International Journal of
Humanities, Education, and Social Sciences

e-ISSN : 3026-0892
p-ISSN : 3026-1422

Index: Harvard, Boston, Sydney, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/IJHESS.v4i3.12544

Generative AI in Mathematics Learning: A Bibliometric Analysis and Narrative Review

Nur Ismiyati
Universitas Negeri Makassar, Indonesia
nurismiyati@unm.ac.id

Submitted:	Revised:	Accepted:	Published:
Sep 21, 2026	Oct 20, 2026	Oct 31, 2026	Nov 5, 2026

Abstract

Since late 2022, generative artificial intelligence (GenAI) has expanded rapidly in mathematics education, yet the research field remains insufficiently mapped, and evidence of its effects on learning is inconsistent. This study mapped the development, structure, and emerging topics of research on GenAI in mathematics learning and synthesized evidence on its uses, effects, and challenges. Bibliometric analysis was combined with a narrative review. Searches for records published between 2022 and 2026 in OpenAlex, ERIC, ScienceDirect, Google Scholar, and Garuda yielded 1,480 records, with 1,206 unique records retained after cleaning. Descriptive and text-mining analyses examined the full corpus, while citation and keyword co-occurrence analyses were conducted in VOSviewer using 758 OpenAlex records. The findings show that 95.2% of records were published from 2024 onward, 48.6% mentioned ChatGPT, and 29.5% explicitly addressed primary or secondary education. The keyword network comprised five clusters, with AI literacy, academic integrity, and self-regulation emerging as the newest topics. The narrative synthesis indicates that GenAI can support learning when integrated into scaffolded instruction with safeguards, whereas unrestricted use as an answer source can undermine learning. These findings highlight instructional design as central to the educational value of GenAI and inform the extension of scaffolding perspectives to AI-mediated mathematics learning. The study identifies school-level experimental research and teacher preparation as priorities for advancing evidence and guiding GenAI integration.

IJHESS Journal is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

Keywords: Bibliometric Analysis; ChatGPT; Generative Artificial Intelligence; Mathematics Learning; Narrative Review

INTRODUCTION

In the last decade, artificial intelligence (AI) has become an increasingly visible part of mathematics education, from intelligent tutoring systems to adaptive assessment (Hwang & Tu, 2021; Mohamed et al., 2022; Zawacki-Richter et al., 2019). The public release of ChatGPT in November 2022 marked a turning point, because generative AI (GenAI) built on large language models can explain concepts, produce worked solutions, and give feedback in natural language within seconds (Dwivedi et al., 2023; Kasneci et al., 2023). Within months, educators began to discuss its opportunities and threats for teaching, learning, and assessment (Baidoo-Anu & Owusu Ansah, 2023; Lo, 2023; Rudolph et al., 2023; Tlili et al., 2023), and international bodies issued guidance on the responsible use of GenAI in schools (UNESCO, 2023). In Indonesia, interest is further driven by persistently low mathematical literacy results (OECD, 2023), and early studies have explored ChatGPT in classrooms and teacher training (Dasari et al., 2024; Rifandi et al., 2025; Supriyadi, 2022).

This study argues that the value of GenAI for mathematics learning cannot be judged from the tool alone. GenAI is often associated with improved learning because it offers personalized explanations (Almarashdi et al., 2024; Pepin et al., 2025), supports problem solving through step-by-step scaffolding (Liu et al., 2025; Schorcht et al., 2024), and helps teachers prepare lessons and tasks (Egara & Mosimege, 2024; Taani & Alabidi, 2024). Students also report positive perceptions of ChatGPT as a learning aid (Chan & Hu, 2023; Egara et al., 2025). However, mathematics learning depends on students' own reasoning, so a tool that delivers ready-made answers may also replace the very thinking that instruction aims to develop.

Empirical evidence reflects this tension. ChatGPT-generated help has produced learning gains comparable to those of human tutors (Pardos & Bhandari, 2024), and meta-analyses report positive effects on learning performance and higher-order thinking (Deng et al., 2025; Wang & Fan, 2025; Wu & Yu, 2024). Other studies found no significant improvement when ChatGPT was added to problem-based learning (Dianta & Rizkianto, 2025), lower performance when students relied on ChatGPT without a teacher (Dasari et al.,

2024), and learning loss when students used GPT-4 without guardrails (Bastani et al., 2025). Previous reviews have examined AI or ChatGPT in mathematics education (Luzano, 2025; Muhamad Sani, 2026; Opesemowo & Adewuyi, 2024; Suprihatiningsih et al., 2026; Turmuzi et al., 2025), but most covered studies only up to 2024, rarely included Indonesian publications, and did not map how the field itself is structured. At the same time, the volume of publications is growing so quickly that a map of the whole field is needed before narrower syntheses can be positioned within it.

The novelty of this study lies in combining a bibliometric map of the field up to 2026, including the Indonesian Garuda index, with a narrative review that interprets key empirical studies through learning theory. Theoretically, the study draws on the concept of scaffolding, which is effective when support is gradually withdrawn so that learners take over the task (Wood et al., 1976); on research showing that ready-made solutions can bypass the cognitive processing needed for learning (Kirschner et al., 2006); on self-regulated learning (Zimmerman, 2002); and on technological pedagogical content knowledge (TPACK), which frames the teacher competence required to integrate GenAI (Mishra et al., 2023).

Accordingly, this study aims to map research on GenAI in mathematics learning published from 2022 to 2026 and to synthesize its main lessons. Specifically, it describes the growth and characteristics of the literature, including sources, languages, GenAI tools, and educational levels (RQ1); identifies the most cited documents and the thematic structure and emerging topics of the field (RQ2); and synthesizes what key empirical studies reveal about the uses, effects, and challenges of GenAI in mathematics learning (RQ3).

METHODS

1. Type of Research

This study combined a bibliometric analysis with a narrative literature review. The bibliometric component is quantitative and descriptive, following the guidelines of Donthu et al. (2021), and was used to answer RQ1 and RQ2. The narrative review is qualitative and interpretive (Snyder, 2019; Xiao & Watson, 2019) and was used to answer RQ3. The study relies entirely on secondary data in the form of published records.

2. Research Design

The study was conducted in three stages: (1) retrieval and cleaning of bibliographic records; (2) bibliometric analysis of the cleaned corpus, consisting of descriptive statistics, dictionary-based text mining, citation analysis, and keyword co-occurrence mapping; and (3) a narrative review of key empirical studies organized around the uses, effects, and challenges of GenAI. Unlike a systematic review, the study did not apply study-level screening or quality appraisal; the bibliometric results describe the field as retrieved by the search, while the narrative review interprets selected studies.

3. Data Sources and Corpus

Records were retrieved from five sources. OpenAlex, an open index of scholarly works (Priem et al., 2022), served as the main multidisciplinary database because its searches can be replicated without subscription. ERIC and ScienceDirect were searched directly, Google Scholar was searched through Publish or Perish with title-word queries, and Garuda was searched to capture Indonesian journals. The publication window was 2022 to 2026, and source-specific filters are reported in Table 1. For the narrative review, key studies were selected purposively from the corpus and the wider literature, prioritizing empirical studies conducted in school or mathematics teacher education settings, studies with experimental or comparative designs, and studies frequently cited in the field.

4. Search Strategy and Data Collection

The searches were conducted on 20 October 2026. The search strings, filters, and number of records retrieved from each source are presented in Table 1.

Table 1. Search Strings and Number of Records Retrieved

Source	Search string (field)	Filters	Records
OpenAlex	("generative AI" OR "generative artificial intelligence" OR "ChatGPT" OR "large language model" OR "GPT-4" OR "chatbot") AND ("mathematics education" OR "mathematics learning" OR "mathematics teaching" OR "mathematics instruction") (title, abstract, keywords)	2022-2026; article; English	758
ERIC	("generative AI" OR ChatGPT OR "large language model") AND mathematics	Peer reviewed; 2022 onward	205

Source	Search string (field)	Filters	Records
ScienceDirect	("generative AI" OR "ChatGPT" OR "large language model") AND ("mathematics education" OR "mathematics learning" OR "mathematics teaching") (title, abstract, keywords)	2022-2026; research articles	18
Google Scholar (Publish or Perish)	Title words: ChatGPT mathematics; "generative AI" mathematics; "large language model" mathematics; ChatGPT matematika	2022-2026	100; 174; 13; 61
Garuda	Title: ChatGPT matematika; kecerdasan buatan pembelajaran matematika; artificial intelligence matematika	2022-2026	40; 22; 89
Total			**1,480**

All records were exported in RIS or BibTeX format and archived without modification. Duplicates were identified by matching DOIs and normalized titles, followed by a manual check of partial-title matches, which removed 261 records. A further 13 records were removed during data cleaning because they were other versions of records already in the corpus, such as preprints, repository copies, supplementary appendices, or Google Scholar citation entries. The final corpus therefore contained 1,206 unique records.

5. Validity and Trustworthiness

Several procedures were used to strengthen the trustworthiness of the analysis. The OpenAlex query was saved as a filtered URL so that the search can be replicated, and all raw exports and the duplicate-removal log were retained as an audit trail. The text-mining dictionaries for GenAI tools and educational levels were defined before the analysis, and their matches were checked against random samples of titles and abstracts. In VOSviewer, a thesaurus file merged synonymous keywords before mapping. For the narrative review, the reported findings of each key study were checked against its original abstract or full text.

6. Data Analysis Technique

RQ1 was answered with descriptive frequencies of publication year, source, language, document type, and journal, computed in Python. GenAI tools and educational levels were identified through dictionary-based text mining of titles, abstracts, and keywords; a record could match more than one category, and 244 records (20.2%) had no abstract, so these counts indicate how often each category is mentioned rather than the exact number of studies. RQ2 was answered with OpenAlex citation counts as of 20 October 2026 and a keyword co-occurrence analysis in VOSviewer 1.6.21 (van Eck & Waltman, 2010). Both analyses used the 758 OpenAlex records, which carry standardized keywords and citation

data and were retrieved through the VOSviewer OpenAlex API. Full counting was applied, and a minimum of 20 occurrences retained 42 of 2,316 keywords. The overlay visualization used the average publication year of the documents containing each keyword to identify emerging topics. RQ3 was answered by organizing the findings of the key studies into three themes: uses, effects, and challenges.

RESULTS

1. Growth and Characteristics of the Literature

The annual number of records increased sharply after the release of ChatGPT (Figure 1). Only three records were dated 2022, compared with 43 in 2023, 202 in 2024, 377 in 2025, and 569 in 2026 up to 20 October. In total, 1,148 records (95.2%) were published from 2024 onward. Ten records had no publication year, and two were dated before 2022.

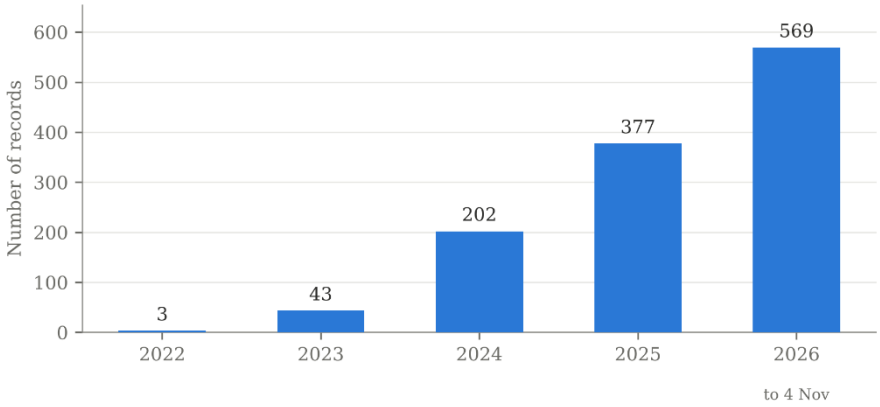


Figure 1. Annual Number of Records on GenAI in Mathematics Learning, 2022-2026 (N = 1,206)

Table 2 summarizes the characteristics of the corpus. Most records came from OpenAlex, were written in English, and were journal articles according to their source metadata. Indonesian-language records accounted for 15.6% of the corpus, and 75 records (6.2%) were reviews or bibliometric studies according to their titles.

Table 2. Characteristics of the Corpus (N = 1,206)

Characteristic	Category	n	%
Source (after deduplication)	OpenAlex	731	60.6
	Google Scholar	203	16.8
	ERIC	139	11.5
	Garuda	115	9.5

Characteristic	Category	n	%
	ScienceDirect	18	1.5
Language	English	983	81.5
	Indonesian	188	15.6
	Other	35	2.9
Document type	Journal article	996	82.6
	Proceedings, books, reports, repositories, preprints	136	11.3
	Unclear from metadata	74	6.1

ChatGPT dominated the tools mentioned in the corpus (Table 3). It appeared in 586 records (48.6%), whereas Gemini or Bard appeared in 59 records (4.9%) and other models in fewer than 2% each. A further 427 records (35.4%) referred only to GenAI or large language models in general. Regarding educational levels, higher education was the most frequently mentioned context (14.8%), followed closely by pre-service teachers (14.3%) and in-service school teachers (12.7%). Only 356 records (29.5%) explicitly mentioned a primary or secondary school context.

Table 3. GenAI Tools and Educational Levels Mentioned in Titles, Abstracts, and Keywords

Category	Mention	n	%
GenAI tool	ChatGPT / GPT models	586	48.6
	Gemini / Bard	59	4.9
	DeepSeek	20	1.7
	Claude	17	1.4
	LLaMA / Mistral / Qwen	11	0.9
	Microsoft Copilot	10	0.8
	GenAI or LLM in general only	427	35.4
Educational level	Higher education	179	14.8
	Pre-service teachers	172	14.3
	In-service school teachers	153	12.7
	Upper secondary	149	12.4
	Primary school	127	10.5
	Lower secondary	102	8.5
	Any primary or secondary school context	356	29.5

The records were published in a wide range of journals. The most productive journals were the *International Journal of Education in Mathematics, Science and Technology* (21 records) and *Pendas: Jurnal Ilmiah Pendidikan Dasar* (17 records), followed by *Frontiers in Education* (13), *Discover Education* (12), and *The Mathematical Education* (11) (Table 4).

Table 4. Most Productive Journals in the Corpus

Rank	Journal	Records
1	International Journal of Education in Mathematics, Science and Technology	21
2	Pendas: Jurnal Ilmiah Pendidikan Dasar	17
3	Frontiers in Education	13
4	Discover Education	12
5	The Mathematical Education	11
6	Digital Experiences in Mathematics Education	10
7	Journal of Computer Assisted Learning	10
8	International Electronic Journal of Mathematics Education	9
9	Social Sciences & Humanities Open	9
10	Education Sciences; ZDM; Education and Information Technologies	8 each

2. Influential Documents and Thematic Structure

The 758 OpenAlex records had received 4,484 citations in total, or 5.92 citations per record on average. Citations were highly concentrated: 466 records (61.5%) had not yet been cited, and 655 records (86.4%) were open access. Table 5 lists the ten most cited documents. The most cited document, a qualitative case study of ChatGPT in mathematics teaching (Wardat et al., 2023), had received 475 citations, more than twice as many as the second-ranked document.

Table 5. Ten Most Cited Documents in the OpenAlex Records

Rank	Document	Focus	Citations
1	Wardat et al. (2023)	Stakeholder perspectives and ChatGPT's limitations in mathematics teaching	475
2	Sánchez-Ruiz et al. (2023)	ChatGPT and blended learning in engineering mathematics	230
3	Lee & Yeo (2022)	AI chatbot for practicing responsive teaching (pre-service teachers)	211
4	Yilmaz et al. (2023)	Students' attitudes toward ChatGPT (technology acceptance)	155
5	Azaria et al. (2023)	ChatGPT's capabilities across domains, including mathematics	130
6	Hu et al. (2024)	Generating and evaluating teaching plans with GPT-4	115

Rank	Document	Focus	Citations
7	Wijaya et al. (2024)	AI literacy, trust, and dependency among mathematics teachers	103
8	Pardos & Bhandari (2024)	ChatGPT-generated help versus human tutor help	98
9	Chauncey & McKenna (2023)	Framework for ethical use of AI chatbots in teaching	91
10	Pepin et al. (2025)	Scoping survey of ChatGPT in mathematics education	89

The keyword co-occurrence map contained 42 keywords grouped into five clusters and connected by 686 links, with a total link strength of 4,742 (Table 6 and Figure 2). The most frequent keywords were *mathematics education* (398 occurrences), *ChatGPT* (337), *generative AI* (225), *large language models* (158), and *artificial intelligence in education* (148).

Table 6. Keyword Clusters in the Co-occurrence Map

Cluster	Items	Keywords	Theme
1	13	academic performance, AI chatbots, AI literacy, ChatGPT, critical thinking, generative AI in education, GeoGebra, high school students, junior high school students, mathematics learning, problem solving, student motivation, student self-regulation	ChatGPT and school students' learning outcomes
2	9	adaptive learning systems, artificial intelligence in education, chatbots, conceptual understanding, educational technology, intelligent tutoring systems, personalized learning, primary school mathematics, student engagement	Adaptive and personalized AI-supported learning
3	8	AI-assisted learning, Gemini, high school mathematics, junior high school mathematics, large language models, mathematical problem solving, mathematical reasoning, prompt engineering	Mathematical reasoning of large language models
4	8	AI in teaching, generative AI, mathematics education, overreliance on AI, pre-service mathematics teachers, pre-service teachers, teacher perceptions, teacher professional development	Teachers, teacher preparation, and over-reliance
5	4	academic integrity, AI ethics, higher education, student perceptions	Ethics and integrity in higher education

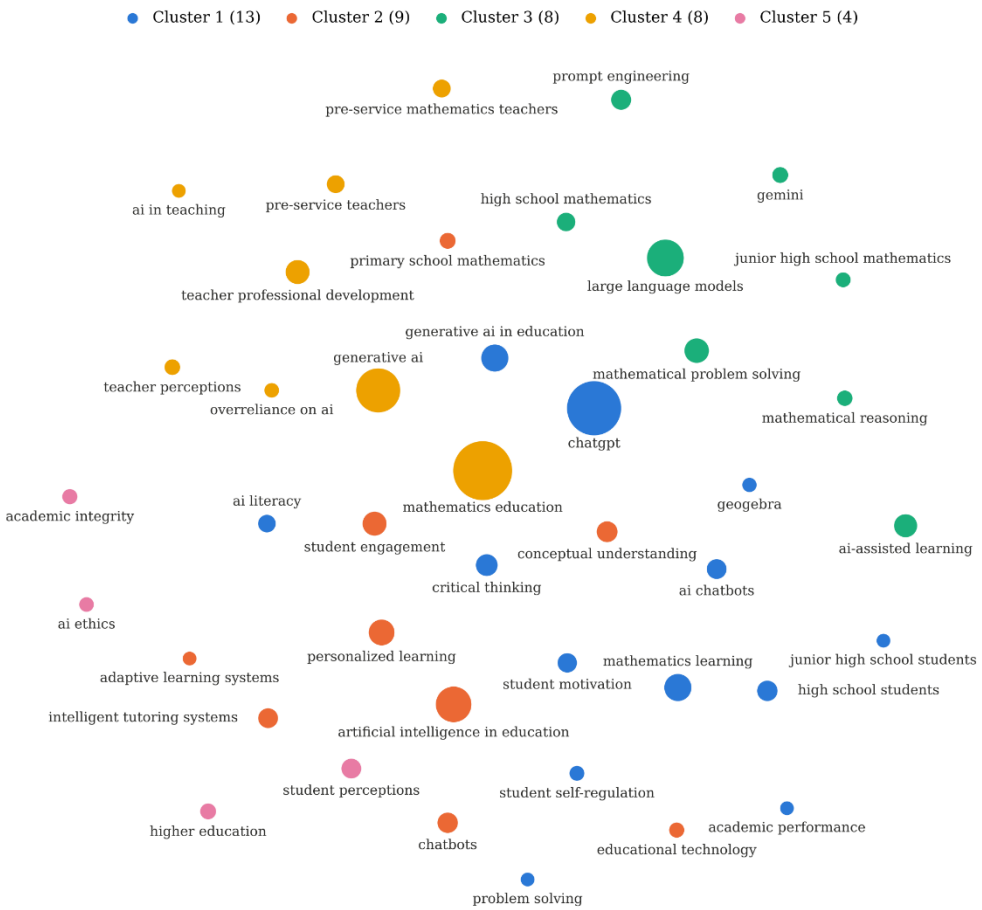


Figure 2. Keyword Co-occurrence Map
(node size reflects occurrences; color indicates cluster)

The overlay visualization showed that the newest keywords, measured by the average publication year of the documents containing them, were *high school students* (2025.65), *academic integrity* (2025.64), *AI literacy* (2025.63), *generative AI* (2025.57), and *student self-regulation* (2025.54). The oldest were *educational technology* (2024.84), *problem solving* (2024.85), and *personalized learning* (2024.94) (Figure 3). In terms of average normalized citations, the keywords with the highest relative impact were *overreliance on AI* (3.32), *pre-service mathematics teachers* (2.61), *teacher professional development* (2.18), and *prompt engineering* (2.13).

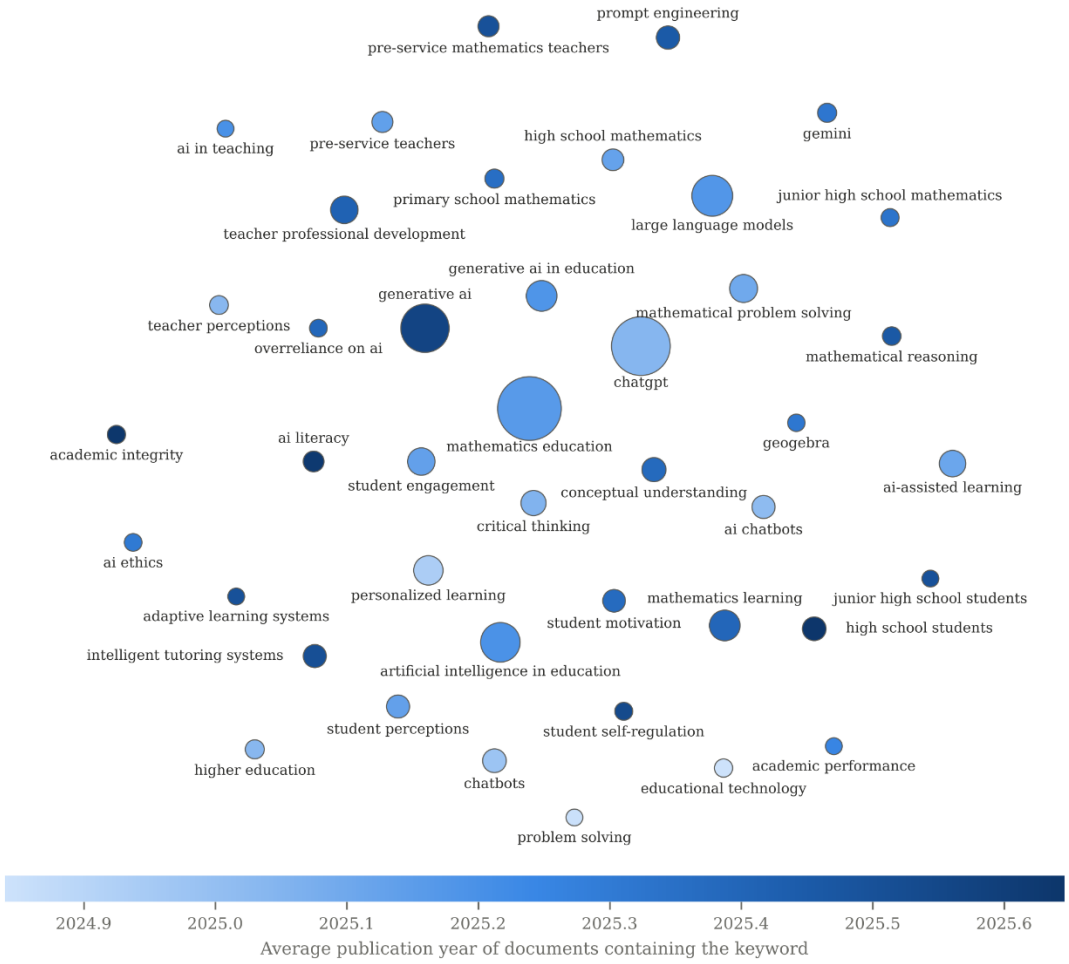


Figure 3. Overlay Visualization by Average Publication Year

3. Uses, Effects, and Challenges in Key Studies

The key studies showed two groups of uses, namely uses by students and uses by teachers (Table 7). For students, the dominant role was that of a tutor that explains concepts and gives stepwise feedback. ChatGPT-generated hints produced learning gains in algebra and statistics equivalent to human-authored hints, although the help failed quality checks on 32% of problems (Pardos & Bhandari, 2024). In primary schools, a GenAI-enabled learning environment improved word problem solving more than conventional instruction (Liu et al., 2025), and ChatGPT-generated worksheets supported two eighth-grade students with special educational needs, who became more engaged (Rizos et al., 2024). For teachers, GenAI served mainly as an assistant for lesson planning, task generation, and assessment design (Egara & Mosimege, 2024; Taani & Alabidi, 2024). Awareness remained limited in some contexts: in one Nigerian sample, only 17% of mathematics teachers were familiar with ChatGPT (Egara & Mosimege, 2024).

Table 7. Uses of GenAI in Mathematics Learning Reported in Key Studies

User	Use	Example studies
Students	Tutor and concept explainer	Bastani et al. (2025); Pardos & Bhandari (2024)
Students	Feedback and problem-solving scaffold	Liu et al. (2025); Schorcht et al. (2024)
Students	Creative writing and discussion aid	Song et al. (2025); Vasconcelos & dos Santos (2023)
Students	Support for special educational needs	Rizos et al. (2024)
Students	Self-study instrument	Egara et al. (2025); Kock et al. (2025)
Teachers	Lesson planning and task generation	Egara & Mosimege (2024); Motseki & Jojo (2025); Taani & Alabidi (2024)
Pre-service teachers	Problem solving and lesson design	Getenet (2024); Mosia et al. (2026)

The effects reported in key studies varied with the conditions of use (Table 8). Positive effects appeared when GenAI was embedded in structured scaffolding or a designed learning environment (Liu et al., 2025; Pardos & Bhandari, 2024), whereas adding ChatGPT to an existing instructional model did not improve outcomes significantly (Dianta & Rizkianto, 2025). Two studies reported unfavorable outcomes. Students who relied solely on ChatGPT performed lower than those taught by teachers, either alone or supported by ChatGPT (Dasari et al., 2024). In a large field experiment in high school mathematics, students with unrestricted access to GPT-4 during practice performed worse on later unassisted tests, whereas a tutor version with pedagogical safeguards largely mitigated this decline (Bastani et al., 2025).

Table 8. Direction of Effects Reported in Key Studies

Study	Context	Condition	Effect
Pardos & Bhandari (2024)	Algebra and statistics, N = 274	ChatGPT hints vs. human hints vs. no help	Positive, equal to human tutor
Liu et al. (2025)	Primary word problems	Designed GenAI learning environment	Positive
Dianta & Rizkianto (2025)	Upper secondary, Indonesia	ChatGPT added to problem-based learning	Not significant
Dasari et al. (2024)	Three learning groups	ChatGPT alone vs. teacher-led	Lower without teacher
Bastani et al. (2025)	High school mathematics	GPT-4 without vs. with guardrails	Negative without guardrails; mitigated with guardrails

Five challenges recurred across the key studies. First, mathematical accuracy remained limited: ChatGPT lacked a deep understanding of geometry and could not correct misconceptions effectively (Wardat et al., 2023), performed well below graduate level in advanced mathematics (Frieder et al., 2023), and solved routine word problems superficially (Strohmaier et al., 2025). Second, over-reliance on AI dialogue systems was associated with weaker critical thinking (Pepin et al., 2025; Zhai et al., 2024). Third, assessment integrity was threatened (Rudolph et al., 2023). Fourth, teacher readiness and AI literacy remained low (Egara & Mosimege, 2024; Mosia et al., 2026). Fifth, ethical issues, privacy, and unequal access to devices and the internet were reported (Dwivedi et al., 2023; UNESCO, 2023).

DISCUSSION

The bibliometric results show a field that is young, fast-growing, and dominated by a single tool. Almost all records appeared from 2024 onward, and ChatGPT was mentioned in nearly half of them, which indicates that research has followed the public release of specific products rather than a stable research agenda. The low share of records addressing primary and secondary schools (29.5%) reveals that the school context, where most mathematics learning takes place, remains underrepresented relative to higher education and teacher preparation. The narrative review adds that the effect of GenAI is conditional on instructional design. Scaffolding is effective when support is gradually withdrawn so that learners take over the task (Wood et al., 1976), and ready-made answers can bypass the cognitive processing needed for learning (Kirschner et al., 2006). GenAI that delivers complete solutions removes this productive effort, whereas GenAI that asks questions and gives hints preserves it, which explains why guardrails reversed the harm observed by Bastani et al. (2025).

The dominance of ChatGPT and the concentration of research in university settings are consistent with earlier reviews of AI and ChatGPT in mathematics education and in education generally (Deng et al., 2025; Heung & Chiu, 2025; Muhamad Sani, 2026; Opesemowo & Adewuyi, 2024). The present map extends these reviews by showing that teacher-related topics carry the highest relative citation impact: *overreliance on AI*, *pre-service mathematics teachers*, and *teacher professional development* all had normalized citation scores above two. This pattern supports the view that teacher competence, framed as TPACK, is central to meaningful integration (Mishra et al., 2023). The emergence of AI literacy, academic

integrity, and self-regulation as the newest topics mirrors the shift in the broader literature from enthusiasm about capabilities toward concerns about responsible use and learner agency (Rudolph et al., 2023; UNESCO, 2023; Zimmerman, 2002). The conditional effects found in key studies are also in line with meta-analyses showing larger gains in problem-based and personalized settings (Wang & Fan, 2025; Wu & Yu, 2024), and with critiques that many efficacy studies confound the tool with the teaching method (Weidlich et al., 2025).

Theoretically, this study extends the concept of scaffolding to AI-mediated mathematics learning by highlighting that the pedagogical behavior of the tool, namely whether it withholds or delivers solutions, matters more than its availability. Methodologically, it shows that open bibliographic sources such as OpenAlex and Garuda can be combined to map a fast-moving field that includes national publications. Practically, three implications follow for Indonesia. Teacher education and professional development should focus on task design and pedagogical prompting rather than on operating the application alone. Schools should regulate GenAI use in assessment and shift toward process-based and oral formats (Rudolph et al., 2023). Funding and research agendas should prioritize experimental studies in primary and secondary classrooms, where evidence remains thin.

This study has several limitations. The corpus was not screened at the study level, so it includes records outside the core topic, such as reviews, conference papers, and studies from other disciplines. The text-mining categories indicate mentions rather than verified study characteristics, and one-fifth of the records lacked abstracts. Citation and keyword analyses were limited to OpenAlex records, citation counts for recent publications are still accumulating, and the 2026 figures cover only part of the year. In addition, the key studies in the narrative review were selected purposively, which may introduce selection bias. Future studies are encouraged to conduct a systematic review with independent screening of school-level empirical studies, to include subscription databases, and to update the map as newer models and more experimental studies become available.

CONCLUSION

This study shows that research on GenAI in mathematics learning grew rapidly after 2023, with 95.2% of the 1,206 records published from 2024 onward. The field has been dominated by ChatGPT and by higher education and teacher preparation contexts, while

only 29.5% of the records addressed primary or secondary schools. The keyword map revealed five thematic clusters, with AI literacy, academic integrity, and self-regulation as the most recent topics and teacher-related topics as the most influential. Key studies indicate that GenAI benefits mathematics learning when it is embedded in scaffolded instruction with pedagogical guardrails, but can harm learning when used as an unrestricted answer source.

The study contributes an up-to-date map of the field that includes Indonesian publications and links this map to the conditions under which GenAI supports learning. Future research should conduct experimental and design-based studies at the primary and lower-secondary levels in Indonesia, develop Indonesian-language GenAI tutors with pedagogical guardrails, examine the long-term effects of GenAI on reasoning and retention, and conduct a systematic review and meta-analysis once sufficient school-level experimental studies are available.

REFERENCES

Almarashdi, H. S., Jarrah, A. M., Abu Khurma, O., & Gningue, S. M. (2024). Unveiling the potential: A systematic review of ChatGPT in transforming mathematics teaching and learning. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(12), Article em2555. https://doi.org/10.29333/ejmste/15739

Azaria, A., Azoulay, R., & Reches, S. (2024). ChatGPT is a remarkable tool—For experts. *Data Intelligence, 6*(1), 240–296. https://doi.org/10.1162/dint_a_00235

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, O., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

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

Chauncey, S. A., & McKenna, H. P. (2023). A framework and exemplars for ethical and responsible use of AI chatbot technology to support teaching and learning. *Computers and Education: Artificial Intelligence, 5*, Article 100182. https://doi.org/10.1016/j.caeai.2023.100182

Dasari, D., Hendriyanto, A., Sahara, S., Suryadi, D., Muhaimin, L. H., Chao, T., & Fitriana, L. (2024). ChatGPT in didactical tetrahedron, does it make an exception? A case study in mathematics teaching and learning. *Frontiers in Education, 8*, Article 1295413. https://doi.org/10.3389/feduc.2023.1295413

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227*, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224

Dianta, R. F., & Rizkianto, I. (2025). Pengaruh Penggunaan ChatGPT dengan Model Problem Based Learning dalam Pembelajaran Matematika terhadap Hasil Belajar Peserta Didik Sekolah Menengah Atas. *Jurnal Pedagogi Matematika, 11*(2), 149–156. https://journal.student.uny.ac.id/jpm/article/view/22104

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

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Egara, F. O., & Mosimege, M. (2024). Exploring the integration of artificial intelligence-based ChatGPT into mathematics instruction: Perceptions, challenges, and implications for educators. *Education Sciences, 14*(7), Article 742. https://doi.org/10.3390/educsci14070742

Egara, F. O., Mosimege, M., & Mosia, M. (2025). Secondary school students' perceptions of their usage of artificial intelligence-based ChatGPT in mathematics learning. *Journal of Education*, (98), 124–146. https://doi.org/10.17159/2520-9868/i98a07

Frieder, S., Pinchetti, L., Chevalier, A., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., & Berner, J. (2023). Mathematical capabilities of ChatGPT. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Advances in neural information processing systems* (Vol. 36, pp. 27699–27744). Curran Associates. https://doi.org/10.52202/075280-1205

Getenet, S. (2024). Pre-service teachers and ChatGPT in multistrategy problem-solving: Implications for mathematics teaching in primary schools. *International Electronic Journal of Mathematics Education, 19*(1), Article em0766. https://doi.org/10.29333/iejme/14141

Heung, Y. M. E., & Chiu, T. K. F. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence, 8*, Article 100361. https://doi.org/10.1016/j.caeai.2025.100361

Hu, B., Zheng, L., Zhu, J., Ding, L., Wang, Y., & Gu, X. (2024). Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design. *IEEE Transactions on Learning Technologies, 17*, 1445–1459. https://doi.org/10.1109/TLT.2024.3384765

Hwang, G.-J., & Tu, Y.-F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics, 9*(6), Article 584. https://doi.org/10.3390/math9060584

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kock, Z.-J., Salinas-Hernández, U., & Pepin, B. (2025). Engineering students' initial use schemes of ChatGPT as an instrument for learning. *Digital Experiences in Mathematics Education, 11*(1), 192–218. https://doi.org/10.1007/s40751-025-00169-w

Lee, D., & Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Computers & Education, 191*, Article 104646. https://doi.org/10.1016/j.compedu.2022.104646

Liu, J., Sun, D., Sun, J., Wang, J., & Yu, P. L. H. (2025). Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools: Learning performance, attitudes and interaction. *Computers and Education: Artificial Intelligence, 9*, Article 100438. https://doi.org/10.1016/j.caeai.2025.100438

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), Article 410. https://doi.org/10.3390/educsci13040410

Luzano, J. F. P. (2025). New frontier in mathematics education: A review of emerging trends and critical issues on artificial intelligence. *International Journal of Technology in Education, 8*(1), 208–219. https://doi.org/10.46328/ijte.1028

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education, 39*(4), 235–251. https://doi.org/10.1080/21532974.2023.2247480

Mohamed, M. Z. B., Hidayat, R., Suhaizi, N. N. B., Sabri, N. B. M., Mahmud, M. K. H. B., & Baharuddin, S. N. B. (2022). Artificial intelligence in mathematics education: A systematic literature review. *International Electronic Journal of Mathematics Education, 17*(3), Article em0694. https://doi.org/10.29333/iejme/12132

Mosia, M., Nannim, F. A., Tsakeni, M., Thamae, M. T., Egara, F. O., & Hamilton, J. (2026). Preservice mathematics teachers' AI adoption intentions: Examining the roles of awareness, trust, and risk. *Eurasia Journal of Mathematics, Science and Technology Education, 22*(9), Article em2912. https://doi.org/10.29333/ejmste/19241

Motseki, P., & Jojo, Z. (2025). Teachers' perceptions on the use of ChatGPT in teaching grade 12 mathematics. *Journal of Education*, (99), 123–141. https://doi.org/10.17159/2520-9868/i99a07

Muhamad Sani, N. S. A., Omar, M. A. A., & Abdullah, M. R. (2026). Exploring the role of generative artificial intelligence in transforming mathematics teaching and learning:

A systematic mapping review. *Innovative Teaching and Learning Journal, 10*(2), 155–174. https://doi.org/10.11113/itlj.v10.256

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Opesemowo, O. A. G., & Adewuyi, H. O. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(7), Article em2478. https://doi.org/10.29333/ejmste/14762

Pardos, Z. A., & Bhandari, S. (2024). ChatGPT-generated help produces learning gains equivalent to human tutor-authored help on mathematics skills. *PLOS ONE, 19*(5), Article e0304013. https://doi.org/10.1371/journal.pone.0304013

Pepin, B., Buchholtz, N., & Salinas-Hernández, U. (2025). A scoping survey of ChatGPT in mathematics education. *Digital Experiences in Mathematics Education, 11*(1), 9–41. https://doi.org/10.1007/s40751-025-00172-1

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts* (arXiv:2205.01833) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2205.01833

Rifandi, M., Muzakir, N. A., & Mayapada, R. (2025). Pelatihan Penggunaan ChatGPT sebagai Media Pembelajaran Matematika untuk Siswa MAN 1 Majene, Sulawesi Barat. *Jurnal Anugerah, 7*(1), 75–86. https://ojs.umrah.ac.id/index.php/anugerah/article/view/7111

Rizos, I., Foykas, E., & Georgakopoulos, S. V. (2024). Enhancing mathematics education for students with special educational needs through generative AI: A case study in Greece. *Contemporary Educational Technology, 16*(4), Article ep535. https://doi.org/10.30935/cedtech/15487

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

Sánchez-Ruiz, L. M., Moll-López, S., Nuñez-Pérez, A., Moraño-Fernández, J. A., & Vega-Fleitas, E. (2023). ChatGPT challenges blended learning methodologies in engineering education: A case study in mathematics. *Applied Sciences, 13*(10), Article 6039. https://doi.org/10.3390/app13106039

Schorcht, S., Buchholtz, N., & Baumanns, L. (2024). Prompt the problem – Investigating the mathematics educational quality of AI-supported problem solving by comparing prompt techniques. *Frontiers in Education, 9*, Article 1386075. https://doi.org/10.3389/feduc.2024.1386075

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Song, Y., Kim, J., Xing, W., Liu, Z., Li, C., & Oh, H. (2026). Elementary school students' and teachers' perceptions toward creative mathematical writing with generative AI. *Journal of Research on Technology in Education, 58*(3), 615–637. https://doi.org/10.1080/15391523.2025.2455057

Strohmaier, A. R., Van Dooren, W., Seßler, K., Greer, B., & Verschaffel, L. (2025). *Large language models don't make sense of word problems. A scoping review from a mathematics education perspective* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2506.24006

Suprihatiningsih, S., Uripno, G., Sujadi, I., & Rangkuti, R. K. (2026). Emerging artificial intelligence in mathematics education: A systematic literature review. *Journal of Learning for Development, 13*(2), 298–313. https://doi.org/10.56059/jl4d.v13i2.2200

Supriyadi, E. (2022). Eksplorasi Penggunaan ChatGPT dalam Penulisan Artikel Pendidikan Matematika. *Papanda Journal of Mathematics and Science Research, 1*(2), 54–68. https://doi.org/10.56916/pjmsr.v1i2.252

Taani, O., & Alabidi, S. (2025). ChatGPT in education: Benefits and challenges of ChatGPT for mathematics and science teaching practices. *International Journal of Mathematical Education in Science and Technology, 56*(9), 1748–1777. https://doi.org/10.1080/0020739X.2024.2357341

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*, Article 15. https://doi.org/10.1186/s40561-023-00237-x

Turmuzi, M., Azmi, S., & Kertiyani, N. M. I. (2026). ChatGPT in school mathematics education: A systematic review of opportunities, challenges, and pedagogical implications. *Teaching and Teacher Education, 170*, Article 105286. https://doi.org/10.1016/j.tate.2025.105286

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Vasconcelos, M. A. R., & dos Santos, R. P. (2023). Enhancing STEM learning with ChatGPT and Bing Chat as objects to think with: A case study. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(7), Article em2296. https://doi.org/10.29333/ejmste/13313

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis [Retracted]. *Humanities and Social Sciences Communications, 12*, Article 621. https://doi.org/10.1057/s41599-025-04787-y

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(7), Article em2286. https://doi.org/10.29333/ejmste/13272

Weidlich, J., Gašević, D., Drachsler, H., & Kirschner, P. (2025). ChatGPT in education: An effect in search of a cause. *Journal of Computer Assisted Learning, 41*(5), Article e70105. https://doi.org/10.1111/jcal.70105

Wijaya, T. T., Yu, Q., Cao, Y., He, Y., & Leung, F. K. S. (2024). Latent profile analysis of AI literacy and trust in mathematics teachers and their relations with AI dependency and 21st-century skills. *Behavioral Sciences, 14*(11), Article 1008. https://doi.org/10.3390/bs14111008

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology, 55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research, 39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

Yilmaz, H., Maxutov, S., Baitekov, A., & Balta, N. (2023). Student attitudes towards Chat GPT: A technology acceptance model survey. *International Educational Review, 1*(1), 57–83. https://doi.org/10.58693/ier.114

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments, 11*, Article 28. https://doi.org/10.1186/s40561-024-00316-7

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