RIPAFD
REVISTA INTERNACIONAL DE PEDAGOGÍA DE LA ACTIVIDAD FÍSICA Y DEPORTES

Vol. 2 Núm. 1 (2026)
Innovación educativa y Actividad Física para un futuro inclusivo

Recibido: Aceptado: Publicado:

Automated systems that integrate generative AI tools and writing tutorials for scientific papers: a systematic review

Darley Jhosue Burgos Angulo [1]
David Job Morales Neira [2]
[1] University of Guayaquil, Faculty of Sciences of Physical Activity, Guayaquil, Ecuador.
[1] State University of Milagro, Faculty of Educational Sciences, Milagro, Ecuador; AITEC - Admiral Illingworth Technological Institute, Higher Technology in Physical Activity, Sports and Recreation, Guayaquil, Ecuador.

[*] Corresponding author: Darley Jhosue Burgos Angulo
darley.burgosan@ug.edu.ec

Abstract:
Scientific writing demands precision, coherence, and adherence to certain academic standards. This systematic review aimed to analyze automated systems that combine generative artificial intelligence and tutoring to improve the production of scientific papers. A search was conducted in Scopus (n=287), Web of Science (n=254), and ERIC (n=188). After applying inclusion and exclusion criteria, 14 studies were selected for further analysis. The results showed that tools such as Grammarly, ChatGPT, Writefull, Scholarcy, and Scispace Copilot help with draft generation, grammar checking, section structuring, and automatic summarization. Combining these tools with personalized tutoring enhances the clarity, coherence, and efficiency of texts, although expert supervision is necessary to ensure scientific rigor and ethical authorship. These systems represent a promising strategy for strengthening academic writing and guiding future research on their responsible use.

Keywords: Scientific writing, artificial intelligence, computer-asisted instruction, educational technology, higher education.

Resumen:
La redacción científica exige precisión, coherencia y el cumplimiento de ciertos estándares académicos. Esta revisión sistemática se propuso analizar los sistemas automatizados que combinan inteligencia artificial generativa y tutorías para mejorar la producción de trabajos científicos. Se llevó a cabo una búsqueda en Scopus (n=287), Web of Science (n=254) y ERIC (n=188). Después de aplicar criterios de inclusión y exclusión, se seleccionaron 14 estudios para un análisis más detallado. Los resultados mostraron que herramientas como Grammarly, ChatGPT, Writefull, Scholarcy y Scispace Copilot ayudan en la generación de borradores, revisión gramatical, estructuración de secciones y resúmenes automáticos. La combinación de estas herramientas con tutorías personalizadas mejora la claridad, coherencia y eficiencia de los textos, aunque es necesario contar con supervisión experta para asegurar el rigor científico y la ética en la autoría. Estos sistemas representan una estrategia prometedora para fortalecer la redacción

Recibido: Aceptado: Publicado:

académica y guiar futuras investigaciones sobre su uso responsable.

Palabras Clave: Redacción científica, inteligencia artificial, enseñanza asistidad por ordenador, tecnologías educativas, Educación superior.

1. Introduction

The integration of generative artificial intelligence in academia is having a significant impact on how students, professors, and researchers generate, analyze, and communicate knowledge. Innovations in natural language processing (NLP) AI are no longer merely assistance tools; they are now mediators in thought, argumentation, and text production. Thus, producing academic text is no longer an exclusively human activity, but rather a practice involving multiple social and technical elements, including sophisticated AI that can generate, transform, and evaluate documents (Bozkurt, 2024; Cohen & Moher, 2025).

This paradigm shift has impacted how scholarly writing is taught and learned, and has fostered new collaborative ecologies of human-machine knowledge. In this context, authors are no longer solitary creators; they are critical negotiators with algorithms, modifying outputs and directing them toward defined communicative and academic goals (Singh et al., 2025). In this sense, writing is a hybrid activity, as human creativity coexists with the predictive power of generative systems, raising questions of originality, authorship, and the ethics of digital co-authorship (Pividori & Greene, 2024; Tang et al., 2024).

AI for educational purposes has grown rapidly in higher education, particularly in the areas of writing, academic tutoring, and research training. Automated text generation tools such as ChatGPT, Quillbot, and Grammarly are used not only for style editing but also to assist in planning, structuring, and refining scholarly discourse. More recent research has demonstrated that these tools can improve clarity, coherence, and linguistic accuracy, as well as increase students' confidence in tackling challenging writing tasks (Rashid et al., 2024; Tseng & Lin, 2024; Zhou & Du, 2025). However, these instrumental benefits are accompanied by pedagogical challenges, as excessive automation can limit critical thinking, weaken cognitive self-regulation, and induce technological dependence (Nguyen et al., 2024; Zizka, 2025).

In higher education, the use of generative systems has proven particularly useful in contexts where English is a foreign or second language. It has been shown that when AI supports teaching, significant improvements occur in students' grammatical accuracy, text organization, and vocabulary development. Similarly, the study by Thohir et al. (2024) notes that the use of AI, particularly Quillbot, aids in paraphrasing, summarization, and reducing plagiarism, thereby supporting the semantic comprehension of academic texts. In all cases, appropriate mediation, ethical guidance, and clarity of learning objectives have been key factors in the positive impact observed.

From the perspective of curriculum design, generative systems allow for the adaptation of teaching, provide real-time feedback, and create more inclusive learning spaces. However, their use requires a mastery of digital competencies that both students and teachers must acquire, particularly knowing how to read the outputs of the models, detect and manage biases, and verify the accuracy of the generated information. This is complemented by academic education and broader literacy in the system for the development of critical thinking, digital ethics, and discernment. (Aljuaid, 2024; Pérez et al., 2025).

Recibido: Aceptado: Publicado:

The impact of AI on writing and data analysis has also reached scientific research. According to Yang & Hwang (2025), generative tools make it possible to automate stages of the research process, from drafting the theoretical framework to writing the conclusion, which optimizes the work and enhances the coherence of the text. However, these developments require even more vigilance to prevent errors in methodology and biases that may arise from the training of the models. At the same time, works like those of Melliti (2024) have shown that students' interaction with ChatGPT in writing graduate theses affects their lexicon and discourse patterns, modifying their style and structure of academic thinking.

Psychological and motivational factors are also important in this regard. Oubibi (2025) demonstrates that the use of AI in graduate work increases digital adaptability, academic confidence, and student engagement. Such findings suggest that when AI is used with flexible formative supervision, it could act as a catalyst for self-directed learning. On the other hand, the experimental research conducted by Ironsi & Solomon (2025) has demonstrated statistically significant improvements in the coherence and quality of texts produced with generative AI, thus confirming the effectiveness of these tools in developing stronger writing skills.

From an ethical perspective, the academic community faces an urgent debate. According to Bozkurt (2024), there is a need to review the definitions of authorship and intellectual property, as well as to establish clearer standards regarding the recognition of the use of such technologies in academic communication. In the same vein, Cohen & Moher (2025) reflect on the ambivalence of AI as a friend or foe of academic writing and argue that its usefulness largely depends on how it is integrated pedagogically and ethically. Pividori & Greene (2024) propose editorial infrastructures that allow for the collaboration between humans and generative systems to be recorded and transparent, while maintaining the traceability and integrity of the manuscripts.

The convergence of these approaches illustrates that artificial intelligence is perceived as a language tool and a cognitive agent that transforms the production of knowledge. Recent studies confirm that the incorporation of artificial intelligence in teaching and writing processes increases autonomy, accuracy, and creativity, as long as it is accompanied by human guidance, ethical evaluation, and appropriate instructional design (Akpan et al., 2025; Pack & Maloney, 2023). In distance education settings, Maphoto et al. (2024) have observed increases in writing self-efficacy and student retention with the use of generative chatbots as learning mediators. Weidmann (2024) highlights the potential of generative chatbots in more specialized areas, such as pharmaceutical education, when used with teacher supervision.

However, despite the growing body of empirical and theoretical research, the literature continues to exhibit conceptual fragmentation and methodological dispersion. Existing studies tend to focus on isolated aspects of AI-assisted writing, such as linguistic accuracy, user perceptions, and ethical concerns, while overlooking an integrative perspective that encompasses the pedagogical, cognitive, technological, and academic integrity dimensions of AI-supported writing. This lack of conceptual integration limits a comprehensive understanding of the impact that automated systems may have on strengthening research practices, higher education, and scientific knowledge production.

To address this knowledge gap, the present study aims to examine the scholarly literature

Recibido: Aceptado: Publicado:

on the use of generative AI-powered automated systems in academic writing and university tutoring. Specifically, it seeks to analyze their pedagogical contributions, ethical implications, and the challenges associated with their integration into contemporary higher education.

2. Metodology

Method

A systematic review was conducted following a rigorously defined, comprehensive, and transparent approach aligned with the PRISMA 2020 guidelines, which represent the international standard for reporting systematic reviews across diverse fields of research (Page et al., 2021). Specifically, this study examines empirical research on the use of generative artificial intelligence tools and automated tutoring systems in the teaching of academic and scientific writing, with particular attention to their effects on writing performance, self-regulation, and university students' perceptions. The methodological approach enabled the identification, synthesis, and comparison of findings from the contemporary literature in order to determine patterns of effectiveness, methodological limitations, and emerging trends within the field of AI-assisted writing.

Data sources and search strategies

Between January and September 2025, a bibliographic search was conducted in databases such as Scopus, Web of Science, ERIC, and Google Scholar, which were chosen for their extensive coverage as well as their international recognition in the dissemination of research in education, technology, and pedagogical innovation. For the search, combined strategies were employed using Boolean operators, and terms in both Spanish and English were used to ensure the complete retrieval of evidence. For the search in English, the expression used was: ("artificial intelligence" OR "generative AI") AND ("education" OR "tutoring system") AND ("academic writing" OR "research writing"), and in Spanish the combination: (AI OR "artificial intelligence") AND ("automated tutoring" OR "assisted learning") AND ("scientific writing" OR "academic writing"). For both, the fields of title, abstract, and keywords were considered.

("inteligencia artificial" OR "IA generativa") AND ("educación" OR "sistema de tutoría") AND ("escritura académica" OR "redacción de investigaciones")

(AI OR artificial intelligence) AND (automated tutoring OR assisted learning) AND (scientific writing OR academic writing)

Inclusion and Exclusion Criteria

The inclusion criteria were set as follows: peer-reviewed publications with full-text availability published between 2020 and 2025, as this period was considered the most representative of the rise of generative artificial intelligence in educational contexts. The criteria established for the review are detailed below.

Recibido: Aceptado: Publicado:

Table 1. *Inclusion and Exclusion Criteria*

PICO element	Inclusion Criteria	Exclusion Criteria
P (Population)	Undergraduate and graduate students or faculty members involved in academic or scientific writing..	Participants in elementary or secondary education, professionals outside the academic or corporate sectors.
I (Intervention)	Integration of generative artificial intelligence tools (ChatGPT, Copilot, Gemini, Quillbot, Claude, etc.) and automated tutorials into the teaching and learning process of academic writing	Studies focusing on non-generative AI or applications of artificial intelligence in writing contexts. Health, marketing, engineering, etc.
C (Control)	Comparison with traditional teaching methods, human tutoring, or no pedagogical intervention.	Lack of a control group, absence of a controlled methodology, or inadequate description of the comparative conditions.
O (Outcomes)	Writing performance, quality of the text produced, perceptions of usefulness, self-regulation, and digital literacy.	Research that does not include quantitative or qualitative findings on outcomes, lacks impact indicators, and is purely descriptive.
Study design	Empirical studies—whether quantitative, qualitative, or mixed-methods—published in indexed peer-reviewed journals.	Reviews or other types of documentation within non-peer-reviewed frameworks.

Note: Authors' own elaboration.

Study selection procedure

The systematic search yielded a total of 732 records, distributed as follows: 287 in Scopus, 254 in Web of Science, 188 in ERIC, and 3 in Google Scholar. After an initial automatic screening, 12 duplicates and 447 ineligible records were removed. The latter were discarded due to lack of relevance to the topic or failure to meet the inclusion criteria. Of the 276 articles remaining after the first screening, a review of the titles and abstracts led to the exclusion of 150 due to low relevance. We then proceeded to retrieve and analyze the full text of 126 articles; 77 of these were excluded due to a lack of empirical evidence, a lack of an intervention design, or the document being unavailable. The methodological eligibility of 49 articles that had been qualitatively prioritized was then evaluated. In this stage, 35 were excluded: 14 for not addressing generative AI or the automation of tutoring, 11 for being outside educational contexts, and 10 for lacking empirical evidence or an educational intervention. Thus, 14 met the inclusion criteria for the final synthesis of the review.

The study selection process was carried out by two independent reviewers, who were responsible for identifying, filtering, and evaluating the articles according to the criteria established in the PICO methodology. This ensured a consistent and rigorous procedure. Each reviewer simultaneously analyzed the titles, abstracts, and full texts, ensuring methodological coherence and the relevance of the selected studies. In the event of discrepancies regarding the inclusion or exclusion of certain records, a third expert reviewer was consulted to mediate and resolve the differences through a reasoned consensus.

Evaluation of methodological quality

The methodology of the selected studies was evaluated using the Critical Appraisal Checklist for Analytical Cross-Sectional Studies from the Joanna Briggs Institute (Aromataris & Munn, 2020), which allows for the assessment of the clarity of the sample description, the validity of the instruments, the rigor of the data

Recibido: Aceptado: Publicado:

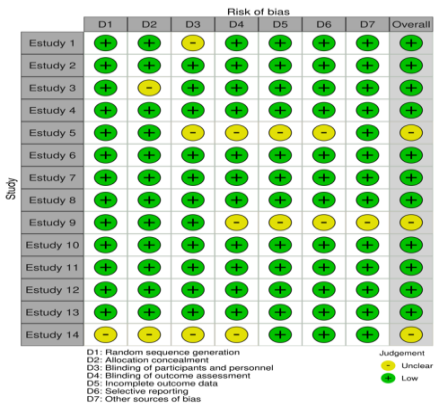
collection procedures, as well as the accuracy of the data analysis. Each study was individually evaluated by two expert reviewers, who resolved their discrepancies by consensus. The results of the bias risk assessment were presented graphically using the "traffic light" model from the ROBVIS software. This model classifies the risk into three levels: low, moderate, and high, according to the domains established by the JBI (McGuinness & Higgins, 2021). According to the analysis, the majority of the studies, around 77%, showed a low risk of bias, while the remaining 23% presented a moderate risk, with no cases of high risk identified. The most common limitations were found in domains related to random sequence generation, allocation concealment, and blinding of participants or assessors, aspects that can affect the internal validity of the studies.

Data extraction and analysis

Data were extracted and analyzed using a matrix that clarified and organized the relevant details of each study. Descriptive variables—such as author, year, country, type of AI, population, methodological design, and instruments—were organized alongside analytical variables pertaining to results and conclusions. The analysis of the studies employed both quantitative and qualitative methodologies, emphasizing thematic coding and narrative synthesis. Quantitative study indicators, such as means, standard deviations, ANOVA, Pearson correlations, and t-tests, were used to compare group means, with a significance threshold of $p < 0.05$. Qualitative value was calculated using content analysis, and emerging categories—such as pedagogical effectiveness, student perceptions, development of self-regulation skills, and ethical use of generative AI—were integrated as qualitative variables. This ensured the internal and external validity of the analysis and demonstrated the benefits of the PRISMA 2020 international standards regarding replicability.

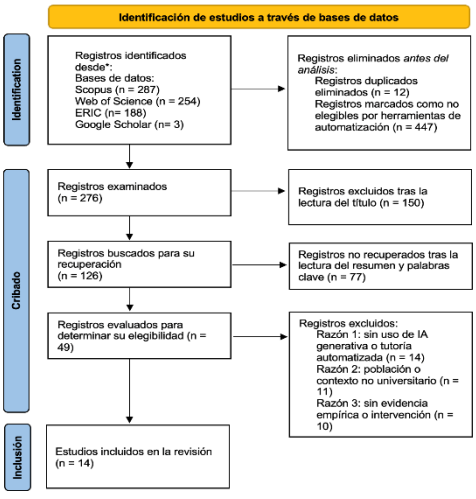
3. Results y discussion

Figure 1. *Results of the risk of bias in the studies*



Source: McGuinness & Higgins (2020).

Figure 2. *PRISMA Flowchart Results*



Source: Page et al. (2021).

Table 2. *Extraction of main data from the selected studies*

Author and year	Database	Variables	Objective	Methodology	Main results	Conclusions
Anani et al. (2025)	Scopus	Attitudes towards AI and its use in academic writing.	To explore the attitudes and usage patterns of AI tools among postgraduate students in Ghana.	A mixed-methods study was conducted with 339 students, using 5-point Likert scale questionnaires and semi-structured interviews. A descriptive analysis was performed, including frequencies, means, and standard deviations, as well as thematic coding. The perceived usefulness of AI was quite high (M=4.2, SD=0.8).	78% of respondents used AI to generate ideas, 85% to correct spelling, and 69% to paraphrase. The high standard deviation indicates a diversity of attitudes. Interviews revealed that the influence of teachers and peers impacts AI use, and concerns also arose regarding creativity and critical thinking. AI improves efficiency and quality in academic writing.	It is suggested to implement ethical training and teacher supervision. It is essential to promote clear institutional policies. There is a self-report bias. It is recommended to conduct future longitudinal studies to evaluate the impact on critical thinking and creativity.
Le et al. (2025)	Scopus	Feedback preferences, interaction with AI, and human tutors.	To analyze students' preferences between human feedback and AI-generated feedback.	In an experiment involving 114 students, four conditions were established: no feedback, human tutor, ChatGPT without restrictions, and an analytical tool. A pre-post ANOVA analysis was conducted, along with qualitative comments.	Initially, 72% of students preferred human tutors; after the intervention, free ChatGPT reached 45%, while the structured interface maintained a 68% preference for humans. ANOVA showed significant differences between the conditions ($F(3,110)=6.21$, $p<0.01$). The feedback suggests that natural interaction with AI reduces aversion to algorithms.	The effectiveness of AI depends on the interface design and user experience. It is recommended to create hybrid environments that balance AI and human teaching. A limitation of the study is its short duration; for future research, a longitudinal follow-up and replication in different educational contexts are suggested.
Teng (2025)	Scopus	Metacognitive awareness and the use of ChatGPT	To analyze how English as a Foreign Language (EFL) students use ChatGPT and their level of metacognitive awareness.	A mixed-methods study was conducted with 452 EFL students, using a validated questionnaire and semi-structured interviews, in addition to confirmatory factor analysis and thematic coding.	The results showed that 38% of students copied generated text, while 42% used it to receive feedback and improve. A positive correlation was also found between the strategic use of the tool and academic performance ($r=0.46$, $p<0.01$). Those with high metacognitive awareness tend to integrate artificial intelligence into their learning process, planning revisions and adjusting their style.	AI has the potential to enrich learning if used strategically. Implementing training programs that foster critical thinking and responsibility is recommended. However, it's important to note that the study's limitations include self-reporting and the EFL context; for future research, controlled pedagogical interventions and long-term follow-up are suggested.

Tseng et al. (2025)	Scopus	Literacy, Artificial Intelligence (AI), and Scientific Writing	To assess the impact of AI on nursing reports is a critical issue.	This quasi-experimental study involved 203 students, comparing a group that used AI (the experimental group) with a control group. Tools such as MAILS and summative assessments were used, and a pre- and post-test ANOVA analysis was conducted.	The results showed significant improvements in understanding of AI ($M = 4.6$ vs. 3.2, $p < 0.01$) and in its application in reports ($M = 4.5$ vs. 3.7, $p < 0.01$). In addition, ChatGPT was used 100% of the time for planning and development, resulting in superior overall performance in the experimental group.	The integration of AI not only improves digital literacy but also academic performance. This implies the need for teacher supervision and ongoing training. However, it should be noted that the generalizability of these results is limited, and it is suggested that future research include multicenter studies and long-term follow-up.
Tarchi et al. (2025)	Scopus	SBW Strategies and ChatGPT Integration.	To analyze how students use ChatGPT for source-based writing and its relationship with text quality.	Twenty-seven students participated; a questionnaire was used to assess their motivation and prior knowledge; SBW was administered using ChatGPT; a retrospective think-aloud session was conducted; and a qualitative and correlational analysis was performed.	Limited use of ChatGPT was observed, primarily due to a lack of experience and ethical concerns; furthermore, a negative correlation was found with the verbatim inclusion of information ($r = -0.42$, $p < 0.05$). Some students used AI to improve the organization and coherence of their texts.	Although ChatGPT is useful for summarizing and organizing information, using it without proper training can lead to dependency and hinder deep learning. Ethical and critical instruction is recommended. A limitation of this study is the small sample size; for future research, it is suggested that the study be replicated with a larger sample size.
Barrot (2025)	Scopus	The use of Google Gemini in academic writing is an interesting topic.	To examine how this tool can be a valuable resource for academic writing.	In a preliminary study involving 35 students and teachers, interviews were conducted and academic papers were analyzed.	Students appreciate the automated assistance it provides, while teachers tend to be somewhat skeptical about its accuracy; however, the overall perception is positive. Google Gemini has the potential to enhance both teaching and academic writing.	This implies that it is necessary to train teachers and conduct a critical evaluation of the content produced. One limitation of the study is that the sample size is small and the context is unique, suggesting that further research is needed, such as multicenter experimental studies.
Moorhouse et al. (2025)	Scopus	L2 Writing and Empowerment Agency.	To investigate how L2 writers use generative AI and institutional policies.	Qualitative study: 21 second-language writers at the graduate level; interviews and preference tasks; thematic analysis.	Writers are adopting generative AI and feel empowered; however, they exhibit limited critical awareness and face tensions regarding institutional policies. Generative AI can enhance perceived agency, though critical awareness and transparency in its use must be addressed.	The authors suggest adopting a flexible policy, accompanied by ethics training. Limitations: small sample size and time frame; future research: longitudinal studies on agency and performance.
Yang et al. (2025)	Scopus	CoRL-VT, an English writing competition.	To evaluate the effect of a CoRL-guided AI	A quasi-experimental study was conducted with 61 students, comparing the CoRL-VT intervention	The results showed a significant improvement in macro-organization and lexical richness;	CoRL-VT promotes self-regulation and formative writing; a structured and guided approach is

			virtual tutor on writing performance.	with a standard AI, using the IELTS writing task before and after the intervention, with an ICC of 0.87.	the post-test, adjusted for baseline scores, favored CoRL-VT, a finding supported by teachers' feedback.	recommended. However, it should be noted that the sample size is small, and for future research, it is suggested that replicable studies be conducted in different contexts.
Ossa & Willatt (2023)	WOS	AI-powered formative feedback, academic writing	To apply generative AI to improve academic writing	Qualitative study. 347 students, 578 responses, ChatGPT-4 feedback using the RECaP-GPT process. Criterion-based feedback.	Learning evidence has improved significantly; feedback has become fast, scalable, and accessible; longitudinal follow-up presents challenges.	Generative AI improves the feedback provided and the development of expected competencies. The implication is that the lack of longitudinal monitoring presents a challenge.
Jin et al. (2025)	WOS	Strategies SRL and the Use of AI in Writing	To model the influence of self-regulated learning on AI writing outcomes.	Qualitative study. 1073 postgraduate survey; 3-level SEM framework for AI use: instrumental, text development, and transformation.	Advanced use of AI improves critical thinking, motivation, and writing quality; basic use does not; AI writing self-efficacy predicts self-regulation.	Self-regulation and self-efficacy in the use of AI should be encouraged, as they optimize results. This implies a need for training in advanced techniques. The limitation of this approach is that it is cross-cutting.
Kong et al. (2024)	WOS	Pedagogy 6-P and SRL with AI.	To design pedagogical strategies for academic writing with AI assistance.	Qualitative study. Iterative 6-P instructional design: planning, soliciting, anticipating, producing, peer review, portfolio monitoring; qualitative observation and analysis.	Improved critical thinking and self-regulation; effective proofreading and editing strategies; collaborative feedback optimizes the learning process.	The 6-P framework facilitates the ethical and strategic use of AI; recommendation: guidance on educational intervention. Limitations: initial implementation. Future research: multicenter validation.
Wang & Ren (2024)	WOS	Linguistic writing and the use of generative AI.	To analyze the use of AI in a collaborative multimedia project.	Descriptive study. 140 students; ChatGPT 3.5 and AI image generation on Padlet; 13-week Wikibook project; survey (N=98)	AI was used in the project to organize and summarize content more quickly and coherently. Interactive AI exercises enhance understanding. AI increases the effectiveness of collaborative learning and digital writing.	Guided integration and training on the limitations and strengths of AI are recommended. Limitation: single-center study. Future research: multicenter, replicable studies.
Vanichvasin (2021)	ERIC	Use of an educational chatbot, perceived effectiveness of the chatbot, students' research knowledge.	To create an educational chatbot, evaluate its effectiveness as a digital learning tool, and analyze its impact on university students' research literacy in Thailand.	A quasi-experimental study was conducted with 36 Thai university students. A chatbot designed to support research learning was developed. The data collection instruments included: the chatbot itself, an expert evaluation form, a questionnaire on effectiveness, and research knowledge tests. Descriptive statistical analysis	The expert evaluation rated the chatbot's usability as very high (M = 4.67, SD = 0.08), suggesting improvements such as more research content and interactive elements. Students' perception of the chatbot's effectiveness was high (= 4.43, S.D. = 0.35) during the pilot test and (= 4.37, S.D. = 0.48) in the target group, highlighting	The students appreciated the ability to access information immediately and perform specific searches, although limitations were identified in the system's ability to interpret incorrect keywords and in the need for selection options for predefined questions. Post-test scores were significantly higher than pre-test

				(mean, standard deviation), content analysis, and a t-test were applied to compare the results before and after the intervention. A pilot test was conducted with 14 students outside the target group to adjust content and functionality.	ease of use, clarity of understanding, innovation, and playful appeal.	scores ($p < 0.05$), indicating a genuine increase in research skills.
(Gong, 2022)	ERIC	IRTP Platform, Digital Teaching Research	To analyze the use of IRTP for educational research and teacher professional development.	Case Study: Secondary School Teaching Staff; IRTP for Classroom Analysis, Observation, and Competency Matrix; Big Data and Scales.	Improves the analysis of teaching and the development of teaching competencies; facilitates applied research and progress monitoring. IRTP promotes professional development for teachers and research practice.	It is suggested to integrate AI to strengthen teacher training. Limitation: specific context; future research: implementation at different educational levels.

Note: Authors' own elaboration.

This review suggests that automated systems incorporating generative AI and tutoring modules positively, albeit somewhat variably, affect writing processes and teacher training. The included experimental and quasi-experimental studies (Jin et al., 2025; Tseng et al., 2025; Yang & Hwang, 2025) document gains in instrumental competencies (text improvement, organization, and lexical richness) and self-regulation skills within guided frameworks with a pedagogical focus on AI. Notably, interventions that combine human scaffolding and text generation (CoRL) tend to produce better results than conversational tools in isolation (Kong et al., 2024; Y. Yang et al., 2025). The educational structure and mediation largely explain why AI leads to significant improvements in writing and learning transfer.

One trend evident in the results is the reliance on interaction design. Recent literature by Le et al. (2025) on conversational interfaces shows that configuring a system like ChatGPT in open-dialogue mode can reduce algorithmic aversion and facilitate student acceptance, but this will not always guarantee improved higher-order outcomes if there is no pedagogical intervention to encourage reflection, verification, and critical revision of the text. This may explain the superior results in interventions that combine self-regulation strategies (Jin et al., 2025; Kong et al., 2024) and sequenced pedagogies (6-P). In these cases, AI is used as a catalyst for practice, but it is metacognitive mediation that transforms the support into the development of writing competence.

Contrasting the above with the literature that explores large-scale AI applications without a solid pedagogical design. The implementation of automated mass feedback systems Ossa & Willatt (2023) clearly demonstrate operational viability, efficiency (in terms of scale and speed), and improvements in access and coverage of formative feedback. However, the reports also note limitations including incomplete responses to queries that do not match keywords, a tendency to homogenize feedback, and the risk of overconfidence in students. These limitations align with the findings in Vanichvasin (2021) and the pilot evaluation of chatbots where usability and satisfaction scores were high (average > 4/5) but there were explicit demands to strengthen the content, provide alternative interactive pathways, and improve tolerance to incorrect inputs. Overall, the absence of a well-prepared design does not justify the technological pedagogical limitations. Increasing efficiency can improve reach, but without instructional redesign, it leaves considerable gaps in the development of critical thinking skills and personalized learning.

Regarding AI literacy, Tseng et al. (2025) and Yang et al. (2025) highlight studies that incorporate explicit training in the use of the tools and affirm that improvements in digital skills are associated with a more advanced and strategic use of AI, specifically the use of text transformation versus the use of technical assistance. Jin et al. (2025) and Teng (2025) point out that self-efficacy and self-regulation are strong predictors of uses that generate improvements in critical thinking and writing quality; these mediating relationships are entirely consistent with the qualitative literature by Anani et al. (2025) and Moorhouse et al. (2025) that describes the perception of empowerment and the risk of dependency. The practical implication, therefore, is that incorporating tools is not enough; Users must be trained to critically analyze outputs, validate sources, and ensure that AI proposals align with disciplinary intersections and academic ethics.

Studies on feedback sequences (Tran, 2025) and the integration of human and AI feedback (Anthony et al., 2025; Sparkman & Witt, 2025) indicate that the order and type of feedback determine the effectiveness of subsequent revisions. The sequences “teacher → AI” and “AI → teacher” do not produce comparable results, reinforcing the notion that AI should be incorporated as a component within

intentional learning cycles. Second, case studies and implementation studies (Abraham & Accad, 2025; Barrot, 2025; Pizarro-Romero & Lovón, 2025) illustrate that different tools (Claude, Gemini, Perplexity) offer varying performance and risk profiles: while some support a rapid creative flow, others provide useful tracers or metadata for verification, making it necessary to choose the appropriate tool for each learning objective.

The methodological and critical literature Odeh et al. (2025); Peters (2025); Saqib & Zia (2025) emphasize the importance of academic integrity and the need for automated models for AI generation detection and paraphrasing; these are not only technical, but also ethical and regulatory, and affect the possibility of large-scale institutional adoption.

The variety in the types of works included, such as randomized trials, quasi-experimental studies, case studies, and bibliometric reviews, in this case from a methodological perspective, strengthens the triangulation of evidence although it limits comparability in quantitative analysis. Many of the reviewed studies indicate moderate to large effects, for example, in the advanced use of certain tools and in performance indicators (an r of 0.4 to 0.8 is considered a strong correlation), and in structured interventions, significant pre and post improvements were found. However, suitable meta-analyses that allow for an estimation of the effect in question are still lacking. On the other hand, most bibliographies and reviews frame their research within university contexts of languages or related disciplines and use small samples (between 27 and 1,073 in the study being analyzed), which affects the possibility of generalization to other disciplines and levels.

This review makes three key contributions. The first is conceptual: it suggests viewing automatic generative AI systems not as substitutes for human tutors, but as co-agents whose effectiveness depends on the pedagogical architecture (scaffolding, feedback loops, verification practice). The second is empirical: it synthesizes evidence showing that interventions combining AI and pedagogy (CoRL, 6-P, AI literacy training) lead to sustained improvements in writing quality and metacognitive practices, with practically significant effect sizes in some studies. The third is normative: it outlines specific expectations for responsible implementation (teacher training, transparency policies, detection of AI-generated content, and ethical evaluation).

Regarding the limitations and future lines of work, these should be multicentric, longitudinal, and experimental. They should compare tools (ChatGPT, Gemini, Claude) and protocols for pedagogical integration. They should evaluate not only the written products but also the transfers in critical thinking, creativity, and investigative behaviors. It is urgent to create standard metrics that assess the "verifiable quality" of the texts produced and the critical appropriation that the user makes. Also, research must address the bias of inequity. That is, to evaluate the impact of these technologies on students with different levels of digital literacy and in low-resource contexts. Finally, the integration of academic governance (transparent use of AI, authorship, plagiarism control) and pedagogical practices must be balanced so that adoption can scale without sacrificing the integrity and critical development of the student.

This study reaffirms that the adoption of automated systems with generative AI combined with the implementation of instructional designs, AI literacy training, and the ethical training of institutions offers transformative potential in academic writing and teacher training. Future studies should focus on the field of pedagogy, which still requires optimization work, and the evaluation of medium- and long-term outcomes.

3. Conclusions

The systematic review made it possible to identify, analyze, and characterize automated systems that integrate generative AI tools and writing guidance for scientific papers. Some of the most widely used and established systems in practice are Grammarly, ChatGPT, Writefull, Scholarcy, and Scispace Copilot, which offer a wide range of features, such as automatic draft generation, suggestions on style and coherence, section structuring, in-depth grammar checking, and the creation of multiple abstracts. Writing becomes more organized and coherent, and errors in scientific reasoning and tables are reduced.

Analyses show that the impact on user learning is enhanced by implementing these tools alongside tutorial support or user guides. Tutorials provide the educational component by guiding users on certain methodological criteria, scientific structures, and the ethical use of AI, which reduces the risk of overreliance and ensures that the content remains rigorous. However, limitations are evident. AI can provide incomplete and inaccurate responses; some systems lack contextual understanding, and carelessness can affect the originality and attribution of manuscripts.

The most effective automated tools for scientific writing and the responsible use of each of these tools, generative AI along with tutoring. Generative AI not only improves the quality of manuscripts but also optimizes production times and the acquisition of academic writing skills. This becomes more relevant for students, researchers, and educators who seek to increase the productivity of their work.

Future research should provide robust empirical evidence regarding the effects of AI-driven automation and integration systems on the quality, originality, and scholarly value of academic manuscripts. In addition, further studies should explore the development of more personalized and hybrid models that incorporate adaptive tutoring systems tailored to users' proficiency levels and specific learning needs. Furthermore, continued progress in the development of ethical frameworks and guidelines for the responsible adoption of artificial intelligence in academic writing is essential to ensure transparency, reliability, and alignment with the principles of scientific integrity. Such advancements will enable greater levels of automation while effectively addressing the associated risks, thereby fostering a model of knowledge production that is both technologically supported and ethically balanced.

REFERENCES

Abraham S. Accad. (2025). Ingrating perplexity ai in pre-writing process: a smart approach to boosting academic writing. *International Journal of Applied Mathematics*, *38*(3s), 347–360. https://doi.org/10.12732/ijam.v38i3s.151

Akpan, I. J., Kobara, Y. M., Owolabi, J., Akpan, A. A., & Offodile, O. F. (2025). Conversational and generative artificial intelligence and human–chatbot interaction in education and research. *International Transactions in Operational Research*, *32*(3), 1251–1281. https://doi.org/10.1111/itor.13522

Aljuaid, H. (2024). The Impact of Artificial Intelligence Tools on Academic Writing Instruction in Higher Education: A Systematic Review. *Arab World English Journal*, *1*(1), 26–55. https://doi.org/10.24093/awej/ChatGPT.2

Anani, G. E., Nyamekye, E., & Bafour-Koduah, D. (2025). Using artificial intelligence for academic writing in higher education: the perspectives of university students in Ghana. *Discover Education*, *4*(1), 46.

https://doi.org/10.1007/s44217-025-00434-5

Anusha, A., Sharma, S., & Sharma, S. (2025). Is ChatGPT the Future of Academic Writing? A Sequential Explanatory Study to Explore Generative Conversational AI as an Academic Writing Support Tool for Research Scholars. *Bulletin of Science, Technology & Society*, *45*(1–2), 23–39. https://doi.org/10.1177/02704676251353105

Aromataris, E., & Munn, Z. (2020). *Joanna Briggs Institute Reviewer's Manual.* . The Joanna Briggs Institute. https://reviewersmanual.joannabriggs.org

Barrot, J. S. (2025). Leveraging Google Gemini as a Research Writing Tool in Higher Education. *Technology, Knowledge and Learning*, *30*(1), 593–600. https://doi.org/10.1007/s10758-024-09774-x

Bozkurt, A. (2024). GenAI et al.: Cocreation, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. *Open Praxis*, *16*(1), 1–10. https://doi.org/10.55982/openpraxis.16.1.654

Cohen, J. F., & Moher, D. (2025). Generative artificial intelligence and academic writing: friend or foe? *Journal of Clinical Epidemiology*, *179*, 111646. https://doi.org/10.1016/j.jclinepi.2024.111646

Gong, J. (2022). Digital Teaching Research Based on the Intelligent Research and Training Platform: Citing the Practice of the Chinese Teaching and Research Group of Senior Secondary School Affiliated to Xingyi Normal University for Minorities as a Case Study. *Science Insights Education Frontiers*, *12*(2), 1741–1748. https://bonoi.org/index.php/sief/article/view/820

Ironsi, C. S., & Solomon Ironsi, S. (2025). Experimental evidence for the efficacy of generative AI in improving students' writing skills. *Quality Assurance in Education*, *33*(2), 237–252. https://doi.org/10.1108/QAE-04-2024-0065

Jin, F., Lin, C.-H., & Lai, C. (2025). Modeling AI-assisted writing: How self-regulated learning influences writing outcomes. *Computers in Human Behavior*, *165*, 108538. https://doi.org/10.1016/j.chb.2024.108538

Jin, X., Yuan, G., Wang, S., & Yu, J. (2025). Shareholding arrangement within controlling family and <scp>ESG</scp> performance: Insights from succession planning in <scp>C</scp> hinese family businesses. *International Review of Finance*, *25*(3). https://doi.org/10.1111/irfi.70028

Kong, S.-C., Lee, J. C.-K., & Tsang, O. (2024). A pedagogical design for self-regulated learning in academic writing using text-based generative artificial intelligence tools: 6-P pedagogy of plan, prompt, preview, produce, peer-review, portfolio-tracking. *Research and Practice in Technology Enhanced Learning*, *19*, 030. https://doi.org/10.58459/rptel.2024.19030

Le, H., Shen, Y., Li, Z., Xia, M., Tang, L., Li, X., Jia, J., Wang, Q., Gašević, D., & Fan, Y. (2025). Breaking human dominance: Investigating learners' preferences for learning feedback from generative <scp>AI</scp> and human tutors. *British Journal of Educational Technology*, *56*(5), 1758–1783. https://doi.org/10.1111/bjet.13614

McGuinness, L. A., & Higgins, J. P. T. (2021). Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research*

Synthesis Methods, *12*(1), 55–61. https://doi.org/10.1002/jrsm.1411

Melliti, M. (2024). AI in MA Thesis Writing: The Use of Lexical Patterns to Study the ChatGPT Influence. *International Journal of TESOL Studies*. https://doi.org/10.58304/ijts.20240305

Moorhouse, B. L., Wan, Y., Wu, C., Wu, M., & Ho, T. Y. (2025). Generative AI tools and empowerment in L2 academic writing. *System*, *133*, 103779. https://doi.org/10.1016/j.system.2025.103779

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, *49*(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593

Odeh, A., Salameh, H. B., & Elrefae, G. (2025). A high-accuracy classification model for AI-generated and AI-paraphrased texts. *Cluster Computing*, *28*(15), 952. https://doi.org/10.1007/s10586-025-05664-8

Ossa, C., & Willatt, C. (2023). Uso de Inteligencia Artificial Generativa para retroalimentar escritura académica en procesos de Formación Inicial Docente. *European Journal of Education and Psychology*, 1–16. https://doi.org/10.32457/ejep.v16i2.2412

Oubibi, M. (2025). Generative artificial intelligence: postgraduate student's digital adaptability, engagement and academic confidence. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-025-00364-z

Pack, A., & Maloney, J. (2023). Using Generative Artificial Intelligence for Language Education Research: Insights from Using <scp>OpenAI</scp> 's <scp>ChatGPT</scp>. *TESOL Quarterly*, *57*(4), 1571–1582. https://doi.org/10.1002/tesq.3253

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., … Alonso-Fernández, S. (2021). Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas. *Revista Española de Cardiología*, *74*(9), 790–799. https://doi.org/10.1016/j.recesp.2021.06.016

Pérez, A., McClain, S. K., Roa, A. F., Rosado-Mendinueta, N., Trigos-Carrillo, L., Robles, H., & Campo, O. (2025). Aplicaciones de inteligencia artificial en la escritura y la corrección académicas en la universidad: una revisión sistemática. *Íkala, Revista de Lenguaje y Cultura*, *30*(1). https://doi.org/10.17533/udea.ikala.355878

Peters, M. (2025). Leveraging Generative Artificial Intelligence with Transparency: Enhancing Academic Integrity in Higher Education. *Journal of Scholarly Publishing*, *56*(2), 510–528. https://doi.org/10.3138/jsp-2024-1124

Pividori, M., & Greene, C. S. (2024). A publishing infrastructure for Artificial Intelligence (AI)-assisted academic authoring. *Journal of the American Medical Informatics Association*, *31*(9), 2103–2113. https://doi.org/10.1093/jamia/ocae139

Pizarro-Romero, J., & Lovón, M. (2025). El uso de la IA en cursos de redacción e investigación universitaria en el aula: una experiencia de caso. *Desde El Sur*, *17*(1), e0015. https://doi.org/10.21142/DES-1701-2025-0015

Rashid, S., Malik, S., Abbas, F., & Khan, J. A. (2024). Pakistani students' perceptions about knowledge, use and impact of artificial intelligence (AI) on academic writing: a case study. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-024-00338-7

Saqib, M. B., & Zia, S. (2025). Evaluation of AI content generation tools for verification of academic integrity in higher education. *Journal of Applied Research in Higher Education*, *17*(4), 1430–1440. https://doi.org/10.1108/JARHE-10-2023-0470

Singh, J. K., Daniel, B. K., & Koh, J. H. L. (2025). Empowering Authorship with AI: a Novel Academic Writing Technology for Authorial Voice. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-025-00503-8

Sparkman, M., & Witt, A. (2025). Claude AI and Literature Reviews: An Experiment in Utility and Ethical Use. *Library Trends*, *73*(3), 355–380. https://doi.org/10.1353/lib.2025.a961199

Tang, A., Li, K., Kwok, K. O., Cao, L., Luong, S., & Tam, W. (2024). The importance of transparency: Declaring the use of generative artificial intelligence (<scp>AI</scp>) in academic writing. *Journal of Nursing Scholarship*, *56*(2), 314–318. https://doi.org/10.1111/jnu.12938

Tarchi, C., Zappoli, A., Casado Ledesma, L., & Brante, E. W. (2025). The Use of ChatGPT in Source-Based Writing Tasks. *International Journal of Artificial Intelligence in Education*, *35*(2), 858–878. https://doi.org/10.1007/s40593-024-00413-1

Teng, M. F. (2025). Understanding EFL student writers' metacognitive awareness in utilizing ChatGPT. *System*, *135*, 103848. https://doi.org/10.1016/j.system.2025.103848

Thohir, L., Udin, U., Isnaeni, M., & Putera, L. J. (2024). The Effectiveness of Integrating Quillbot for Paraphrasing in Writing Instruction. *Journal of Languages and Language Teaching*, *12*(4), 2112. https://doi.org/10.33394/jollt.v12i4.12780

Tran, T. T. T. (2025). Enhancing EFL Writing Revision Practices: The Impact of AI- and Teacher-Generated Feedback and Their Sequences. *Education Sciences*, *15*(2), 232. https://doi.org/10.3390/educsci15020232

Tseng, L.-P., Huang, L.-P., & Chen, W.-R. (2025). Exploring artificial intelligence literacy and the use of ChatGPT and copilot in instruction on nursing academic report writing. *Nurse Education Today*, *147*, 106570. https://doi.org/10.1016/j.nedt.2025.106570

Tseng, Y.-C., & Lin, Y.-H. (2024). Enhancing English as a Foreign Language (EFL) Learners' Writing with ChatGPT: A University-Level Course Design. *Electronic Journal of E-Learning*, *22*(2), 78–97. https://doi.org/10.34190/ejel.21.5.3329

Vanichvasin, P. (2021). Chatbot Development as a Digital Learning Tool to Increase Students' Research Knowledge. *International Education Studies*, *14*(2), 44. https://doi.org/10.5539/ies.v14n2p44

Wang, L., & Ren, B. (2024). Enhancing Academic Writing in a Linguistics Course with Generative AI: An Empirical Study in a Higher Education Institution in Hong Kong. *Education Sciences*, *14*(12), 1329.

https://doi.org/10.3390/educsci14121329

Weidmann, A. E. (2024). Artificial intelligence in academic writing and clinical pharmacy education: consequences and opportunities. *International Journal of Clinical Pharmacy*, *46*(3), 751–754. https://doi.org/10.1007/s11096-024-01705-1

Yang, J. J., & Hwang, S.-H. (2025). Transforming hematological research documentation with large language models: an approach to scientific writing and data analysis. *Blood Research*, *60*(1), 15. https://doi.org/10.1007/s44313-025-00062-w

Yang, Y., Huang, L., Lin, W., Li, Y., Xu, Y., & Cheng, L. (2025). Enhancing Sustainable English Writing Instruction Through a Generative AI-Based Virtual Teacher Within a Co-Regulated Learning Framework. *Sustainability*, *17*(19), 8770. https://doi.org/10.3390/su17198770

Zhou, S., & Du Preez, G. (2025). Bringing Cinderella to spotlight? Developing grammar through ChatGPT in writing. *Language Learning & Technology*, *29*(1), 1–25. https://doi.org/10.64152/10125/73639

Zizka, L. (2025). "It Looks Good Enough": Recognizing the Quality of Generative AI Output in Academic Writing Tasks in Higher Education. *Journal of Hospitality & Tourism Education*, 1–10. https://doi.org/10.1080/10963758.2025.2496663

CONFLICTS OF INTEREST

The authors declare no conflict of interest.