

Artificial Intelligence in Academic Writing in Higher Education: A Systematic Literature Review of Pedagogical Integration, Assessment Practices, and Teachers' Perspectives

Saidatul Akmar Zainal Abidin., *Noor Hanim Rahmat

Akademi Pengajian Bahasa, Universiti Teknologi MARA, Shah Alam, Malaysia

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.100700779>

Received: 01 August 2026; Accepted: 06 August 2026; Published: 12 August 2026

ABSTRACT

Artificial intelligence (AI) is rapidly transforming academic writing in higher education by enhancing teaching practices, writing assessment, and student learning. Despite the growing adoption of AI-powered tools such as ChatGPT, Grammarly, PaperPal, and Automated Writing Evaluation (AWE) systems, existing research remains fragmented across pedagogical, assessment, and teacher-related perspectives. This study systematically reviews the current evidence on AI in academic writing using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines. Guided by four research questions formulated through the PICO framework, a comprehensive search of Scopus and Web of Science identified 236 records, from which 24 peer-reviewed studies published between 2021 and 2026 met the inclusion criteria following screening, eligibility assessment, and quality appraisal. The findings reveal three dominant themes: (1) AI integration in academic writing teaching and learning, (2) AI-assisted assessment, feedback, and writing evaluation, and (3) teachers' perspectives, ethical issues, and AI adoption. Collectively, the reviewed studies demonstrate that AI enhances writing quality, learner engagement, self-regulated learning, and formative assessment through timely and personalized feedback. However, AI remains limited in evaluating higher-order writing competencies, including critical thinking, originality, and contextual reasoning, reinforcing the continued importance of human expertise in writing assessment. The review further highlights the influence of teachers' perceptions, ethical concerns, and institutional readiness on successful AI implementation. The study contributes a comprehensive synthesis of current research by integrating pedagogical, assessment, and human perspectives through the TPACK, Assessment for Learning, and Technology Acceptance Model frameworks. These findings provide practical implications for educators, institutions, and policymakers while identifying priorities for future research on responsible and pedagogically sound AI integration in academic writing within higher education.

Keywords: artificial intelligence, academic writing, higher education, writing assessment, pedagogical integration, systematic literature review

INTRODUCTION

Background of the study

Artificial intelligence (AI) has rapidly transformed higher education by reshaping teaching, learning, and assessment practices, particularly in the domain of academic writing. The emergence of generative AI applications, such as ChatGPT, Grammarly, PaperPal, and Automated Writing Evaluation (AWE) systems, has expanded opportunities for students to receive immediate feedback, improve writing quality, and engage in more personalized learning experiences. These technologies support multiple stages of the writing process, including idea generation, drafting, revision, editing, and self-assessment, while also assisting educators in providing timely and consistent feedback. Consequently, AI is increasingly viewed as a valuable pedagogical resource that can enhance writing instruction and formative assessment when integrated with appropriate teaching strategies. As higher education institutions continue to adopt AI technologies, there is growing interest in understanding

how these tools influence academic writing practices, assessment approaches, and educators' roles within digitally enhanced learning environments.

Statement of Problem

Despite the increasing adoption of AI in higher education, several challenges remain regarding its effective implementation in academic writing. Existing studies have reported concerns related to the reliability of AI-generated feedback, academic integrity, plagiarism, ethical use, transparency, and students' potential overreliance on AI-generated content. Although AI has demonstrated considerable potential to improve writing performance and reduce instructors' assessment workload, it cannot fully evaluate higher-order writing skills such as critical thinking, creativity, disciplinary reasoning, and contextual understanding. Moreover, teachers' willingness to adopt AI varies according to their perceptions of its usefulness, ease of use, institutional support, and professional preparedness. These issues indicate that technological advancement alone is insufficient; successful implementation requires sound pedagogical integration, responsible assessment practices, and adequate institutional policies that support educators in using AI effectively.

Research Gap

Although the body of research on AI in education has expanded considerably, the literature remains fragmented across instructional, assessment, and human dimensions. Many previous studies have focused on a single aspect of AI use, such as automated feedback, teachers' perceptions, or AI-assisted assessment, with limited attempts to synthesize these perspectives into a unified understanding.

Furthermore, there is a lack of comprehensive systematic literature reviews that simultaneously examine AI integration in academic writing teaching and learning, AI-assisted assessment and writing evaluation, and teachers' perspectives, ethical issues, and AI adoption within higher education. To address this gap, this study employs the PRISMA framework to systematically review empirical studies published between 2021 and 2026 and synthesizes the evidence into these three interconnected themes.

By integrating pedagogical, assessment, and human perspectives, this review provides a comprehensive overview of the current state of AI in academic writing, identifies emerging research trends, and offers evidence-based recommendations for future research, educational policy, and classroom practice.

Research Questions

Research questions are important to a systematic literature review (SLR). The questions provide the foundation and direction for the entire review process. A well-defined research question ensures that the literature search is exhaustive and systematic, covering all relevant studies that address key aspects of the topic (Kitchenham, 2007). The research questions in this SLR are formulated based on the PICo framework. This framework is a mnemonic style used to formulate research questions, particularly in qualitative research, proposed by Lockwood et al. (2015). PICo stands for Population, Interest, and Context. This SLR is done to answer the four research questions below;

RQ1 How are artificial intelligence technologies integrated into academic writing teaching and learning in higher education?

RQ2 How are artificial intelligence technologies integrated into academic writing teaching and learning in higher education?

RQ3 What are educators' and pre-service teachers' perceptions, ethical concerns, and experiences regarding the adoption of artificial intelligence in academic writing?

LITERATURE REVIEW

Theoretical Framework

Assessment for Learning (AfL)

Assessment for Learning (AfL), proposed by Black and Wiliam (1998), is a formative assessment theory that emphasizes using assessment as a continuous process to enhance learning rather than merely measuring students' achievement. Unlike traditional summative assessment, AfL encourages timely, constructive, and actionable feedback that enables learners to identify their strengths, recognize areas requiring improvement, and actively engage in revising their work. The theory promotes self-assessment, peer assessment, and reflective learning, thereby fostering learner autonomy and self-regulation throughout the learning process. Within the context of academic writing, AfL provides a strong theoretical foundation for integrating artificial intelligence (AI) into writing assessment because AI-powered tools, such as ChatGPT, Grammarly, PaperPal, and Automated Writing Evaluation (AWE) systems, can deliver immediate and personalized feedback on grammar, coherence, organization, vocabulary, citation practices, and writing style. Such real-time feedback supports iterative revision, allowing students to refine their writing before final submission. Moreover, AI-assisted assessment aligns with AfL's emphasis on continuous improvement by enabling students to monitor their progress and make informed revisions independently. Rather than replacing instructors, AI complements teacher feedback by increasing the frequency, accessibility, and consistency of formative assessment, ultimately promoting writing development, self-regulated learning, and improved academic writing performance in higher education.

Technological Pedagogical Content Knowledge (TPACK)

Technological Pedagogical Content Knowledge (TPACK), developed by Mishra and Koehler (2006), is a comprehensive framework that explains how teachers can effectively integrate technology into teaching and learning by combining three interrelated forms of knowledge: content knowledge (CK), pedagogical knowledge (PK), and technological knowledge (TK). The framework emphasizes that successful technology integration occurs not through technological expertise alone, but through the dynamic interaction of these knowledge domains, enabling educators to design meaningful and contextually appropriate learning experiences. In the context of academic writing in higher education, TPACK provides a robust theoretical foundation for understanding the pedagogical integration of artificial intelligence (AI) tools such as ChatGPT, Grammarly, PaperPal, and Automated Writing Evaluation (AWE) systems. These technologies can support brainstorming, drafting, revision, feedback, and collaborative writing; however, their educational effectiveness depends on teachers' ability to integrate them with sound pedagogical strategies and disciplinary writing objectives. TPACK therefore highlights the critical role of educators in selecting appropriate AI tools, designing AI-supported learning activities, and ensuring that technology enhances rather than replaces the writing process. By aligning AI capabilities with instructional methods and academic writing outcomes, the TPACK framework supports learner-centered pedagogy, promotes meaningful technology integration, and enables educators to maximize the educational benefits of AI while maintaining academic rigor and writing quality.

Technology Acceptance Model (TAM)

Technology Acceptance Model (TAM), proposed by Davis (1989), is one of the most influential theories for explaining individuals' acceptance and adoption of new technologies. TAM posits that users' intention to adopt a technology is primarily determined by two key constructs: perceived usefulness, which refers to the degree to which individuals believe that using a technology will enhance their job performance, and perceived ease of use, which reflects the extent to which the technology is perceived as effortless to learn and operate. In the context of higher education, TAM provides a valuable theoretical lens for understanding teachers' perspectives toward the integration of artificial intelligence (AI) in academic writing and AI-assisted assessment. Educators are more likely to adopt AI tools such as ChatGPT, Grammarly, PaperPal, and Automated Writing Evaluation (AWE) systems when they perceive these technologies as improving teaching efficiency, enhancing feedback quality, reducing assessment workload, and supporting students' writing development. Conversely, concerns about academic integrity, reliability, bias, transparency, and the potential overreliance of students on AI-generated content may negatively influence teachers' acceptance of these technologies. Therefore, TAM effectively

explains the factors shaping educators' attitudes, behavioural intentions, and actual use of AI in writing instruction and assessment. By emphasizing perceived usefulness and ease of use, the model justifies why teachers' acceptance is a critical determinant of the successful pedagogical integration of AI and the effective implementation of AI-assisted assessment practices in higher education.

Conceptual Framework

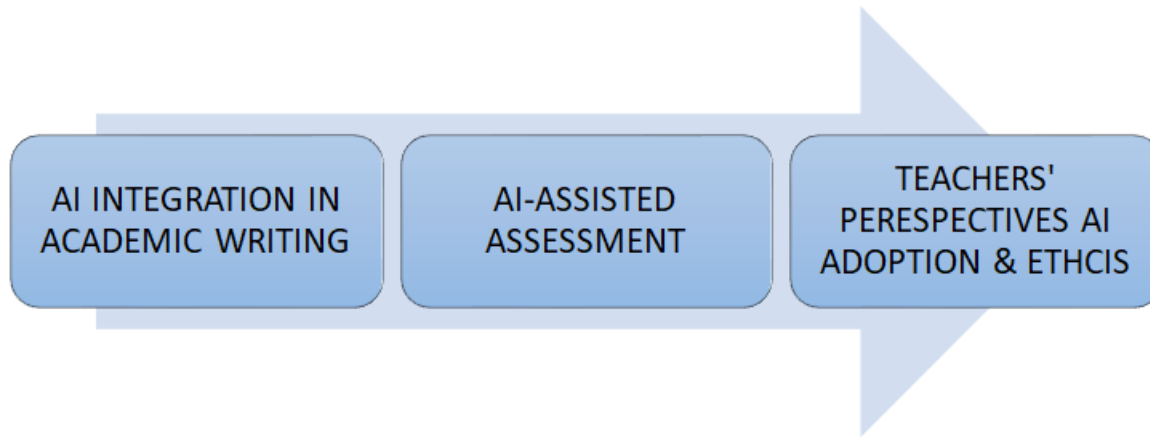


Figure 1- Conceptual Framework

The conceptual framework (refer to Figure 1) for this study is based on three interconnected dimensions of artificial intelligence (AI) use in academic writing in higher education: AI integration in academic writing, AI-assisted assessment, and teachers' perspectives on AI adoption and ethics, as illustrated in the framework. These dimensions are theoretically anchored in the Technological Pedagogical Content Knowledge (TPACK) Framework (Mishra & Koehler, 2006), Assessment for Learning (AfL) (Black & Wiliam, 1998), and the Technology Acceptance Model (TAM) (Davis, 1989). The framework proposes that the effective integration of AI into academic writing practices depends on educators' ability to combine technological knowledge, pedagogical approaches, and content expertise. Through the TPACK lens, AI tools such as ChatGPT, Grammarly, and automated writing platforms can be incorporated into writing instruction to facilitate idea generation, drafting, revision, and feedback processes. Thus, pedagogical integration serves as the foundation for understanding how AI technologies are embedded within academic writing instruction in higher education.

The second component of the framework focuses on AI-assisted assessment, which is supported by the principles of Assessment for Learning (AfL). Black and Wiliam (1998) argue that assessment should promote learning through timely feedback, self-regulation, and continuous improvement. In this context, AI-powered assessment tools provide immediate feedback on grammar, coherence, argumentation, and overall writing quality, enabling students to revise and improve their work more effectively.

The third dimension, teachers' perspectives on AI adoption and ethics, is anchored in the Technology Acceptance Model (TAM). According to Davis (1989), technology adoption is influenced by users' perceptions of usefulness and ease of use. Applied to higher education, teachers' acceptance of AI is likely to shape their willingness to integrate AI into writing instruction and assessment practices. At the same time, concerns related to academic integrity, ethical use, transparency, and overreliance on AI influence adoption decisions. Collectively, these three theories provide a comprehensive framework for systematically examining how AI is integrated into academic writing, how it transforms assessment practices, and how educators perceive its pedagogical and ethical implications.

MATERIAL AND METHODS

For conducting systematic literature reviews, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach is a widely accepted standard that guarantees transparency, completeness, and consistency throughout the procedure by Page et al. (2021). guide how to systematically identify, screen, and

include studies in their review. In the context of this SLR, the writers involved in this article act as researchers and work together to analyse the articles deemed relevant for the title. Two important databases, Web of Science and Scopus, were used in this analysis because of their wide coverage and robustness.

The PRISMA approach is organized into four key stages: (i) identification, (ii) screening, (iii) eligibility, and (iv) data abstraction. In the identification phase, databases are searched to locate all relevant studies. The screening phase then involves evaluating these studies against predefined criteria to eliminate irrelevant or low-quality research. During the eligibility phase, the remaining studies are thoroughly assessed to confirm they meet the inclusion criteria. Finally, data abstraction focuses on extracting and synthesizing data from the included studies, which is essential for deriving meaningful and reliable conclusions. This structured method ensures that the systematic review is conducted with rigor, leading to trustworthy results that can guide future research and practice. The PRISMA framework lists three main stages for analysis: (i) identification, (ii) screening, (iii) eligibility, (iv) data abstraction and analysis, as well as (v) quality appraisal.

Identification

In accordance with the PRISMA framework, the identification phase represents a critical starting point in the systematic literature review process, as it is designed to ensure exhaustive coverage of relevant studies while reducing the risk of selection bias. In this study, a structured and comprehensive search was conducted across two leading bibliographic databases, Scopus and Web of Science (WoS), using the keywords “challenges in academic writing assessment,” “universities,” and “AI technology.” Table 1 below shows the search string.

These databases were deliberately selected because of their extensive multidisciplinary coverage, stringent indexing criteria, and strong representation of high-quality, peer-reviewed journals. The search process resulted in the retrieval of 136 records from Scopus and 100 records from Web of Science, yielding a combined total of 236 records prior to the removal of duplicates.

The higher number of results obtained from Scopus can be attributed to its broader inclusion of journals and stronger coverage in the social sciences, whereas WoS contributed a more targeted collection of highly cited and methodologically rigorous studies. As shown in Table 1, the identification of 236 records reflects the growing academic interest in writing within higher education and highlights the conceptual diversity of research spanning linguistic, pedagogical, and psychological perspectives. Employing multiple databases enhances the methodological robustness of the review by minimizing database-specific bias and increasing the likelihood of capturing diverse research traditions and geographical contexts.

Additionally, the use of carefully designed search terms ensured sensitivity to terminological variation while maintaining alignment with the review’s objectives. Overall, this transparent and systematic identification process establishes a strong foundation for subsequent screening and eligibility stages, consistent with PRISMA standards and expectations for rigorous, replicable research

Table 1 The search string.

Scopus	TITLE-ABS-KEY ((challenge* OR issue* OR difficulty* OR barrier*) AND ("academic writing" OR "writing assessment" OR "writing evaluation" OR "writing skills") AND (teacher* OR educator* OR faculty OR instructor*) AND (univers* OR "higher education" OR "tertiary education") AND ("AI technology" OR "artificial intelligence" OR "machine learning" OR "automated assessment" OR "digital tools"))
Web of Science	TITLE-ABS-KEY ((challenge* OR issue* OR difficulty* OR barrier*) AND ("academic writing" OR "writing assessment" OR "writing evaluation" OR "writing skills") AND (teacher* OR educator* OR faculty OR instructor*) AND (univers* OR "higher education" OR "tertiary education") AND ("AI technology" OR "artificial intelligence" OR "machine learning" OR "automated assessment" OR "digital tools")) Retrieved July 2026

Screening

During the screening stage of the systematic literature review, potentially relevant studies were evaluated to ensure their alignment with the four established research questions. This phase focused on refining the initial pool of records by removing duplicates and applying predefined inclusion and exclusion criteria. Following the initial review, 172 publications were excluded for failing to meet the scope and relevance requirements of the study, leaving 64 (Scopus=21; Wos; 43) articles for further consideration, as outlined in Table 2. The primary inclusion criterion was the type of literature, with priority given to scholarly sources that offer substantive and empirical insights. As a result, materials such as book reviews, book series, meta-syntheses, meta-analyses, conference proceedings, and book chapters not directly related to the core focus of the review were excluded. To ensure consistency and relevance to contemporary academic discourse, only peer-reviewed publications written in English and published between 2021 and 2026 were retained. Additionally, a duplication check was done but no repeated records were found across the databases. This systematic and criteria-driven screening process enhanced the focus and quality of the evidence base, ensuring that only the most relevant and methodologically sound studies progressed to the subsequent stages of analysis.

Eligibility

For the third step, known as eligibility, a total of 64 articles have been prepared. All articles' titles and key content were thoroughly reviewed at this stage to ensure that the inclusion requirements were fulfilled and fit into the present study with the current research aims. Finally, 24 articles are available for review (see Table 4).

Table 2 -The selection criterion is searching

Criterion	Inclusion	Exclusion
Language	English	Non-English
Time line	2021 – 2026	< 2021
Literature type	Journal (Article)	Conference, Book, Review
Publication Stage	Final	In Press
Subject Area	Academic Writing Assessment, Higher Education	Besides Academic Writing Assessment, Higher Education

Data Abstraction and Analysis

An integrative analysis was used as one of the assessment strategies in this study to examine and synthesize a variety of research designs (quantitative, qualitative, and mixed methods). To develop themes, the authors carefully analyzed 64 publications for relevant material related to the study's topics. They examined the methodology and research results of significant studies on PIU and mental health in children. The authors collaborated to develop themes based on the evidence presented throughout the study. They kept a log of their data analysis process to document relevant perspectives, analyses, or ideas. The authors ensured consistency in theme design by comparing the results and discussing disagreements between themselves. This expert review phase established the validity of each subtheme by ensuring clarity, importance, and suitability (Refer to Table 4).

Quality of Appraisal

According to Kitchenham and Charters (Kitchenham, 2007), the researchers need to assess the quality of the research chosen. The next stage is applying quality assessment suggested by Abouzahra et al., (2020), which consists of six questions complying to a scoring procedure for evaluating each criteria. The possible ratings are

"Yes" (Y) with a score of 1 if the criterion is fully met, "Partly" (P) with a score of 0.5 if the criterion is somewhat met but contains some gaps or shortcomings, and "No" (N) with a score of 0 if the criterion is not met at all.

Table 3- Possible ratings for 6 questions by Abouzahra et al., (2020),

No	Question
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?

Each writer independently assesses the study according to these criteria, and the scores are then totalled across all experts to determine the overall mark. For a study to be accepted for the next process, the total mark, derived from summing the scores from all three experts, must exceed 3.0. This threshold ensures that only studies meeting a certain quality standard proceed further. The results are shown in Table 4.

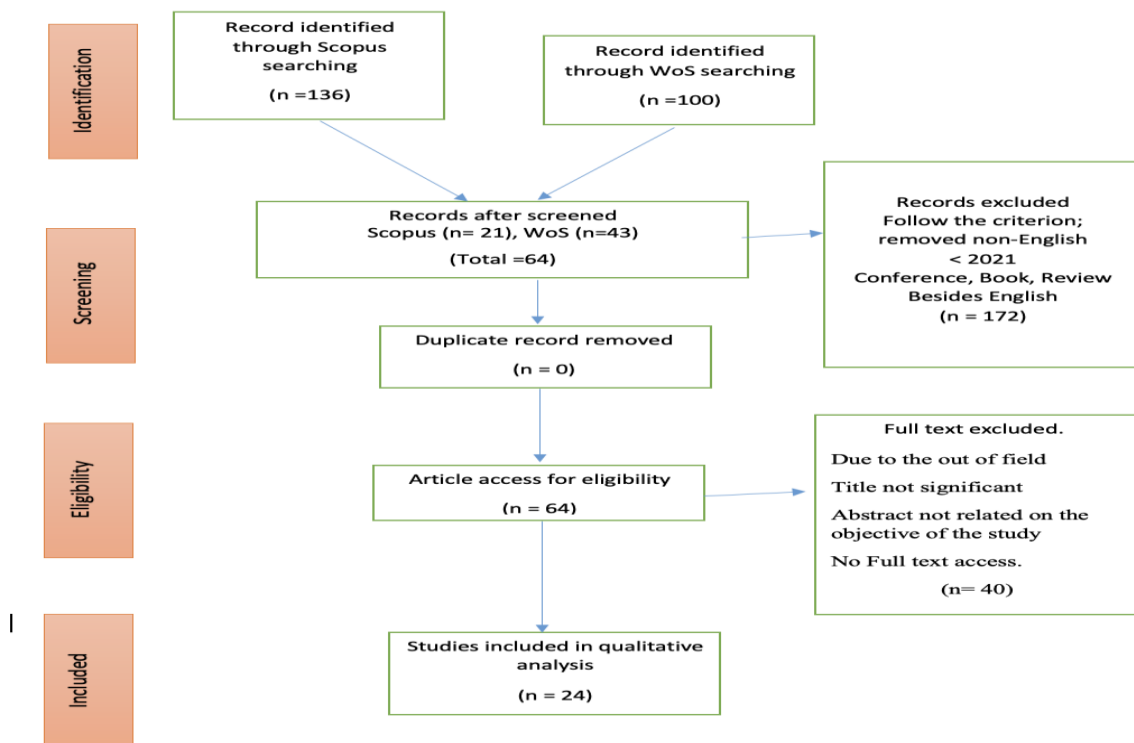


Figure 2. Flow diagram of the proposed searching study (Moher D, Liberati A, Tetzlaff J, 2009)

RESULT AND FINDINGS

Quality of Assessment

Table 4 shows the result of the assessment performance for 24 selected primary studies. The overall quality of the articles is as follows;

- High quality ($\geq 80\%$): Majority of studies
- Moderate quality (60–79%): Conceptual or descriptive studies
- Lower quality ($< 60\%$): Primarily survey-only or exploratory abstracts

Table 4- Quality of Assessment of 24 documents

PS	Author (Year)	QA1	QA2	QA3	QA4	QA5	QA6	Total Mark	Percentage (%)
PS1	Nascimento et al. (2025)	Y	Y	Y	P	N	P	4.0	66.7
PS2	Palacios-Núñez et al. (2025)	Y	Y	Y	Y	N	N	4.0	66.7
PS3	Eusebio-Hermira et al. (2026)	Y	Y	Y	Y	Y	P	5.5	91.7
PS4	Akhmedjanova et al. (2026)	Y	Y	Y	Y	N	P	4.5	75.0
PS5	Oladokun et al. (2026)	Y	Y	Y	P	N	N	3.5	58.3
PS6	González-Cervera et al. (2026)	Y	Y	Y	Y	N	N	4.0	66.7
PS7	Alawad (2025)	Y	Y	Y	Y	Y	N	5.0	83.3
PS8	Çela (2025)	Y	Y	Y	Y	Y	P	5.5	91.7
PS9	Karthika & B (2026)	Y	Y	Y	Y	N	N	4.0	66.7
PS10	Hamamra (2026)	Y	Y	Y	Y	N	Y	5.0	83.3
PS11	Ding, Zou & Kohnke (2025)	Y	Y	Y	Y	Y	P	5.5	91.7
PS12	Steiss et al. (2024)	Y	Y	Y	Y	Y	P	5.5	91.7
PS13	Mahdi & Alkhateeb (2025)	Y	Y	Y	Y	N	N	4.0	66.7
PS14	Bedington et al. (2024)	P	Y	P	P	N	P	3.0	50.0
PS15	Lin & Hwang (2026)	Y	Y	Y	Y	Y	N	5.0	83.3
PS16	Nguyen & Barrot (2024)	Y	Y	Y	Y	Y	P	5.5	91.7
PS17	Marzuki et al. (2023)	Y	Y	Y	Y	N	Y	5.0	83.3
PS18	De Wilde (2024)	Y	Y	Y	P	Y	P	5.0	83.3
PS19	Du & Nordin (2026)	Y	Y	Y	Y	Y	N	5.0	83.3

PS20	Agrati et al. (2026)	Y	Y	Y	Y	N	P	4.5	75.0
PS21	Yildiz & Asma (2026)	Y	Y	Y	Y	Y	N	5.0	83.3
PS22	Chen & Pan (2022)	Y	Y	Y	Y	Y	P	5.5	91.7
PS23	Li (2021)	Y	Y	Y	Y	N	P	4.5	75.0
PS24	Miranty & Widiati (2021)	Y	Y	Y	P	N	P	4.0	66.7

Table 4 presents the quality assessment results for the 24 selected studies (PS1–PS24) using six predefined quality assessment (QA) criteria adapted for this systematic literature review. Each criterion was evaluated using a three-point scale: Yes (Y = 1), Partly (P = 0.5), and No (N = 0), resulting in a maximum possible score of 6 points per study. Overall, the selected papers demonstrated satisfactory methodological quality, with total scores ranging from 3.0 (50.0%) to 5.5 (91.7%). Six studies achieved the highest score of 5.5 (91.7%), indicating that they clearly stated their research objectives, established sound methodologies, defined key concepts, and included comparisons with related work, although limitations were only partially discussed. Another six papers scored 5.0 (83.3%), while seven studies obtained 4.0 (66.7%), reflecting generally acceptable quality with minor shortcomings, particularly regarding comparisons with previous studies and explicit discussion of research limitations. Only one study received 3.5 (58.3%) and one study scored 3.0 (50.0%), suggesting relatively weaker reporting. Overall, the findings indicate that the majority of the selected studies are of moderate to high quality and are suitable for inclusion in the systematic literature review, while acknowledging that several studies could improve the reporting of comparative analyses and study limitations.

Analysis of all Three Themes

Table 5- Analysis for Themes

Theme	Number of Studies	Percentage
Theme 1: AI Integration in Academic Writing Teaching and Learning	12	50.0%
Theme 2: AI-Assisted Assessment, Feedback, and Writing Evaluation	7	29.2%
Theme 3: Teachers' Stance, Ethical Issues, and AI Adoption	5	20.8%
Total	24	100%

This three-theme framework aligns well with the research trends observed in the selected studies and provides a coherent structure for an SLR. Theme 1 focuses on how AI technologies are incorporated into academic writing instruction and their impact on learning outcomes.

Theme 2 examines AI's role in writing assessment, automated feedback, and evaluation, including comparisons between AI-generated and human-generated assessments. Theme 3 addresses the human and institutional dimensions of AI adoption, emphasizing educators' perceptions, ethical considerations, AI literacy, and implementation challenges.

Together, these themes comprehensively cover all 24 selected studies while avoiding redundancy, making them suitable as the main discussion sections in a PRISMA-compliant systematic literature review.

Theme 1. AI Integration in Academic Writing Teaching and Learning

Table 6- Theme 1

Theme Description	Papers
Studies investigating how AI tools are integrated into academic writing instruction, learning activities, writing development, self-assessment, and pedagogical practices.	PS2, PS4, PS8, PS9, PS10, PS11, PS14, PS15, PS17, PS19, PS23, PS24

Theme 1: AI Integration in Academic Writing Teaching and Learning comprises 12 studies (PS2, PS4, PS8, PS9, PS10, PS11, PS14, PS15, PS17, PS19, PS23, and PS24), making it the largest thematic group in this review. These studies primarily investigate how artificial intelligence (AI) technologies are integrated into academic writing instruction to enhance teaching practices, student learning, and writing performance.

The reviewed articles examine a wide range of AI applications, including ChatGPT, automated writing evaluation (AWE) systems, AI-supported feedback, procedural scaffolding, Grammarly, and other digital writing assistants. Collectively, the studies demonstrate that AI can improve various aspects of academic writing, such as grammar, vocabulary, organization, revision, self-assessment, critical thinking, and learner engagement.

Several studies also highlight the effectiveness of combining AI-generated feedback with instructor guidance, suggesting that a human–AI collaborative approach yields better learning outcomes than relying solely on AI. Moreover, the findings emphasize the importance of pedagogically sound integration, digital literacy, and appropriate instructional design to maximize the benefits of AI-assisted writing. Overall, this theme illustrates that AI has evolved from a simple writing support tool into an integral component of academic writing pedagogy, offering opportunities to personalize learning while fostering students' writing competence and autonomy.

Theme 2. AI-Assisted Assessment, Feedback, and Writing Evaluation

Table 7- Theme 2

Theme Description	Papers
Studies focusing on AI as an assessment, scoring, feedback, or evaluation tool, including comparisons between AI and human evaluators and the development of AI-based assessment frameworks.	PS3, PS7, PS12, PS13, PS16, PS21, PS22

Theme 2: AI-Assisted Assessment, Feedback, and Writing Evaluation comprises seven studies (PS3, PS7, PS12, PS13, PS16, PS21, and PS22) that examine the application of artificial intelligence in assessing academic writing, providing automated feedback, and evaluating students' performance.

The studies primarily compare AI-generated assessments with human evaluations to determine their reliability, consistency, and effectiveness. Several papers report that AI can deliver objective, timely, and consistent feedback, making it a valuable tool for formative assessment and reducing instructors' workload. Other studies focus on developing AI-assisted evaluation rubrics and assessment frameworks to improve scoring accuracy and support writing instruction.

Despite these advantages, the findings consistently emphasize that AI cannot fully replace human evaluators, particularly in assessing higher-order writing skills such as critical thinking, creativity, contextual understanding, and conceptual depth. Consequently, the reviewed studies advocate a hybrid assessment approach that combines AI efficiency with human expertise to achieve more comprehensive and pedagogically meaningful evaluation outcomes.

Theme 3. Teachers' Perspectives, Ethical Issues, and AI Adoption in Academic Writing

Table 8- Theme 3

Theme Description	Papers
Studies examining educators' and pre-service teachers' perceptions, ethical concerns, institutional readiness, AI literacy, responsible AI use, and challenges associated with AI adoption in academic writing.	PS1, PS5, PS6, PS18, PS20

Theme 3: Teachers' Perspectives, Ethical Issues, and AI Adoption includes five studies (PS1, PS5, PS6, PS18, and PS20) that explore educators' and pre-service teachers' attitudes toward the adoption of artificial intelligence in academic writing. The studies consistently highlight that teachers recognize AI's potential to enhance teaching, learning, and writing support while expressing concerns about academic integrity, originality, overreliance on AI, and the possible decline in students' critical thinking skills. Ethical issues, including responsible AI use, citation practices, and AI literacy, emerge as central topics across the studies. Furthermore, the findings emphasize that successful AI integration depends on adequate institutional support, clear policies, and professional development to strengthen teachers' digital competencies. Overall, these studies suggest that while educators generally hold positive views toward AI-assisted writing, its effective implementation requires balanced pedagogical practices, ethical guidelines, and continuous training to ensure that AI complements rather than replaces human judgment and academic integrity.

CONCLUSION

Summary of Findings

This systematic literature review synthesized evidence from 24 peer-reviewed studies published between 2021 and 2026 to examine the current state of artificial intelligence (AI) in academic writing within higher education. Guided by the PRISMA framework, the review identified three dominant themes: AI integration in academic writing teaching and learning, AI-assisted assessment, feedback, and writing evaluation, and teachers' perspectives, ethical issues, and AI adoption. The findings indicate that AI has evolved from a supplementary writing aid into a pedagogical technology that supports multiple stages of the writing process, including brainstorming, drafting, revision, feedback, and self-assessment. Most studies reported that AI-powered tools, such as ChatGPT, Grammarly, and Automated Writing Evaluation (AWE) systems, contribute positively to students' writing quality, learner engagement, and formative assessment by providing timely and personalized feedback. However, the review also found that AI cannot fully replace educators in assessing higher-order writing competencies, including critical thinking, originality, creativity, and contextual understanding. Furthermore, teachers generally acknowledged the educational value of AI while expressing concerns regarding academic integrity, ethical use, student dependency, transparency, and the need for institutional policies and professional development to support responsible implementation. These findings collectively demonstrate that successful AI integration depends on balancing technological innovation with sound pedagogical practices and human expertise.

Implications and Suggestions for Future Research

The findings of this review have several important implications for higher education institutions, educators, and researchers. From a pedagogical perspective, institutions should integrate AI into academic writing curricula using evidence-based instructional strategies grounded in the TPACK framework while leveraging AI-assisted feedback in accordance with Assessment for Learning principles. At the institutional level, clear policies, ethical guidelines, and continuous professional development are essential to foster responsible AI use and enhance educators' confidence in adopting AI technologies. The review also reinforces the importance of maintaining human oversight in writing assessment to ensure that AI complements rather than replaces instructor judgment. Future research should move beyond descriptive investigations by conducting longitudinal, experimental, and cross-cultural studies that evaluate the long-term impact of AI on students' writing performance, critical thinking,

academic integrity, and learning outcomes. In addition, future studies should explore the effectiveness of emerging generative AI tools across different disciplines, educational contexts, and student populations while examining ethical governance, assessment validity, and institutional readiness. Such research will contribute to developing sustainable and pedagogically sound models for integrating AI into academic writing in higher education.

REFERENCES

1. Agrati, L. S., Vinci, V., Beri, A., & Berardi, P. (2026). Use of AI-assisted writing tools for student self-assessment at school: Survey on teachers' perceptions and practices. *Frontiers in Education*, 11, Article 1701597. <https://doi.org/10.3389/educ.2026.1701597>
2. Akhmedjanova, D., Suchkova, S., & Zharkova, N. (2025). Teaching research writing with AI: A case study of academic development courses in higher education. *Journal of Language and Education*, 11(4), 112–130. <https://doi.org/10.17323/jle.2025.26589>
3. Alawad, E. A. (2025). Transformative assessment in academic writing: AI-based feedback systems with adaptive rubrics. *Journal of Teaching English for Specific and Academic Purposes*, 13(1), 141–149. <https://doi.org/10.22190/JTESAP250113012A>
4. Bedington, A., Halcomb, E. F., McKee, H. A., Sargent, T., & Smith, A. (2024). Writing with generative AI and human-machine teaming: Insights and recommendations from faculty and students. *Computers and Composition*, 71, 102833. <https://doi.org/10.1016/j.compcom.2024.102833>
5. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
6. Chen, H. M., & Pan, J. (2022). Computer or human: A comparative study of automated evaluation scoring and instructors' feedback on Chinese college students' English writing. *Asian-Pacific Journal of Second and Foreign Language Education*, 7(1), Article 34. <https://doi.org/10.1186/s40862-022-00171-4>
7. Çela, E. (2025). Exploring the potential and implications of digital tools in teaching academic writing. *Sodobna Pedagogika—Journal of Contemporary Educational Studies*, 76(4). <https://doi.org/10.63384/sptB54z804a>
8. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
9. De Wilde, V. (2024). Can novice teachers detect AI-generated texts in EFL writing? *ELT Journal*, 78(4), 414–422. <https://doi.org/10.1093/elt/ccae031>
10. Ding, L. Q., Zou, D., & Kohnke, L. (2025). ChatGPT as an automated writing evaluation tool: How students perceive it and how it affects their writing. *Education and Information Technologies*, 30(18), 26777–26799. <https://doi.org/10.1007/s10639-025-13775-3>
11. Du, J. P., & Nordin, N. R. M. (2026). The impact of the iWrite automated writing evaluation system on university EFL students' writing performance and writing anxiety. *Education Sciences*, 16(3), Article 411. <https://doi.org/10.3390/educsci16030411>
12. Eusebio-Hermira, S., Cuberos-Vicente, R., & Doquin de Saint-Preux, A. (2026). Academic assessment with artificial intelligence: Comparison between teachers and ChatGPT in master's theses. *Revista de Investigación en Educación*, 24(2), 309–326. <https://doi.org/10.35869/reined.v24i2.6942>
13. González-Cervera, A., Burgueño-López, J., & Urosa-Sanz, B. (2026). Ethical use of artificial intelligence tools by future teachers. *Canadian Journal of Learning and Technology*, 52(1). <https://doi.org/10.21432/cjlt29045>
14. Hamamra, B. (2026). Assessed academic writing outcomes in an AI-permitted research methodology course: Evidence from EFL university students in Palestine. *Asian-Pacific Journal of Second and Foreign Language Education*, 11(1), Article 36. <https://doi.org/10.1186/s40862-026-00413-9>
15. Karthika, V. K., & B, H. M. (2026). Exploring the impact of AI feedback in academic writing self-assessment: A duoethnographic study. *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2026.2619502>
16. Kitchenham, B. (2007). Guidelines for performing systematic literature reviews in software engineering (Version 2.3). EBSE Technical Report.

17. Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, 99, 102505. <https://doi.org/10.1016/j.system.2021.102505>
18. Lin, C. J., & Hwang, G. J. (2026). Artificial intelligence-supported procedural scaffolding for promoting EFL learners' writing performance in flipped peer assessment activities. *Interactive Learning Environments*, 34(3), 1839–1853. <https://doi.org/10.1080/10494820.2025.2532629>
19. Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis: Methodological guidance for systematic reviewers utilizing meta-aggregation. *International Journal of Evidence-Based Healthcare*, 13(3), 179–187. <https://doi.org/10.1097/XEB.0000000000000062>
20. Mahdi, H. S., & Alkhateeb, A. (2025). Revolutionising essay evaluation: A cutting-edge rubric for AI-assisted writing. *International Journal of Computer-Assisted Language Learning and Teaching*, 15(1), Article 368226. <https://doi.org/10.4018/IJCALLT.368226>
21. 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), Article 2236469. <https://doi.org/10.1080/2331186X.2023.2236469>
22. Miranty, D., & Widiati, U. (2021). Automated writing evaluation (AWE) in higher education: Indonesian EFL students' perceptions about Grammarly use across student cohorts. *Pegem Eğitim ve Öğretim Dergisi*, 11(4), 126–137. <https://doi.org/10.47750/pegegog.11.04.12>
23. Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
24. Nascimento, K. A. S., Fialho, L. M. F., & da Costa, M. A. A. (2025). Pedagogy professors' perspectives on the use of artificial intelligence in academic work. *Educação e Pesquisa*, 51, e294604. <https://doi.org/10.1590/S1678-4634202551294604en>
25. Nguyen, L., & Barrot, J. S. (2024). Detecting and assessing AI-generated and human-produced texts: The case of second language writing teachers. *Assessing Writing*, 62, 100899. <https://doi.org/10.1016/j.asw.2024.100899>
26. Oladokun, B. D., Ukubeyinje, E. S., Oyighan, D., Obim, I. E., & Oloniruha, E. A. (2026). What if AI wrote it: Exploring the perspectives of university teachers toward AI-generated literature by undergraduates in library schools. *College & Undergraduate Libraries*, 33(1), 42–64. <https://doi.org/10.1080/10691316.2025.2607372>
27. Page, M. J., McKenzie, J. E., Bossuyt, P., 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. *Medicina Fluminensis*, 57(4), 444–465. https://doi.org/10.21860/medflum2021_264903
28. Palacios-Núñez, M. L., Mendoza-García, E. M., Zarate, J. W. N., & Deroncele-Acosta, A. (2025). ChatGPT in the teaching of academic writing in higher education: Teachers' perspectives on its uses, challenges, and future in personalized learning. *Educat*, 93, 33–51. <https://doi.org/10.21556/edutec.2025.93.3995>
29. Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J. L., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
30. Yildiz, Ü., & Asma, B. (2026). Comparing artificial intelligence (AI) and instructor scoring in writing skills assessment. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 66. <https://doi.org/10.9779/pauefd.1574813>