REVISTA INVECOM
ISSN en línea: 2739-0063

Generative AI and academic writing: A bibliometric analysis

Inteligencia artificial generativa y escritura académica: un análisis bibliométrico

Received: 17/08/2026 - **Accepted:** 28/10/2026

Gressia Victoria Valenzuela Támara
https://orcid.org/0000-0001-9890-9973
grvalenzuelata@ucvvirtual.edu.pe
Universidad César Vallejo. Lima, Perú

Jose Luis Grimaldo Rejas
https://orcid.org/0000-0002-7157-1252
jgrimaldo@ucvvirtual.edu.pe
Universidad César Vallejo. Lima, Perú

Gloria Milagros Ocampo Guerra
https://orcid.org/0000-0003-0757-5598
gocampog@ucvvirtual.edu.pe
Universidad César Vallejo. Lima, Perú

Jorge Luis Tomas Florez Salas
https://orcid.org/0009-0000-4102-498X
jflorezs@unam.edu.pe
Universidad Nacional de Moquegua. Tacna, Perú

How to cite this article:
Valenzuela Támara, G. V., Grimaldo Rejas, J. L., Ocampo Guerra, G. M., & Florez Salas, J. L. T. (2026). Generative AI and academic writing: A bibliometric analysis. *Revista InveCom*, *6*(Núm. Especial 3), 1-12. https://doi.org/10.5281/zenodo.24264225

Abstract

Although generative artificial intelligence (GenAI) has demonstrated growing potential to support academic writing, its relationship with metacognitive control has not yet been sufficiently explored across different educational levels. Therefore, this study examines the scientific literature on GenAI-supported academic writing through a descriptive bibliometric and thematic analysis. Twenty full-text reports were reviewed to determine their eligibility. Of these, sixteen articles published between 2024 and 2026 met the established criteria. Subsequently, descriptive statistics and a thematic co-occurrence analysis were employed. The data were organized according to their bibliographic characteristics and five potentially overlapping conceptual categories. Of the included articles, 11 (68.8%) were published in 2025, 13 (81.3%) involved higher education students, and only three (18.8%) focused on secondary school students. Likewise, nine articles (56.3%) were related to L2 writing and to metacognition or self-regulation. The affiliations of the first authors represented ten different geographical locations. The results reveal a concentration of evidence in the university setting and limited availability of comparable studies involving adolescents. Moreover, the adapted PRISMA 2020 flow diagram makes clear that the full-text assessment was limited and that no exhaustive search was conducted in indexed databases. Consequently, this study offers a descriptive and delimited analysis, not a global assessment of the impact of GenAI. Finally, it is a priority to examine how adolescents reread, revise, and critically evaluate their written work in real writing contexts.

Keywords: generative artificial intelligence, metacognitive monitoring, academic writing.

Resumen

Aunque la inteligencia artificial generativa (IAG) ha demostrado un potencial creciente para facilitar la escritura académica, su relación con el control metacognitivo aún no ha sido suficientemente explorada en los distintos niveles educativos. De allí que, esta investigación examina la producción científica sobre escritura académica apoyada en IAG mediante un análisis bibliométrico y temático descriptivo. Se revisaron veinte informes de texto completo para determinar su elegibilidad. De ellos, dieciséis artículos publicados entre 2024 y 2026 cumplieron los criterios establecidos. Posteriormente, se emplearon estadísticas descriptivas y un análisis de coocurrencia temática. Los datos se organizaron según sus características bibliográficas y cinco categorías conceptuales que podían superponerse. De los artículos incluidos, 11 (68,8%) se publicaron en 2025, 13 (81,3%) involucraron a estudiantes de educación superior y solo tres (18,8%) se centraron en estudiantes de secundaria. Asimismo, nueve artículos (56,3%) se relacionaron con la escritura en lengua/L2 y con la metacognición o autorregulación. Las afiliaciones de los primeros autores representaron diez ubicaciones geográficas distintas. Los resultados evidencian una concentración de la evidencia en el ámbito universitario y una escasa disponibilidad de estudios comparables con adolescentes. Además, el diagrama adaptado de PRISMA 2020 deja claro que la evaluación de textos completos fue limitada y que no se realizó una búsqueda exhaustiva en bases de datos indexadas. Por consiguiente, este estudio ofrece un análisis descriptivo y delimitado, no una valoración global del impacto de la IAG. Finalmente, resulta prioritario examinar cómo los adolescentes releen, revisan y evalúan críticamente sus producciones en contextos reales de escritura.

Palabras clave: inteligencia artificial generativa, monitoreo metacognitivo, escritura académica.

Introduction

Large language models have transformed the practical settings in which students generate ideas, compose paragraphs, revise arguments, and receive feedback. Writing with GenAI may make feedback and linguistic expression easier, but it should not be mistaken for independent judgment. A learner may produce more polished writing while exerting minimal control over the evidence, authorship, or reasoning required to assess machine-generated ideas. Thus, the educational challenge is not only whether GenAI is used, but how students monitor its use across different phases of the writing process (Teng, 2025; Stofiana et al., 2025; Fan et al., 2025).

Metacognition refers to knowledge and monitoring of one's cognition, while self-regulated learning involves the planning, strategic action, and reflection through which learners pursue their objectives (Flavell, 1979; Zimmerman, 2002). In AI-assisted writing, these processes involve selecting appropriate prompts, evaluating different outputs, identifying unsupported claims, assessing the reliability of feedback, and determining whether revisions strengthen the argument. This is not the same as being able to generate fluent writing or to use a chatbot frequently. This distinction is particularly important in educational studies that report changes in the quality of a final text without examining students' decision-making processes.

Recent empirical research has documented both possibilities. AI-supported metacognitive prompts have been explored in both collaborative and individual writing contexts (Haidar et al., 2025; Apriyadi et al., 2025). Teng (2025) examined students' knowledge of using ChatGPT for academic-writing feedback, while El Hosayny et al. (2026) investigated postgraduate students' strategic engagement through qualitative interviews. Fan and colleagues (2025) investigated the effects of different learning aids on motivation and performance in a writing task. Other studies complement the feedback system and examine essay quality, including research on AI-generated feedback (Zhang, 2025), picture-cued school writing (Zhuang et al., 2025), and ESL writing instruction (Mahapatra, 2024). These approaches are not interchangeable. For example, a feedback intervention may address product quality, whereas a metacognitive intervention addresses students' monitoring and control, as well as the completed product.

There is also a significant educational difference in the ages of the participants. Research involving undergraduate and postgraduate participants can illuminate processes and inform design decisions, but it cannot directly establish the same effects among adolescents. In secondary-school settings, teacher supervision, assessment procedures, digital access, age-related development, and school policies may affect students' capacity to use AI critically. Therefore, a well-scoped review should report the educational level of the sample studied, rather than generalizing university-level findings to secondary schools. This distinction is particularly important for quality education and for the United Nations Sustainable Development Goal 4, which seeks equitable learning opportunities and the development of relevant skills.

Bibliometric analysis typically quantifies research outputs and can examine authors, publication sources, collaboration patterns, citation relationships, and conceptual associations (Aria & Cuccurullo, 2017; Donthu et al., 2021). The quality and breadth of these approaches depend on the bibliographic records on which they rely. Moreover, when a comprehensive database collection is not available, it is still possible to define a set of publications through a smaller, explicitly bounded bibliometric profile, provided that the scope of inference is limited. For instance, a small purposive sample should not be presented as representative of the total output of Scopus, Web of Science, or the international research community.

Against this background, this study asks: What patterns of publication timing, authorship, institutional geography, educational level, and thematic emphasis can be observed in a focused set of research publications on GenAI-supported academic writing during 2024–2026? The objectives are to (a) describe the distribution of publications over time and by source; (b) examine author participation and the geographical locations of first-author affiliations; (c) distinguish research involving secondary-school students from research involving higher-education students; and (d) map recurring conceptual themes and their co-occurrence. This analytical scope is intentionally descriptive and does not make claims about causal effects or global scientific impact.

Materials and methods

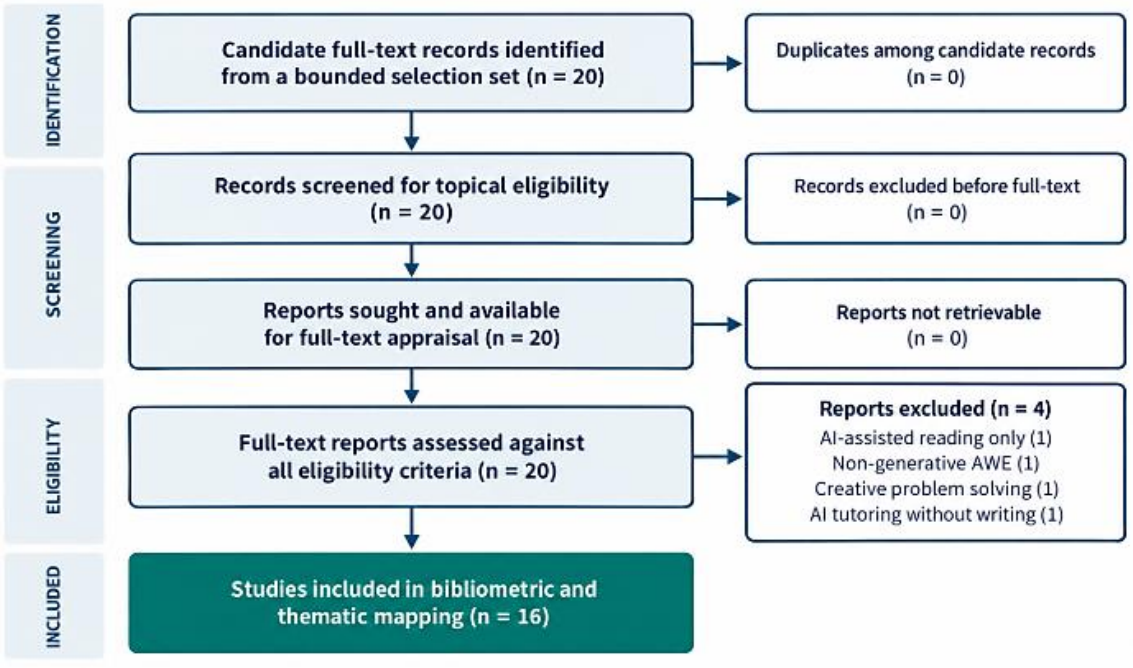
A descriptive, retrospective bibliometric profile and a structured thematic map were developed using article titles, author keywords, and abstracts. A single journal article was used as the unit of analysis. The observation period comprised articles assigned to journal issues dated 2024, 2025, or 2026 and available by 18 October 2026. The corpus was deliberately constructed from a specified collection of 20 candidate works at the intersection of GenAI, student writing, and educational practice; it was not compiled through a documented, exhaustive search of an indexed database. Therefore, the study establishes its own sample frame and does not calculate field-wide publication rates, citation impact, or coverage-adjusted national productivity.

Eligibility criteria and full-text verification

Records were eligible if they addressed a generative artificial intelligence (AI) technology used in an educational writing task or a closely related writing-learning activity and had an identifiable journal-article record, including a title, author list, publication year, and publication venue. Studies conducted in both school and university settings were retained to enable comparison across educational levels. Studies were excluded if they addressed only AI-assisted reading comprehension, traditional automated evaluation without recognizable generative AI, creative problem-solving without a writing-learning focus, or general AI tutoring without a writing objective. Titles and DOIs were verified against the published front matter and, where available, journal issue information was prioritized over the year of an earlier accepted manuscript or repository version. For example, the published version of Chen et al. (2026) was used instead of the preprint year, and the issue year was used for publications released online before an issue was assigned.

In total, 20 distinct potential full-text reports were examined. This restricted set contained no duplicates. All 20 reports proceeded to full-text screening, and four were excluded, each for one primary eligibility reason: one study was limited to AI-assisted reading comprehension, one examined traditional automated writing evaluation without clear use of generative AI, one addressed creative problem-solving rather than writing learning, and one was a general AI-tutoring study focused on programming tasks. The final analytical corpus consisted of the remaining 16 records. The available numbers are presented in Figure 1 as an adapted PRISMA 2020 flow diagram (Page et al., 2021). This figure should not be described as a complete Scopus or Web of Science search flow, as the history of the initial bibliographic search and the number of records identified in each database were not documented.

Figure 1
PRISMA 2020-adapted eligibility flow for the bounded full-text corpus



Metadata and thematic coding

For each included article, the following information was extracted: document identifier, publication year, title, publication venue, DOI, number of listed authors, first-author institutional location, participant educational level, and research design. Where the journal publication differed from an earlier version of the article, the bibliographic information was harmonized with the journal version of record. Institutional geography was recorded only on the basis of the first author's affiliation to avoid double-counting internationally co-authored publications. For descriptive purposes, Hong Kong SAR and Macao SAR were retained as separate affiliation locations and were not treated as sovereign countries.

Five non-exclusive thematic codes were defined: (1) metacognition/self-regulated learning, for explicit monitoring, awareness, strategic regulation, or self-regulation; (2) feedback/evaluation, for AI-supported formative feedback or writing evaluation; (3) collaboration/analytics, for peer or human–AI collaboration, dashboards, and learning analytics; (4) argumentation/criticality, for argumentative writing, critical writing, and evidence-based reasoning; and (5) language/L2, for explicitly language-learning or second-/foreign-language writing contexts. Codes were assigned according to the clear emphasis of each paper, as reflected in its title, abstract, and keywords, rather than by automatically inferring all concepts present in the background literature. An article could receive more than one code.

Indicators and analytical boundaries

Frequencies, percentages, mean authorship, and a 5 × 5 document-level co-occurrence matrix were calculated from the verified record sheet. The co-occurrence value for two themes corresponds to the number of publications coded for both themes; diagonal cells correspond to the number of publications coded for a given theme. Because the themes overlap, theme totals may legitimately exceed 16. The analysis does not report total citations, *h*-indices, co-citation clusters, bibliographic coupling, or institutional citation rankings, as reliable, standardized cited-by counts and cited-reference metadata were not available for this corpus. Apparent increases across 2024–2026 are interpreted as patterns within the included set, not as growth rates for the global scientific literature. The year 2026 was incomplete at the cut-off date.

Results

The total dataset comprised 16 journal articles published in 15 separate journals. Eleven articles (68.75%) were assigned to 2025, three (18.75%) to 2026, and two (12.50%) to 2024. The distribution is heavily concentrated in 2025 (Figure 2), but this concentration is related to the selection frame and should not be interpreted as a global increase in indexed publishing. The tally for 2026 is incomplete. *The International Journal of Learning, Teaching and Educational Research* is the only venue represented twice, while all other venues contributed one article each. The wide range of venues indicates that the selected studies cover education, applied linguistics, learning analytics, and educational technology, and are not restricted to a single specialized publication.

Figure 2

Annual publication counts in the selected corpus

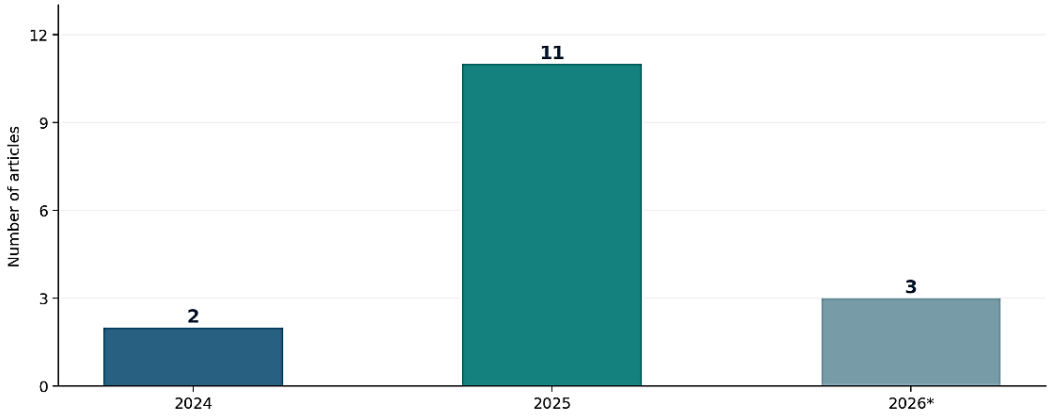


Table 1

Core descriptive indicators of the analyzed publications

Indicator	Value
Publications in analytic corpus	16
Journal titles	15
Publication years	2024: 2 (12.50%); 2025: 11 (68.75%); 2026: 3 (18.75%)
Named authorship occurrences	46 across 16 papers
Mean authors per paper	2.88
Single-author articles	5 (31.25%)
Secondary-level studies	3 (18.75%)
Higher-education studies	13 (81.25%)

Note. Percentages are based on the 16 included publications unless otherwise indicated. Authorship occurrences are not disambiguated researcher identities

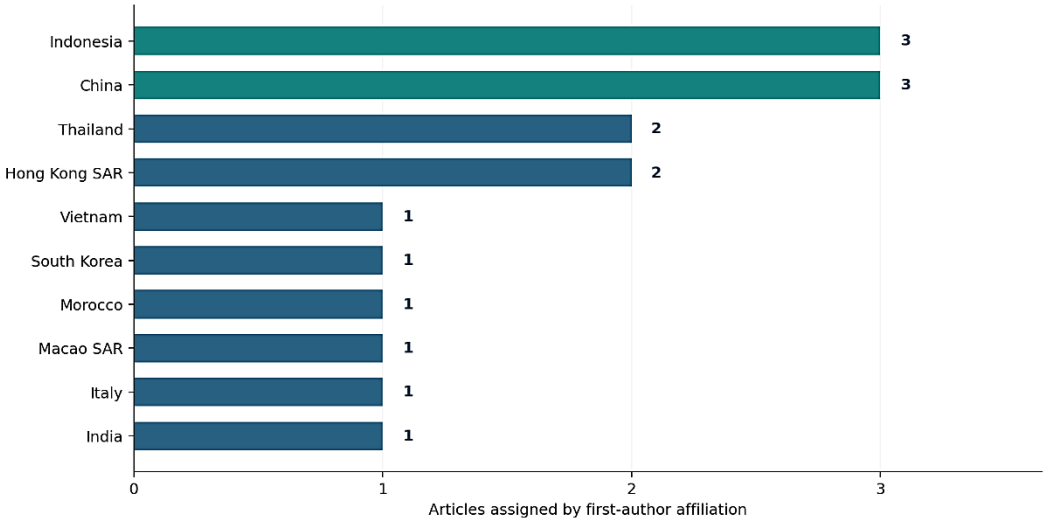
Authorship and geographic distribution

The 16 articles included 46 named authorship occurrences, corresponding to 2.88 authors per article. Eleven articles (68.75%) were multi-authored, while five were single-authored (31.25%). A conservative name-based comparison identified one author who was evidently repeated, Gunawan Suryoputro, who appeared in publications led by Apriyadi and Haidar. Accordingly, this recurrent authorship was normalized as one individual, resulting in 45 distinct author names. This name-based count is not equivalent to a disambiguated Scopus Author ID analysis, as verified author identities were not available across all records. The observed distribution does not support a classification of the most prominent researchers in the field.

The first authors' affiliations were located in 10 different places (Figure 3). China and Indonesia each accounted for three articles (18.75%), while Thailand and Hong Kong SAR each accounted for two (12.50%). Six other locations each accounted for one article (6.25%): India, Italy, Macao SAR, Morocco, South Korea, and

Vietnam. These are descriptors of the selected sample, not measures of population-adjusted productivity or national research capacity. Co-authorship across multiple locations, involving authors from different institutional contexts, is still evident in the metadata of individual articles but was not double-counted in Figure 3.

Figure 3
Publication counts by first-author institutional location



Note. Each article is counted once by first-author affiliation. Hong Kong SAR and Macao SAR are affiliation locations and are not classified here as sovereign countries

Table 2
First-author affiliation locations of the included articles

Institutional location	Articles	Proportion
China	3	18.8%
Indonesia	3	18.8%
Hong Kong SAR	2	12.5%
Thailand	2	12.5%
India	1	6.2%
Italy	1	6.2%
Macao SAR	1	6.2%
Morocco	1	6.2%
South Korea	1	6.2%
Vietnam	1	6.2%

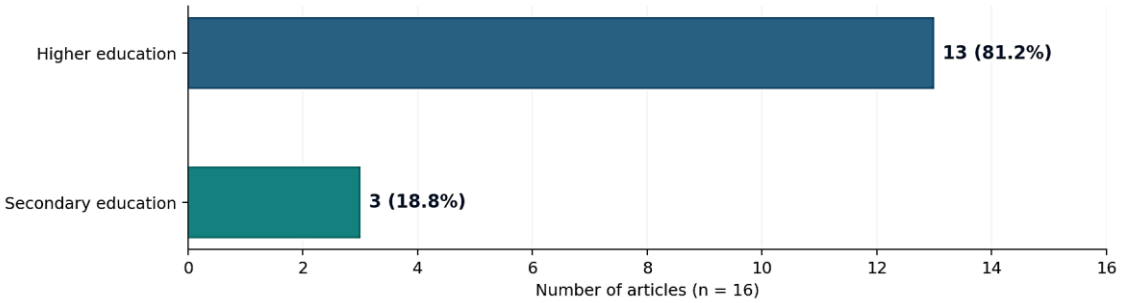
Note. Publication counts are whole-article counts, with one location assigned per article. This is not a measure of citation impact or international collaboration

Educational setting and secondary-school research

The most important distinction in this sample concerns the educational level of the participants: 13 studies (81.25%) examined students enrolled in higher education, while only three articles (18.75%) focused on secondary-school contexts. The three secondary-school studies involved participants in a wide range of writing scenarios. Zhuang et al. (2025) examined the use of a system implemented in secondary schools throughout Hong Kong to explore generative feedback for picture-cued writing. Apriyadi et al. (2025) conducted a study on the use of digital metacognitive prompts and ethnopedagogical strategies with secondary-school students in Indonesia. Haidar et al.

(2025) examined the implementation of metacognitive prompts in writing tasks for high-school students. These projects included both collaborative and independent writing tasks. They differ in their explicit approaches to monitoring and intervention, as well as in how they assess outcomes, but they share a common interest in AI-supported writing.

Figure 4
Distribution of studies by educational level



Note. The classification refers to the participants studied.
Findings concerning university students are not extrapolated to adolescents

Table 3
Included studies involving secondary-school learners

Study	Setting	Relevant learning task
Zhuang et al. (2025)	Hong Kong secondary schools	GenAI feedback for picture-cued English writing
Apriyadi et al. (2025)	Indonesian high school	Digital metacognitive prompts and contextualized writing
Haidar et al. (2025)	Indonesian high school	GenAI metacognitive prompting in collaborative and independent writing

Note. These three publications directly address secondary schooling, but their designs, outcomes, and contexts differ; this table is descriptive rather than a pooled-effect comparison

These three studies support the existence of intervention research in school writing, but they do not establish the overall prevalence of GenAI use, long-term cognitive outcomes, or causal mediation through monitoring across secondary-school systems. In particular, findings from undergraduate studies should not be treated as direct evidence about adolescents. A targeted database search restricted to secondary schooling would be required to determine whether the observed scarcity is due to publication coverage, study selection, or a genuine evidence gap.

Topical profile and document-level co-occurrence

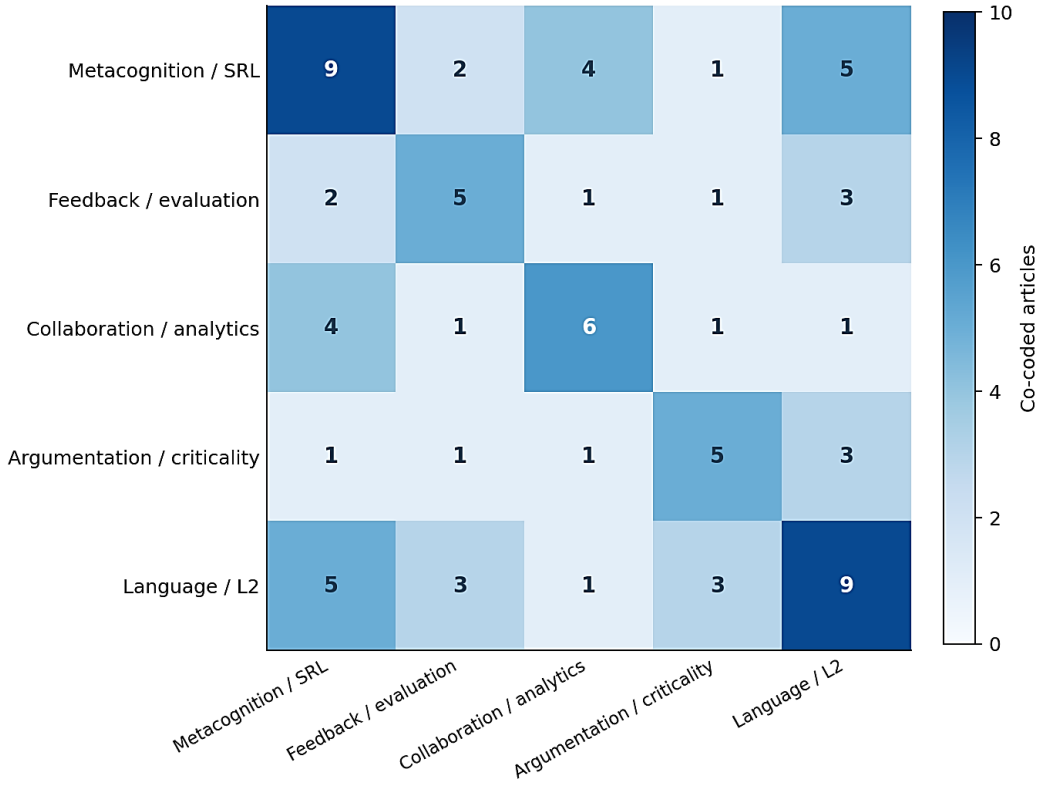
Nine publications (56.25%) were coded for metacognition/self-regulated learning, and nine (56.25%) for language/L2 settings. Collaboration/analytics was coded in six articles (37.50%), while feedback/evaluation and argumentation/criticality were each coded in five (31.25%). These are descriptive codes based on the specified emphases of the research, not bibliometric author-keyword frequencies downloaded directly from Scopus. The codes should be understood as transparent, content-based descriptors rather than probabilistic subject classifications generated by an algorithm. Table 4 explains the coding rationale, and Figure 5 shows the intersections between coded categories.

Table 4
Frequency of non-exclusive thematic categories

Theme	Articles	Share	Coding rule
Metacognition / SRL	9	56.25%	Monitoring, reflection, regulation or SRL
Language / L2	9	56.25%	Language-learning or EFL/ESL/L2 writing
Collaboration / analytics	6	37.50%	Human/peer–AI collaboration or dashboards
Feedback / evaluation	5	31.25%	AI-based writing feedback or assessment
Argumentation / criticality	5	31.25%	Argument, evidence or critical writing

Note. Codes were assigned manually according to explicit study emphases.
Articles may belong to more than one category; therefore, the summed shares exceed 100%

Figure 5
Co-occurrence matrix of manually coded research themes



Some studies examined the cognitive oversight of human–AI collaboration, while others focused on feedback and the linguistic quality of the output. Methodological and conceptual integration across studies is represented in the co-occurrence matrix. Some studies examined the cognitive supervision of human–AI collaboration. An article may span several subfields and may not fully convey the relative importance of each theme individually. Within the existing body of work, studies of metacognition and language learning frequently occur within the broader topic of writing instruction. However, this descriptive overlap should not be interpreted as evidence of mediation, explanatory direction, or instructional effectiveness.

Eligibility decisions and excluded studies

Full-text exclusions were determined according to the stated eligibility criteria. Exactly one primary reason was assigned to each excluded report, and no excluded study contributed to the 16-record denominator.

Table 5
Full-text articles excluded during eligibility assessment

Report	Primary exclusion reason
Supardan & Capati (2026), AI as a Metacognitive Mirror	Reading-comprehension monitoring; no writing-learning task
Wale (2024), Artificial intelligence in education: automated writing evaluation	Conventional automated writing evaluation; GenAI not established as the writing intervention
Urban et al. (2024), ChatGPT improves creative problem-solving	Creative problem-solving rather than writing instruction
Yusuf et al. (2026), The impact of generative AI tutors	AI-supported programming/problem solving, without an educational writing task

Note. The authors and titles are provided to document the four excluded full-text reports, not as studies supporting the results

Bibliographic record audit

Table 6
Bibliographic record identifiers, years and DOI values

ID	Author	Year	Education	DOI
B01	Fan	2025	Higher	10.1111/bjet.13544
B02	Chen et al.	2026	Higher	10.1016/j.iheduc.2025.101054
B03	Hong et al.	2025	Higher	10.30564/fls.v7i7.10072
B04	Wang & Ren	2024	Higher	10.3390/educsci14121329
B05	Zhang	2025	Higher	10.3390/bs15050600
B06	Zhuang et al.	2025	Secondary	10.1016/j.caeai.2025.100450
B07	Chen & Shim	2026	Higher	10.26803/ijlter.25.3.1
B08	Mahapatra	2024	Higher	10.1186/s40561-024-00295-9
B09	El Hosayny et al.	2026	Higher	10.1007/s44217-026-01282-7
B10	Tran	2025	Higher	10.54855/ijaile.25237
B11	Khampusaen	2025	Higher	10.70730/PGCQ9242
B12	Apriyadi et al.	2025	Secondary	10.23887/jpi-undiksha.v14i3.95320
B13	Tarchi et al.	2025	Higher	10.1007/s40593-024-00413-1
B14	Teng	2025	Higher	10.1016/j.system.2025.103848
B15	Stofiana et al.	2025	Higher	10.1016/j.amper.2025.100242
B16	Haidar et al.	2025	Secondary	10.26803/ijlter.24.6.11

Note. DOI identifiers follow the article-level coded dataset.
Educational level is determined from the study sample; full citations appear in the reference list

Discussion

The main observation is a thematic imbalance rather than a bibliometric impact hierarchy. Metacognition and language-related writing are salient themes across the restricted collection of articles, although comparatively few studies focus on adolescent learners. This is significant because secondary-school writing involves attention to the development of academic judgment and pedagogical monitoring. Evidence relevant to the explicit design of metacognitive prompts is provided by Apriyadi et al. (2025) and Haidar et al. (2025), while Zhuang et al. (2025) offer a complementary perspective on feedback-mediated revision. The three studies using school-based data do

not provide a sufficient basis for general statements about long-term independence or the transfer of self-regulation across disciplines.

The conceptual picture emerging from research in higher education is more differentiated. Teng (2025) and El Hosayny et al. (2026) focus their investigations on awareness and strategic engagement, whereas Chen and Shim (2026) evaluate writing performance in terms of self-efficacy and metacognitive strategy use. Fan et al. (2025) consider the possibility that external assistance may alter the way students expend cognitive effort. These contrasts help explain why the instructional benefits of GenAI cannot be reduced to a single outcome, such as an increase in writing scores. Patterns of change in product quality, cognitive engagement, source evaluation, learning motivation, and student autonomy may differ depending on the task structure and the assistance provided to the learner.

There is also a commonality in the way interventions are designed. GenAI feedback can support the editing process and provide similar writing assistance (Mahapatra, 2024; Zhang, 2025; Zhuang et al., 2025). According to Chen et al. (2026), Apriyadi et al. (2025), and Haidar et al. (2025), prompts and learning analytics can support learners in explaining their decisions or monitoring their progress. This has been shown by several studies. These findings point to a practical design principle: both reflective habits and creative products should be assessed. Students may be asked to discuss the advantages and disadvantages of ideas accepted or rejected with the assistance of artificial intelligence (AI), support arguments with evidence from external sources, track changes, or compare independently produced texts with AI-assisted drafts. The effectiveness outcomes of these proposals are the product of a synthesis, rather than the result of new effectiveness assessments, gathered through this bibliometric search.

There are also concerns regarding the geographical distribution of the sample. The concentrated character of the publications is not surprising, given that Asian institutions account for a large number of first-author affiliations. It is important to emphasize that the absence of some regions does not imply a lack of participation by scholars working in other contexts. More comprehensive regional comparisons could be obtained through improved retrieval from Scopus and Web of Science, while preserving query dates, language filters, and institutional identifiers. A complete bibliometric analysis would enable bibliographic coupling, co-citation analysis, standardized citation metrics, and the longitudinal evolution of themes. These indicators cannot be credibly reconstructed from a selective set of article records alone (Donthu et al., 2021).

This study has several important methodological limitations. First, the 20 candidate documents do not constitute a comprehensive or random sample of the population, and the full history of the database search is not provided. Second, the small number of retained articles makes differences in rank order unstable. Third, the institutional location of the first author is only one of several possible ways of attributing geography and overlooks some transnational collaborations. Fourth, topic assignment was rule-based but manually interpreted and may reasonably differ from another coder's classification of borderline studies. Fifth, official journal publication dates may differ from earlier online publication dates. Finally, citation counts and reference lists were not included in a standardized bibliographic export; therefore, we deliberately chose not to assess scientific influence, field-wide collaboration, or citation networks. These safeguards protect against misleading precision and enhance the interpretability of the descriptive data.

For a future comprehensive review, the search should be documented in at least one indexed database using fielded keywords and specific filters. Scopus TITLE-ABS-KEY searches could combine *generative AI* or *ChatGPT* with writing-related terms and *metacognition* or *self-regulation*, using a separate set of terms for secondary school. Retrieved citation records, including authors and affiliations, author keywords, citations, and references, should be exported. The records should be deduplicated by DOI and made auditable. A larger corpus could then be analyzed using Bibliometrix or VOSviewer and compared with the article-level categories described above. This is guidance for future replication, not a claim that these database procedures generated the current dataset.

Conclusions

The 16 articles considered here provide an interdisciplinary, but uneven, evidence profile. Most of the selected publications were published in 2025, and most involved university students. Only three explicitly examined secondary-school writing. Metacognition/self-regulated learning and language-oriented writing were the most prominent conceptual focuses in the coded sample. These results address the descriptive mapping objectives for the sampled articles and identify areas where secondary-education-focused research could make a meaningful contribution.

The main scientific conclusion is that AI-supported writing should be studied as both a textual and a metacognitive activity. Improvements in essays or in the perceived usefulness of feedback do not, in themselves,

demonstrate increased autonomous monitoring, critical evaluation, or sustained learning. Studies that link visible revision decisions, appropriate reliance on AI, and explicit measures of metacognitive control would provide stronger evidence than comparisons based on outcomes alone.

A further particularly valuable step would be a registered and reproducible bibliometric study conducted in an indexed database, followed by empirical research in secondary schools that combines writing assessments with think-aloud data, revision histories, and teacher-guided verification tasks. Such a research program would determine whether the underrepresentation of school settings persists under broader retrieval conditions and whether metacognitive support changes the educational outcomes of generative assistance.

References

Apriyadi, H., Suryoputro, G., & Tarmini, W. (2025). The role of digital metacognitive prompts assisted by Gen-AI and ethnopedagogical approaches to improve writing skills. *Jurnal Pendidikan Indonesia, 14*(3), 657–668. https://doi.org/10.23887/jpi-undiksha.v14i3.95320

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11(*4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Chen, A., Lian, J., Kuang, X., & Jia, J. (2026). Can theory-driven learning analytics dashboard enhance human-AI collaboration in writing learning? Insights from an empirical experiment. *The Internet and Higher Education, 68*, 101054. https://doi.org/10.1016/j.iheduc.2025.101054

Chen, Y., & Shim, J. (2026). From GenAI use to L2 writing performance: A moderated sequential mediation model of self-efficacy and metacognitive strategy use. *International Journal of Learning, Teaching and Educational Research, 25*(3), 1–18. https://doi.org/10.26803/ijlter.25.3.1

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

El Hosayny, O., Razkane, H., & Yeou, M. (2026). Metacognitive regulation and ChatGPT-assisted writing: Introducing an interactive model for cultivating strategic engagement. *Discover Education, 5*, Article 246. https://doi.org/10.1007/s44217-026-01282-7

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Haidar, H., Suryoputro, G., & Safi'i, I. (2025). Impact of the integration of metacognitive prompts by generative artificial intelligence (GenAI) in collaborative and individual learning in improving writing skills and metacognitive awareness. *International Journal of Learning, Teaching and Educational Research, 24*(6), 232–250. https://doi.org/10.26803/ijlter.24.6.11

Hong, H., Vate-U-Lan, P., & Viriyavejakul, C. (2025). Cognitive offload instruction with generative AI: A quasi-experimental study on critical thinking gains in English writing. *Forum for Linguistic Studies, 7*(7), 325–334. https://doi.org/10.30564/fls.v7i7.10072

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

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

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

Stofiana, T., Sunendar, D., Mulyati, Y., & Sastromiharjo, A. (2025). Writing with AI, thinking with Toulmin: Metacognitive gaps and the rhetorical limits of argumentation. *Ampersand, 15*, 100242. https://doi.org/10.1016/j.amper.2025.100242

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

Tran, T. T. (2025). The effects of AI-supported metacognitive strategy instruction on the writing performance of non-English majors. *International Journal of AI in Language Education, 2*(3), 101–116. https://doi.org/10.54855/ijaile.25237

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

Zhang, K. (2025). Enhancing critical writing through AI feedback: A randomized control study. *Behavioral Sciences, 15*(5), 600. https://doi.org/10.3390/bs15050600

Zhuang, Y., Zhao, R., Xie, Z., & Yu, P. L. H. (2025). Enhancing language learning through generative AI feedback on picture-cued writing tasks. *Computers & Education: Artificial Intelligence, 9*, 100450. https://doi.org/10.1016/j.caeai.2025.100450

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

AUTHOR CONTRIBUTIONS: All authors contributed to the conception and design of the bibliometric study, the definition of the search strategy, the processing and analysis of the corpus, the interpretation of the results, and the writing or critical revision of the manuscript. All authors approved the final version of the article.

CC BY