, M. N., Pham, T. T. T., Trung, H. T., & Xuan, S. H. (2023). Race with the machines: Assessing the capability of generative AI in solving authentic assessments. *Australasian Journal of Educational Technology*, *39*(5), 59–81. https://doi.org/10.14742/ajet.8902

Thangthong, P., Phiromsombut, J., & Imsa-ard, P. (2024). Navigating AI writing assistance tools: Unveiling the insights of Thai EFL learners. *THAITESOL Journal, 37*(1), 111–131. https://so05.tci-thaijo.org/index.php/thaitesoljournal/article/view/270168

Waltzer, T., Pilegard, C., & Heyman, G. D. (2024). Can you spot the bot? Identifying AI-generated writing in college essays. *International Journal for Educational Integrity, 20(2024)*, Article 11. https://doi.org/10.1007/s40979-024-00158-3

Westfall, C. (2023, January 28). Educators battle plagiarism as 89% of students admit to using OpenAI's ChatGPT for homework. *Forbes*. https://www.forbes.com/sites/chriswestfall/2023/01/28/educators-battle-plagiarism-as-89-of-students-admit-to-using-open-ais-chatgpt-for-homework/

Biodata

Dau Thi Thanh Luy is currently a lecturer at the University of Management and Technology, Ho Chi Minh City. She has more than 12 years of experience in teaching English to all ages, designing curricula, and conducting assessments. She also serves as a Speaking Examiner for Cambridge Assessment English, Cambridge University Press, and Assessment. She is actively involved in many international conferences on TESOL and language teaching, both as a presenter and an attendee. Her research interests include the integration of technology in English language teaching, remote teaching, and language assessment.

Lam Thuy Trang is currently a Deputy Head of the Department of English, University of Management and Technology, Ho Chi Minh City. She obtained a Master's degree in TESOL in 2009. She has been teaching English at the university level for more than 15 years. She is also particularly involved in research and professional work on program development, language testing, and assessment. Her many interests include professional development, curriculum development, language teaching, and assessment.