EUROPEAN JOURNAL OF STEM EDUCATION

European Journal of STEM Education,
2026, 11(1), Article 43
eISSN: 2468-4368

lectito

Research paper

Artificial Intelligence in Mathematics Education: A PRISMA-Based Systematic Literature Review (2021-2025)

Fariza N. Omirzakova *, Sarsenkul Sh. Tleubai

Korkyt Ata Kyzylorda University, KAZAKHSTAN

*__Corresponding Author:__ omirzakovafariza@korkyt.kz

This paper is part of the special issue: AI and STEM: Exploring the Impact of Artificial Intelligence on Science, Technology, Engineering, and Mathematics Education

Citation: Omirzakova, F. N., & Tleubai, S. S. (2026). Artificial intelligence in mathematics education: A PRISMA-based systematic literature review (2021-2025). *European Journal of STEM Education, 11*(1), Article 43. https://doi.org/10.20897/ejsteme/19221

Published: September 5, 2026

ABSTRACT

This systematic literature review examines research on artificial intelligence (AI) in mathematics education published between January 1, 2021, and August 1, 2025. Searches of Scopus and Google Scholar identified 922 records; after deduplication, screening, and full-text eligibility assessment, 42 peer-reviewed journal articles and conference proceedings were included. The review used descriptive quantitative summaries and a deductive-inductive thematic synthesis. Two independent reviewers conducted screening (Cohen's kappa = 0.88), and methodological quality was assessed using the Mixed Methods Appraisal Tool (MMAT). The included literature indicates increasing attention to generative AI, personalised support, feedback, teacher practice, academic integrity, and equity. Evidence for educational benefits varies substantially across study designs and contexts, and technical capability should not be equated with demonstrated classroom effectiveness. Geographic patterns in the selected sample are described without attributing them to regulatory, economic, or infrastructural causes that were not directly tested. Because the 2025 search covered only January through August 2025, publication counts are treated as partial-year data and are not directly comparable to complete prior years. Key limitations include reliance on two databases, English- and Russian-language restrictions, reproducibility constraints in Google Scholar, methodological heterogeneity, and limited long-term evidence. Overall, AI shows potential to support mathematics teaching and learning, but stronger longitudinal and comparative evidence is needed to establish effectiveness, equity, and sustainable implementation.

Keywords: artificial intelligence, mathematics education, systematic review, personalised learning

Introduction without heading introduction without heading.

Artificial intelligence (AI) has become increasingly prominent in mathematics education research, with studies examining intelligent tutoring systems, generative AI, automated assessment, learning analytics, and teacher-facing tools. Existing reviews have documented geographical and methodological patterns (Hwang & Tu, 2021), emerging forms of AI integration (Nguyen & Pham, 2025), current practices (Awang et al., 2025), and broader trends in mathematics education (Mohamed et al., 2022). These studies provide an important foundation, but ongoing developments in generative AI and classroom adoption warrant an updated synthesis that distinguishes technological capability from educational evidence and accounts for methodological quality when interpreting reported outcomes. More broadly, recent AI-in-education research identifies personalised learning, intelligent tutoring systems and chatbots, and automated assessment as key areas of educational AI development (Acar et al., 2025; Lee et al., 2025).

Evidence on the educational effects of AI is promising yet heterogeneous. A meta-analysis by Yi et al. (2025) reported positive effects of AI-based tools on K-12 mathematics learning, while Gusti and Amastini (2025) discussed the pedagogical potential of deep-learning approaches. However, the strength and interpretation of reported benefits depend on study design, learner group, intervention duration, and outcome measures. Teachers' attitudes and pedagogical readiness are also important for the classroom integration of tools such as ChatGPT (Uğraş et al., 2024). Other reviews identify technical and infrastructural barriers (D. Li et al., 2025; Sutomo & Turmudi, 2025), while Panqueban and Huincahue (2024) discuss academic-integrity and digital-equity concerns. AI-supported assessment likewise raises questions about validity, reliability, and fairness (Zacharis & Papadakis, 2025).

These scholars warn that regions with limited access to technology and rural schools risk falling behind in the global educational landscape, and that students' capacity for independent thinking may be diminished.

Accordingly, AI in mathematics education offers both potential benefits and significant constraints. Some studies report improved performance, feedback, or learning support, whereas others identify technical limitations, risks to academic integrity, and unequal access. These findings should be interpreted in light of the design and evidential purpose of the underlying studies rather than generalised to AI technologies as a whole.

It is precisely at this point that a key scholarly contradiction arises: the mechanisms for balancing AI's cognitive benefits against its socio-ethical shortcomings remain poorly organised. Most existing studies focus either on technological effectiveness or on ethical constraints, but not both. However, for developing countries and regions affected by digital divides, there is a noticeable lack of integrated methodological models that reconcile technological innovation with traditional mathematical didactics without exacerbating social inequality. This study aims to bridge the gap between "opportunity and risk".

Research aim: To systematically review scholarly evidence published between January 2021 and August 1, 2025, on the use of AI in mathematics education, with particular attention to technological developments, pedagogical opportunities and limitations, ethical concerns, and contextual differences.

Research questions:

1. **RQ1**: Which stages define the development of artificial intelligence technologies in mathematics education at school and higher education levels, and how do cross-national differences and subject-specific application patterns emerge in this process?
2. **RQ2:** What opportunities and limitations are associated with using AI technologies in mathematics education, and how are ethical issues, academic integrity, and social inequality addressed in recent research, including proposed solutions to these concerns?

This review synthesises recent evidence on AI in mathematics education and identifies areas in which further empirical research is needed to support effective and ethically responsible implementation.

METHODOLOGY

The review process followed the PRISMA reporting framework (Page et al., 2021). Its principles informed the organisation of the search and selection workflow, from record identification and eligibility assessment to data extraction and evidence synthesis. This procedure was applied consistently across both research questions to document the construction of the final evidence base.

The retrieved evidence was summarised quantitatively using frequencies and percentages. The qualitative component combined deductive coding with inductive theme development, enabling predefined analytical dimensions to be examined while also capturing patterns emerging from the selected studies (Gough et al., 2017).

During extraction, information reported directly by the included studies was distinguished from review-level coding and interpretation. Descriptive characteristics were organised into categories established before analysis, whereas additional themes were developed from patterns identified across the evidence. Interpretations were included only when they could be traced to the reviewed material and were not presented as directly reported findings.

Data sources and search strategy

The study employed an advanced search strategy covering the period from January 1, 2021, to August 1, 2025, and encompassing a broad range of pedagogical and technological concepts. The extended search string used for Scopus was as follows:

TITLE-ABS-KEY (("artificial intelligence" OR "generative artificial intelligence" OR "generative AI" OR "machine learning" OR "intelligent tutoring system" OR "adaptive learning" OR "chatbot" OR "chatgpt" OR "large language model") AND ("mathematics education" OR "mathematical learning" OR "mathematics teaching" OR "numeracy" OR "algebra" OR "geometry" OR "calculus")) AND (LIMIT-TO (PUBYEAR , 2025) OR LIMIT-TO (PUBYEAR , 2024) OR LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021)).

For Google Scholar, given the platform's syntactic and ranking constraints, the following Boolean query was used. The search was conducted on 1 August 2025, using relevance-ranked results and the same 2021-August 1, 2025 date boundary applied to the review. Search results were screened against the same eligibility criteria used for Scopus. Because Google Scholar rankings are dynamic, the procedure is reported as a reproducibility limitation rather than assumed to be fully replicable:

("artificial intelligence" OR "generative AI" OR "machine learning" OR "intelligent tutoring system" OR "chatgpt") AND ("mathematics education" OR "mathematics teaching" OR "mathematics learning").

The search terms were selected to cover major AI concepts (artificial intelligence, generative AI, machine learning, intelligent tutoring systems, adaptive learning, chatbots, ChatGPT, and large language models) and mathematics-education concepts (mathematics education, mathematical learning, mathematics teaching, numeracy, algebra, geometry, and calculus). Inclusion was restricted to peer-reviewed scholarly publications directly related to mathematics teaching, learning, assessment, teacher education, or the educational use of AI. Non-peer-reviewed grey literature was excluded.

Following the initial database queries, pre-screening was applied to exclude irrelevant disciplinary fields (such as chemistry, engineering, pharmacology, immunology, and agriculture) and to restrict publication language to English and Russian. The initial search and database filtering identified 922 records (594 from Scopus and 328 from Google Scholar). After removing duplicates (n = 228), 694 unique records proceeded to screening (Scopus: 415, Google Scholar: 279).

Study selection

In the initial stage of the analysis, unique publications were evaluated against the predefined selection criteria in **Table 1.**

Table 1

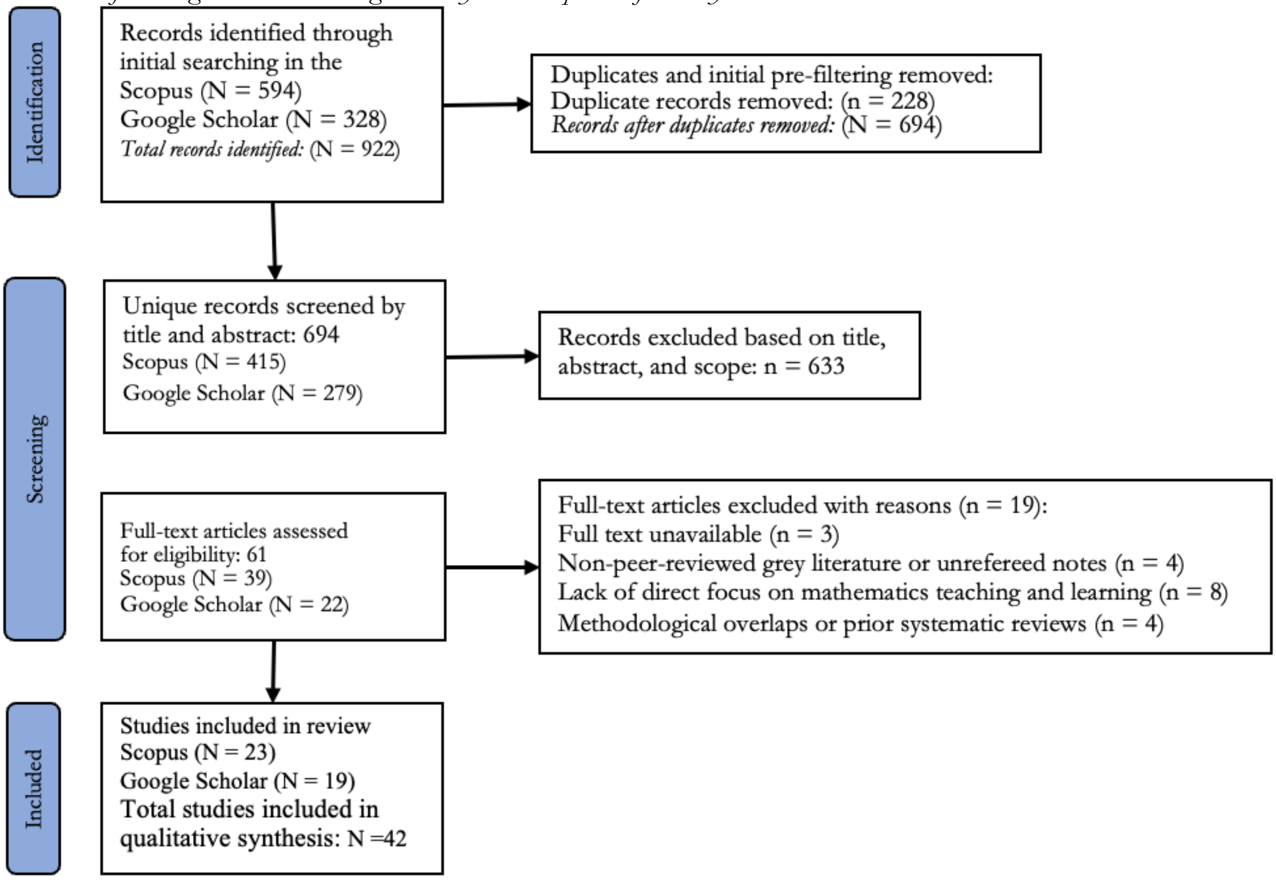
Inclusion and exclusion criteria

Inclusion	Exclusion
Studies describing mathematics teaching practices at school and higher education levels.	Studies examining AI applications in subjects other than mathematics or without analyzing mathematical aspects.
Studies addressing the adoption of AI technologies and their use in educational processes.	Studies unrelated to educational contexts or focusing exclusively on the technical parameters (algorithms) of AI.
Analyses of the impact of AI on educational quality, learner motivation, competencies, methodological models, or didactic applications.	Duplicate records identified across databases.
Empirical (quantitative and/or qualitative), theoretical, or mixed-methods studies.	Non-peer-reviewed blog posts, promotional materials, and low-credibility sources.

Both reviewers (FO and ST) independently evaluated all 694 unique records against the predefined eligibility criteria. After title and abstract screening, 633 records were excluded for not meeting the mathematics-education scope, leaving 61 for full-text evaluation. At this stage, three reports were unavailable in full text, four were non-peer-reviewed or preprints, eight lacked a direct focus on mathematics teaching or learning, and four were methodologically overlapping or previous systematic reviews. Reviewer discrepancies were discussed until agreement was reached. Inter-rater agreement was high (Cohen’s kappa = 0.88). The resulting evidence base comprised 42 publications, with 23 retrieved from Scopus and 19 from Google Scholar (**Figure 1**).

Figure 1

PRISMA flow diagram demonstrating the study selection process for the systematic review



Analysis of selected studies

The 42 included publications were coded as S1–S23 for Scopus records and G24–G42 for Google Scholar records, with the complete study inventory presented in Appendix A. Analysis proceeded in two complementary stages. First, predefined categories were used to map publication year, geographic setting, educational context, academic level, disciplinary area, and forms of AI application. Second, patterns emerging across the studies were examined inductively to identify recurring pedagogical approaches, limitations, and methodological responses. The resulting themes were then interpreted in relation to the two research questions (Gough et al., 2017).

Methodological quality assessment and risk identification

Given the heterogeneity of the included studies, evidence was grouped by research design and evidential purpose rather than treated as equivalent. The analytical groups comprised: (1) intervention and empirical outcome studies; (2) perception and survey studies; (3) design-based and technical studies with a direct educational connection; and (4) conceptual studies. Prior systematic reviews were used for framing and reference checking but were not analysed as units in the final sample.

Methodological appraisal was conducted using the Mixed Methods Appraisal Tool (MMAT 2018) for studies to which its design-specific criteria applied. Two reviewers independently recorded criterion-level judgements as Yes, No, or Cannot tell. Agreement between reviewers was high (Cohen's kappa = 0.88). Technical benchmarking, dataset development, and conceptual papers that did not fit an MMAT empirical design category were designated as Not applicable. Rather than producing a composite quality percentage, appraisal results were used to assess the strength of support each type of evidence provided for the review's conclusions. Detailed study-level decisions are reported in Appendix B.

The systematic data extraction, coding definitions, and synthesis rules used across these domains are detailed in **Table 2** below.

Table 2

Coding framework for data extraction and thematic synthesis

Domain	Coding unit	Definition / decision rule	Use in synthesis
Study characteristics	Whole study	Publication year, country/setting, educational level, mathematics domain, AI type, and study design as explicitly reported.	Descriptive mapping for RQ1.
Pedagogical opportunity	Finding or author-reported outcome	Evidence concerning learning support, feedback, adaptation, teacher support, assessment, or related educational use.	Thematic synthesis for RQ2; claims calibrated to study design.
Limitation / risk	Finding or author-reported concern	Evidence concerning accuracy, overreliance, academic integrity, privacy, fairness, access, or other constraints.	Thematic synthesis for RQ2.
Evidence type	Whole study	Intervention/outcome; perception/survey; qualitative; design/technical; conceptual.	Determines what type of inference the study can support.
Contextual interpretation	Review-level interpretation	Higher-level interpretation not directly extracted from a study. Causal wording is avoided unless supported by direct evidence.	Used cautiously in Discussion/Conclusion and identified as interpretation.

RESULTS AND DISCUSSION

This section presents the findings of the systematic review, organised by research question. To address RQ1, a deductive analytical approach was used, based on predefined categories derived from the literature (Gough et al., 2017). This method enabled systematic comparison of studies and helped identify recurring themes and patterns across various educational contexts. However, relying on predefined parameters may limit the analytical scope and lead to the overlooking of emerging aspects beyond the initial framework. To address this limitation, RQ2 was explored inductively, allowing for a more nuanced and flexible interpretation of the data.

RQ1: Which stages define the development of artificial intelligence technologies in mathematics education at school and higher education levels, and how do cross-national differences and subject-specific application patterns emerge in this process?

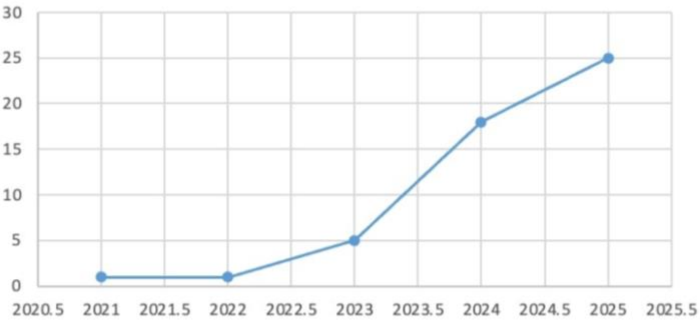
Temporal patterns in the included literature (2021-august 2025)

The small number of studies published in 2021-2022 reflects the composition of the final sample rather than an established historical stage of the field. Studies from these years covered intelligent tutoring systems and early AI-supported mathematics learning. Because the review was not designed as a longitudinal periodisation study, publication counts are used descriptively and are not treated as evidence of discrete developmental phases.

The number of included publications increased in 2023 and 2024, coinciding with growing attention to large language models and generative AI. Studies increasingly examined mathematical problem-solving, lesson planning, teacher use, feedback, and classroom implications. This pattern is described as a shift in emphasis within the selected literature rather than as a formally validated developmental phase (**Figure 2**).

Figure 2

Stages of AI research in mathematics education (2021–2025)



First, a comprehensive evaluation of capabilities was conducted, emphasising the strengths and limitations of large language models (LLMs) in mathematical modelling, geometric problem-solving, and university-level test tasks. Second, studies analysing pedagogical impact examined teachers' flexible use of LLMs in lesson planning and their influence on the adoption of new instructional methodologies. Consequently, the growth in the number of publications reflected an academic response to the increasing accessibility of AI technologies and the growing demand for new research.

2025: Partial-year evidence and emerging themes

The 2025 search covered publications available only through August 1, 2025. Consequently, the 2025 count reflects a partial publication year and is not directly comparable with the complete annual counts for 2021-2024. Within the studies retrieved during this period, recurring themes included methodological quality, academic integrity, teacher preparation, responsible design, and the educational use of generative AI. These themes indicate areas of current research attention but do not establish that the field has reached a stage of strategic stabilisation or maturity.

Firstly, ethical and safety concerns became central. The academic community has increasingly examined the potential negative impacts of AI tools on learning and the risks to academic integrity, emphasising the need to implement safeguards. Secondly, research on incorporating AI tools into teacher education and professional training programmes grew stronger, reflecting a shift towards recognising AI as an essential component of pedagogical skills rather than merely an auxiliary resource. Thirdly, responsible design topics gained importance, with research focusing not only on the development of intelligent tutoring systems (ITS) and chatbots but also on pedagogically effective and ethically sound design methods, including induction, scaffolding, clarification, and example-driven learning strategies. The included studies were geographically diverse, but their distribution reflects the review's search strategy, database coverage, language restrictions, and eligibility criteria (**Figure 3**)

Figure 3

Geographic distribution of publications by country

Geographic distribution of publications on the use of AI technologies in mathematics education

The included studies were geographically diverse, but their distribution reflects the review's search strategy, database coverage, language restrictions, and eligibility criteria. Accordingly, country and regional counts are treated as characteristics of the selected sample rather than as direct measures of overall research activity or of causal effects of economic, regulatory, or infrastructural conditions.

Asia (including Türkiye) accounted for 18 studies in the selected sample. The studies covered a range of topics, including teacher preparation, AI-supported learning, mathematical problem-solving, and educational access. Because the review did not systematically collect national economic or policy indicators, differences among countries are described without attributing publication patterns to economic or social causes.

Studies from South Korea and China in the selected sample examined student motivation, higher-order thinking, and the integration of AI-supported approaches in educational settings. These findings illustrate the thematic diversity of the regional sample but should not be taken as evidence of a uniform regional strategy.

North America, particularly the United States, was strongly represented in the selected sample [S1, S2, S6, S8, S9, S16, G31-G36, G38]. The studies covered intelligent tutoring systems, learning analytics, responsible AI design, and student outcomes. The review did not test whether publication volume was driven by technological leadership, financial investment, or funding policy; such explanations therefore fall outside the evidence generated by this review.

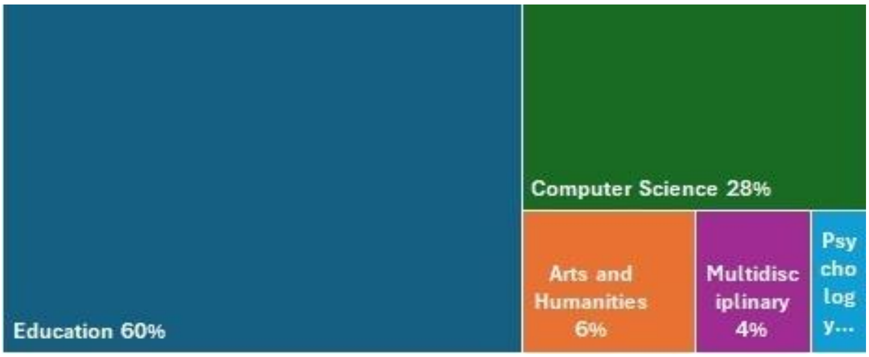
European institutions were underrepresented in the selected sample [S12, S14, S17, S22]. Some studies and contextual literature address responsible and ethical AI use, but the present review does not establish a causal relationship between European regulatory frameworks and the volume of publications.

Studies from African contexts were limited in the selected sample [S19, G28, G29]. They covered AI-supported mathematics learning, teacher awareness, and accessible tutoring models. Given the review's database and language restrictions, the small number of included studies should not be taken as a direct measure of research capacity or as evidence of a specific infrastructural cause.

South America was represented by a small number of studies in the selected sample. For example, the Argentine study [S15] examined the strengths and limitations of generative AI for geometric reasoning in Spanish. The limited representation should be interpreted cautiously, as the review was restricted to two databases and to English- and Russian-language publications. The distribution of studies across disciplinary fields is as follows: Education, 60%; Computer Science, 28%; Arts and Humanities, 6%; Psychology, 2%; and Multidisciplinary research, 4% (**Figure 4**).

Figure 4

Applications of artificial intelligence in mathematics education: In-subject application areas



The use of artificial intelligence in mathematics education: Intra-disciplinary application areas

The distribution of studies across disciplinary fields is as follows: Education, 60%; Computer Science, 28%; Arts and Humanities, 6%; Psychology, 2%; and Multidisciplinary research, 4%. These results highlight the disciplinary areas in which the application of artificial intelligence technologies to mathematics education has been most extensively explored.

Most publications focus on the field of Education [S1, S2, S4, S7- S9, S11, S12, S14, S16, S17-S19, S22, G24, G27-G29, G32-G35, G39, G40, G42]. These studies emphasise the pedagogical potential of AI technologies and confirm their active integration into educational systems. Specifically, research has examined AI's role in personalising learning and its impact on student motivation.

Furthermore, several studies have explored teachers' perceptions, experiences, and challenges in integrating AI into mathematics education, including the use of ChatGPT to develop motivational lessons. Other research has examined AI's impact on students' learning outcomes, creative writing, and problem-solving skills. Additionally, the role of AI-supported collaborative learning in mathematics teaching has been examined.

The Computer Science domain [S3, S5, S6, S13, S15, S21, S23, G25, G26, G31, G36] primarily explores the algorithmic and computational capabilities of AI systems. For instance, [S6] examines the problem-solving performance of ChatGPT and other AI tools from an algorithmic perspective. Additionally, [S6] evaluates the effectiveness of a fair AI model in predicting students' learning outcomes. Collectively, these studies investigate the technical and practical aspects of AI technologies and their implications for mathematics education. As a result, this research area focuses on the development, optimisation, and adaptation of AI systems for educational purposes.

The two studies within the Arts and Humanities domain [G30, G37] focus on the philosophical, social, and human-centred aspects of integrating AI into mathematics education. By examining users' perceptions of AI tools, trust levels, and practical use, these works emphasise the importance of the human element in adopting AI within educational systems. They also explore how AI-supported learning environments affect students' emotional well-being and attitudes towards the subject, including helping to reduce mathematics anxiety. Research in this area highlights the necessity of responsible, ethical, and reflective use of AI in education.

Only one study falls within the Psychology domain [G38]. This research describes an intervention experiment using AI technologies to reduce mathematics anxiety among primary school students. The scarcity of studies in this area indicates that the relationship between AI-mediated learning and psychological factors remains underexplored in the current literature.

Two studies are classified as Multidisciplinary [S10, S20]. [S10] provides evidence that unregulated generative AI may harm learning and raises concerns about ethical implementation and the preservation of educational quality. In contrast, [S20] offers a conceptual framework for improving mathematics teachers' creative pedagogical skills through the ChatGPT model.

Overall, 88% of the reviewed studies are situated at the intersection of education and computer science, highlighting artificial intelligence as a convergent force that integrates pedagogical and technological innovation in mathematics education. At the same time, the relatively limited representation of humanities and psychological research reveals a significant gap in the literature, particularly regarding the human-centred aspects of AI use, including ethics, trust, and emotional impact. This imbalance suggests that current research priorities remain largely solution-driven and technologically focused, while the broader social, ethical, and affective implications of AI-supported mathematics learning remain insufficiently explored.

RQ2: What opportunities and limitations are associated with using AI technologies in mathematics education, and how are ethical issues, academic integrity, and social inequality addressed in recent research, including proposed solutions to these concerns?

Personalisation emerged as a recurring educational application of AI across the reviewed literature. Structured tutoring systems and dialogue-based language models support this goal in different ways. RadarMath, MathGPT, and Flexi 2.0 [G24, S13], for example, use adaptive mechanisms to identify areas of difficulty and provide staged support. Their functions include diagnosing gaps in mathematical understanding, providing guided practice, and adjusting learning pathways in response to learners' needs.

LLMs, in turn, encompass general-purpose platforms such as ChatGPT, as well as specialised Math-LLM models, and function as flexible, dialogue-based cognitive support tools. ChatGPT effectively handles standard and intermediate-level mathematical problems, offering students clear solution pathways, while specialised Math-LLM models [G37] demonstrate greater precision in complex mathematical reasoning and explanation.

Furthermore, AI technologies support personalised learning by analysing students' prior errors, partially understood topics, and performance patterns [S9, S23]. Based on this analysis, optimal individual learning trajectories are proposed that match each learner's level of understanding [S9, S12, G25]. This involves dynamically adjusting task complexity: students are offered explanations and exercises suited to their current competency level, neither too simple nor too complex [S8, G30]. For example, if a student struggles to solve algebraic equations, the system does not immediately present more complex problems; instead, it provides additional tasks to reinforce foundational skills [G32, G36]. As a result, AI-assisted adaptation of content to individual needs helps create a deeper learning experience. Within such a supportive environment, students' attitudes towards learning improve [G38], making the learning process more effective. Research beyond mathematics education similarly suggests that AI-supported and multimodal technologies can extend personalised learning opportunities when their design incorporates appropriate feedback, teacher mediation, and equitable access (Martin, 2026).

AI tools also support cognitive interest in learning mathematics. Chatbots based on large language models, such as ChatGPT, are distinguished by their dialogue-based teaching functions. Using an interactive chatbot format [S22], these tools engage students through personalised questions and explanatory dialogue. This approach ensures constructive engagement in the learning process [G38], allowing students to acquire knowledge in a natural and comfortable communicative environment.

In addition, Intelligent Tutoring Systems employ their own support strategies. Systems such as RICE AlgebraBot [S1] are designed to provide consistent support throughout the learning process. Alongside correcting students' mistakes, these systems offer encouraging feedback that helps build learners' confidence in their abilities. Furthermore, the active integration of gamification elements is an important feature of AI-based educational tools. Presenting complex tasks in a game-like format makes it easier to master mathematical content. Thus, while LLMs primarily motivate learners through dialogue, ITS platforms provide support through task design and interactive structure.

One of the key advantages of artificial intelligence in the learning process is its rapid feedback mechanisms. In this regard, large language models (LLMs) demonstrate faster response times than Intelligent Tutoring Systems (ITS). However, this speed does not automatically translate into higher pedagogical quality. Students primarily use tools such as ChatGPT or specialised Math-LLM models [G37] to obtain quick answers to questions and to deepen their understanding of explanations. Nevertheless, the speed and convenience of these generative tools require careful monitoring of their pedagogical quality and methodological depth.

That said, certain limitations remain. For example, while mobile platforms such as Rori [S19] provide accessible feedback mechanisms, the use of limited WhatsApp-based interfaces can make it difficult to manage complex graphical or visual mathematical tasks. Overall, AI tools are not designed to replace teachers; rather, their primary function is to support teaching by providing timely, real-time feedback [S4, S6, S7, S17, S19, G33]. This feedback system enables the early detection of misunderstandings and errors.

The reviewed studies also describe teacher-facing uses of AI. Automating routine activities can free up time for direct instructional work [S5, S20], while generative and specialised systems can assist with lesson preparation and resource development [S3]. MATH41 [S3], for instance, supports the rapid production of mathematics tasks for learners at different levels.

In addition, AI provides teachers with data that supports timely pedagogical intervention. Hybrid AI models such as CognifyNet [S22] analyse students' activity patterns, enabling educators to detect emerging difficulties early. This illustrates AI's predictive role in learning management. AI algorithms also help forecast students' academic outcomes [G35] and automate assessment processes [S16]. Automated assessment not only saves time but also increases the objectivity of evaluation. Taken together, these functions demonstrate that AI technologies not only reduce administrative burden but also contribute to data-driven management of the learning process.

Limitations of using AI technologies in mathematics education

The evidence also highlights risks when AI support is inaccurate, poorly supervised, or over-relied on. Two recurring concerns are the erosion of independent mathematical activity and exposure to incorrect outputs. Heavy dependence on ready-made AI solutions may reduce opportunities for learners to practise problem-solving and critical evaluation [S15]. If students bypass the reasoning required to construct solutions themselves, their capacity to handle more demanding tasks may be affected, a concern reflected in reported learning-performance outcomes [S11].

Furthermore, AI systems, including those specifically designed for mathematics, are not immune to errors in tasks requiring high academic accuracy. When addressing complex mathematical problems or algorithmic challenges, AI tools may produce incorrect solutions or flawed strategies [S14], raising doubts about their reliability. Mistakes in geometric reasoning or in executing advanced algorithms, for example, can lead students away from correct conceptual understanding. Even specialised systems such as Math-LLMs can generate inaccuracies when constructing long chains of logic [S14].

Another concern is students' tendency to accept AI-generated outputs uncritically. Evidence shows that learners often pay little attention to verifying the accuracy of AI-provided solutions [S15], allowing misconceptions to take root in the learning process. Therefore, uncritical trust in AI-generated information is unjustified, and active oversight by educators remains essential to uphold academic accuracy and conceptual integrity. Furthermore, integrating artificial intelligence technologies into mathematics education raises a complex array of ethical, academic integrity, and social inequality concerns. These issues raise significant concerns about the fairness, trustworthiness, and legitimacy of AI-supported educational methods.

Privacy and fairness raise further concerns. AI-supported learning environments may process sensitive information about achievement, errors, learning trajectories, and, in some applications, affective states [G35]. The educational value of such data therefore needs to be weighed against questions of access, security, and responsible use. Fairness is equally important for predictive and assessment systems [G35, G39], because insufficient validation or biased training data can produce systematically different outcomes for learner groups defined by characteristics such as gender, ethnicity, or socioeconomic background. From a broader sociotechnical perspective, concerns about bias also extend to how gender stereotypes may be reproduced through the design and use of conversational AI technologies (Moratti et al., 2026; Sultan et al., 2025).

With the rapid development of generative AI, academic integrity has become a significant ethical issue. AI tools enable students to access ready-made answers or complete solution pathways for mathematical problems [S14, G40], encouraging academic misconduct that prioritises results over the learning process. This practice undermines the validity of assessments, as high grades obtained through AI-assisted work may not truly reflect students' knowledge or skills [S14]. Consequently, educators may struggle to accurately assess learners' genuine understanding when evaluating AI-generated or AI-supported assignments [S10]. These challenges weaken the credibility of assessment systems and, in turn, the integrity of the educational process. Therefore, recent research increasingly focuses on developing AI-detection tools and creating pedagogical strategies to reduce academic dishonesty while promoting meaningful learning.

Access to advanced AI tools is also uneven. The costs of high-quality tutoring systems and specialised Math-LLMs can place them beyond the reach of some schools and households. Evidence on the digital access gap [S21, G26] suggests that institutional and material resources determine who can benefit from AI-supported mathematics learning. Without measures to broaden access, technological innovation may reproduce rather than reduce existing educational inequalities. At the same time, emerging evidence indicates that well-designed AI-supported mathematics instruction may also contribute to more inclusive learning opportunities for students with specific accessibility needs (Asanre et al., 2026).

Additionally, recent studies not only examine the ethical, academic integrity, and social inequality challenges associated with implementing AI technologies but also propose practical strategies and solutions to address them. Additionally, recent studies not only examine the ethical, academic integrity, and social inequality challenges associated with implementing AI technologies but also propose practical strategies and solutions to address them (**Table 3**).

Table 3

Key issues in the implementation of AI Technologies

Key Issues	Proposed Solutions	Studies
Ethical issues: data privacy and fairness	Ethical and legal governance. This approach focuses on ensuring the fair, transparent, and secure operation of AI systems.	G35, G39
Academic integrity	Pedagogical redesign to support academic integrity. This involves modifying teaching practices and assessment systems to reduce the risk of academic misconduct associated with generative AI.	G40, G41
Social inequality	Reducing inequality and increasing accessibility. The benefits of AI technologies should be accessible to all students, particularly those from resource-constrained backgrounds.	S21, G26

To reduce the risks associated with AI technologies in mathematics education and to fully harness their potential, recent research recommends the following comprehensive strategic approaches:

Improving data protection policies.

Educational institutions should implement strict regulations aligned with international standards to ensure the secure storage and handling of personal and learning-related data collected by AI platforms [G35]. Students must be provided with clear, accessible information on how their data is collected, processed, and used.

Developing fair AI algorithms. To minimise the risk of bias in AI-driven predictive systems, developers should prioritise the development of fairness-oriented algorithms [G39]. This requires systematic testing to ensure that AI-generated decisions and recommendations remain equitable and objective across diverse student groups.

Reforming assessment practices. Assessment should move away from standardised tasks that AI can easily complete towards activities that demand critical thinking, higher-order analysis, and creativity. For example, students might be asked not only to provide answers but also to critically evaluate AI-generated solutions or to design and justify mathematical models [G40].

Enhancing teachers' professional competence. Ongoing professional development programmes are vital for equipping mathematics teachers with the skills to integrate AI technologies into teaching and to evaluate AI-supported student work [G41]. Such training supports informed pedagogical decisions and promotes the responsible use of AI in the classroom.

Funding AI tools. Governments and educational institutions should support the development and subsidisation of non-commercial, free, and high-quality ITS [S19] for schools. Such initiatives would directly help reduce the digital divide and promote fairer access to AI-enhanced learning opportunities.

Infrastructure provision

Investment in the technical infrastructure crucial for the effective use of AI-based learning platforms, including reliable internet connectivity and access to appropriate digital devices, is necessary [G26].

Evidence from resource-constrained rural educational settings further shows that socioeconomic and infrastructural barriers can substantially shape opportunities for technology-supported teaching and learning (Hasan & Naomee, 2026).

The review identified several key research gaps in assessing the overall impact of AI integration within mathematics education:

1. Lack of long-term impact assessment. There is a scarcity of studies examining how sustained use of AI tools affects students' mathematical reasoning and independent problem-solving skills over extended periods (e.g., 1–3 years). Existing research primarily examines the risk of AI dependency but provides no empirical evidence on how this dependency affects long-term learning outcomes.
2. Limited empirical validation of strategic interventions. Although alternative assessment methods, such as critical analysis tasks and oral examinations, are often proposed to address concerns about academic integrity, there is a lack of comparative and experimental studies assessing how effectively these strategies reduce AI-

assisted misconduct. While the current literature discusses what should be done, it offers limited evidence of their real-world effectiveness.

3. Social inequality and the absence of coordinated policy frameworks. Research into the impact of AI adoption on learning outcomes in low-resource and culturally diverse settings remains scarce. Specifically, the lack of longitudinal socio-economic data demonstrating how AI technologies might exacerbate educational inequalities hampers the development of effective global and local policies to promote equity in AI-supported education.

CONCLUSION

This systematic literature review examined 42 studies on AI in mathematics education published between January 2021 and 1 August 2025. The selected literature shows growing attention to generative AI, intelligent tutoring systems, teacher practice, feedback, assessment, academic integrity, and equity. Because the evidence base is methodologically heterogeneous, conclusions were calibrated to study design and evidential purpose.

Publication patterns suggest a shift in research emphasis over the review period, particularly following the wider availability of large language models. However, the review does not treat these patterns as validated developmental stages. The 2025 evidence is especially provisional because the search covered only January through August, representing a partial publication year.

Geographical differences were observed in the selected sample, with stronger representation from Asia and North America and fewer studies from Europe, Africa, and South America. These distributions should not be taken as evidence that regulation, investment, or infrastructure caused regional differences. Database coverage, language restrictions, and the review's eligibility criteria may also have shaped the observed pattern.

From a pedagogical perspective, the reviewed literature suggests that AI can support diagnostic feedback, adaptive learning activities, lesson planning, and certain forms of assessment. The strength of this evidence varies: intervention studies can inform claims about learning outcomes, whereas perception studies primarily describe attitudes and experiences, and technical studies primarily demonstrate system capabilities. AI-generated speed or functionality should therefore not be equated with instructional quality without direct comparative evidence.

Ethical and societal considerations become central rather than marginal. The review shows that concerns about data privacy, algorithmic fairness, academic integrity, and unequal access are interconnected aspects of AI integration, not isolated problems. Generative AI, in particular, challenges traditional assessment methods by blurring the line between learning support and academic misconduct, prompting a fundamental rethink of how mathematical competence is assessed. Notably, although many studies propose governance frameworks, assessment reforms, and professional development as solutions, robust empirical validation of these strategies remains scarce.

By synthesising technological, pedagogical, ethical, and policy-oriented strands of research, this review offers an integrated analytical perspective on AI in mathematics education. Rather than treating innovation and risk as separate domains, the findings conceptualise AI integration as a tensioned system in which educational value emerges only when pedagogical goals, ethical safeguards, and contextual constraints are addressed together. This perspective helps explain why technologically advanced solutions may yield uneven or even counterproductive results when implemented without appropriate instructional and institutional adaptation.

Several priorities follow from these findings. First, longitudinal evidence is needed to establish how sustained AI-supported instruction affects mathematical reasoning, learner independence, and conceptual development. Second, comparative and experimental studies should test whether proposed assessment redesigns and integrity-preserving practices are effective under authentic educational conditions. Third, research should include more low-resource and culturally diverse settings so that policies for AI-supported mathematics education are informed by evidence on both access and effectiveness.

Overall, the evidence reviewed supports a cautious view of AI in mathematics education. Educational benefit depends on how a technology is designed, used, supervised, and situated within a particular learning context; technical performance alone is insufficient evidence of instructional effectiveness. Future work should therefore strengthen longitudinal and comparative evaluation, examine learning outcomes transparently, support teacher preparation, test approaches to academic integrity, and address equitable access before broad claims of effectiveness or policy are made.

Limitations of the study

This review has several limitations that should be taken into account when interpreting its findings:

Database, search, and language limitations: The review was limited to Scopus and Google Scholar, and to publications in English and Russian. Relevant studies indexed elsewhere or published in other languages may therefore have been missed. Google Scholar uses dynamic ranking, which limits exact reproducibility even when the query, date, and screening procedure are documented.

Methodological heterogeneity and appraisal limitations: The included evidence spans intervention, survey, qualitative, design-based, technical, and conceptual research. These designs support different types of inference and were therefore not treated as equivalent. MMAT was used to inform cautious interpretation of applicable empirical studies; however, heterogeneity limits direct comparison across the full evidence base.

Regional and contextual limitations: Geographic distributions are based on studies retrieved from the review's database, subject to language and eligibility restrictions. The review did not collect sufficient causal evidence to attribute regional publication patterns to regulation, investment, infrastructure, or other macro-level factors.

Time-window and publication-bias limitations: The search ended on August 1, 2025, so 2025 is a partial publication year and cannot be directly compared with complete years. Restricting the review to peer-reviewed sources may also contribute to publication bias and exclude potentially relevant grey literature.

Acknowledgement

The authors gratefully acknowledge Korkyt Ata Kyzylorda University and the Chairman of the Board-Rector for supporting this research through the Rector's Grant. The authors also thank Gulzhan Karibaeva for her valuable support in preparing the manuscript.

Funding

This work was funded by the Grant of the Chairman of the Board-Rector of Korkyt Ata Kyzylorda University under the project "The Future of Mathematics: Methodology for the Application of Artificial Intelligence Technologies" (Contract No. 132-01-26, dated January 9, 2026).

Ethical statement

This systematic review analysed previously published literature and did not involve recruiting human participants, collecting new individual-level personal data, or conducting an intervention. Institutional ethics approval and informed consent were therefore not applicable to this review. No review protocol was prospectively registered.

Competing interests

The authors declare no competing interests in relation to this research, authorship, or publication.

Author contributions

FNO: conceptualisation, literature search, screening, data extraction, analysis, visualisation, and drafting the manuscript. SST: methodology, independent screening and full-text assessment, validation, interpretation, supervision, and critical revision of the manuscript. Both authors resolved screening disagreements by consensus, reviewed the final manuscript, and approved the version submitted for publication.

Data availability

The complete list of the 42 studies included in the review is provided in Appendix A, and the study-level methodological appraisal is presented in Appendix B. Additional review materials are available from the corresponding author upon reasonable request, subject to the records retained by the authors.

AI disclosure

Generative AI tools were used during manuscript revision to support language editing, refine wording, and organise the methodological-appraisal presentation. They were not used to determine study eligibility or to generate, fabricate, or alter study data. The authors verified the source information, reviewed all AI-assisted revisions, and take full responsibility for the accuracy, interpretation, and final content of the manuscript.

Biographical sketch

Fariza N. Omirzakova is a doctoral student in the Department of Mathematics Teacher Training at Korkyt Ata Kyzylorda University, Kazakhstan. She holds a Master of Technical Sciences. Her research focuses on mathematics education and the integration of artificial intelligence into teaching and learning. Her research interests include artificial intelligence in education, mathematics education, educational technologies, and the development and evaluation of AI-supported approaches to mathematics teaching.

Sarsenkul Sh. Tleubai is an Associate Professor at Korkyt Ata Kyzylorda University, Kazakhstan, and holds the academic title of Candidate of Pedagogical Sciences. Her academic work focuses on mathematics teacher education and pedagogy. Her research interests include mathematics teaching methodology, teacher education, innovative teaching approaches, educational technologies, and the pedagogically grounded integration of digital and artificial intelligence technologies into education.

Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and do not necessarily reflect the views of Lectito Publications and/or the editor(s). Lectito Publications and/or the editor(s) disclaim responsibility for any injury to persons or property resulting from any ideas, methods, instructions, or products referred to in the content.

REFERENCES

Acar, E., Yigit, F., & Deiri, Y. (2025). A focused review of artificial intelligence in education: Evolution and challenges. *Journal of Interdisciplinary Research in Artificial Intelligence and Society*, *1*(1), Article 3. https://doi.org/10.20897/jirais/17640

Akosah, E. F. (2025). AI-assisted collaborative learning in mathematics education: A qualitative approach. *International Journal of Technology in Education and Science*, *9*(3), 416–433. https://doi.org/10.46328/ijtes.634

Alhazzani, N. S. (2024). Enhancing mathematics teachers' pedagogical skills by using ChatGPT. *International Journal of Innovative Research and Scientific Studies*, *7*(4), 1614–1626. https://doi.org/10.53894/ijirss.v7i4.3460

Alvarez, J. I. (2024). Evaluating the impact of AI–powered tutors MathGPT and Flexi 2.0 in enhancing calculus learning. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi*, *8*(2), 495–508. https://doi.org/10.22437/jiituj.v8i2.34809

Asanre, A. A., Oguntola, J. I., Ramatea, M. A., & Lawani, A. O. (2026). Enhancing mathematics retention through seeing-AI: A quasi-experimental study among visually impaired junior secondary students in Ogun State, Nigeria. *European Journal of STEM Education*, *11*(1), Article 18. https://doi.org/10.20897/ejsteme/18268

Awang, L. A., Yusop, F. D., & Danaee, M. (2025). Current practices and future direction of artificial intelligence in mathematics education: A systematic review. *International Electronic Journal of Mathematics Education*, *20*(2), Article em0823. https://doi.org/10.29333/iejme/16006

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & 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

Cho, M. K., & Kim, S. (2025). Analyzing AI-based educational platforms for supporting personalized mathematics learning. *International Electronic Journal of Mathematics Education, 20*(4), em0847. https://doi.org/10.29333/iejme/16664

Dahal, N., Lamichhane, B. R., Luitel, B. C., & Pant, B. P. (2023). AI chatbots as math algorithm problem solvers: A critical evaluation of its capabilities and limitations. In *Proceedings of the 28th Asian Technology Conference in Mathematics (ATCM 2023).*

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

Gabriel, F., Kennedy, J., Marrone, R., & Leonard, S. (2025). Pragmatic AI in education and its role in mathematics learning and teaching. *npj Science of Learning*, *10*(1), Article 26. https://doi.org/10.1038/s41539-025-00315-4

Garcia, K. F., Ong, A. K. S., Gumasing, M. J. J., & Delos Reyes, C. R. V. (2025). Engineering students' perceptions and actual use of AI-based math tools for solving mathematical problems. *Acta Psychologica*, *256*, Article 105004. https://doi.org/10.1016/j.actpsy.2025.105004

Gusti, V. Y. K., & Amastini, F. (2025). Bridging pedagogy and AI: A systematic review of deep learning in mathematics education. *Jurnal Pendidikan Matematika (JPM)*, *11*(2), 141–156. https://doi.org/10.33474/jpm.v11i2.24319

Hasan, M. S., & Naomee, I. (2026). ICT teaching-learning practices and barriers in rural secondary schools of Bangladesh. *Asia Pacific Journal of Education and Society*, *14*(2), Article 4. https://doi.org/10.20897/apjes/19181

Henkel, O., Horne-Robinson, H., Kozhakhmetova, N., & Lee, A. (2024). Effective and scalable math support: Experimental evidence on the impact of an AI-math tutor in Ghana. In A. M. Olney et al. (Eds.), *Artificial intelligence in education: Posters and late breaking results* (Vol. 2150, pp. 373–381). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-64315-6_34

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

Inoferio, H. V., Espartero, M., Asiri, M., Damin, M., & Chavez, J. V. (2024). Coping with math anxiety and lack of confidence through AI-assisted learning. *Environment and Social Psychology*, *9*(5), Article 2228. https://doi.org/10.54517/esp.v9i5.2228

Inweregbuh, O. C., & Ugwuanyi, C. C. (2025). Mathematics teachers' self-efficacy in utilizing artificial intelligence (AI) tools as correlate of their professional development. *African Journal of Science, Technology, Mathematics and Education (AJSTME)*, *11*(1), 298-304.

Jancarik, A., Novotna, J., & Michal, J. (2022). Artificial intelligence assistant for mathematics education. *European Conference on E-Learning*, *21*(1), 143–148. https://doi.org/10.34190/ecel.21.1.783

Kanvaria, V. K., & Yadav, A. (2024). Artificial intelligence in the classroom: Experimental research on innovative approaches to mathematics instruction. *Revista Electronica de Veterinaria*, *25*(1), 708–716. https://doi.org/10.69980/redvet.v25i1.678

Kim, Y. R., Park, M. S., & Joung, E. (2025). Exploring the integration of artificial intelligence in math education: Preservice teachers' experiences and reflections on problem-posing activities with ChatGPT. *School Science and Mathematics*, *125*(3), 145–158. https://doi.org/10.1111/ssm.18336

Lawrence, L., Echeverria, V., Yang, K., Aleven, V., & Rummel, N. (2024). How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process. *British Journal of Educational Technology*, *55*(3), 823–844. https://doi.org/10.1111/bjet.13372

Lee, J., Wang, S., & Esposito, A. G. (2025). Content Analysis of College Student Perceptions: Mental Health-Related Support from Generative AI Versus Faculty Mentors. *American Journal of Qualitative Research, 9*(4), 70-100. https://doi.org/10.29333/ajqr/16815

Li, C., & Lyu, B. (2025). Investigating the motivational and knowledge affordances of conversational AI using induction, concretization and exemplification in math learning. *British Journal of Educational Technology*, *56*(5), 1814–1841. https://doi.org/10.1111/bjet.13612

Li, C., Xing, W., & Leite, W. (2024). Using fair AI to predict students' math learning outcomes in an online platform. *Interactive Learning Environments*, *32*(3), 1117–1136. https://doi.org/10.1080/10494820.2022.2115076

Li, C., Xing, W., Song, Y., & Lyu, B. (2025). RICE AlgebraBot: Lessons learned from designing and developing responsible conversational AI using induction, concretization, and exemplification to support algebra learning. *Computers and Education: Artificial Intelligence*, *8*, Article 100338. https://doi.org/10.1016/j.caeai.2024.100338

Li, D., Osman, S., Alhassora, N. S. A., Amantha Kumar, J., & Kartini Said Husain, S. (2025). Empowering K-12 mathematics teachers with artificial intelligence: A systematic review of insights, applications, and challenges. *IEEE Access*, *13*, 117198–117209. https://doi.org/10.1109/ACCESS.2025.3585784

Lu, Y., Pian, Y., Chen, P., Meng, Q., & Cao, Y. (2021). RadarMath: An intelligent tutoring system for math education. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(18), 16087–16090. https://doi.org/10.1609/aaai.v35i18.18020

Lu, Y., Supriya, K., Shaked, S., Simmons, E. H., & Kusenko, A. (2025). Incentivizing supplemental math assignments and using AI-generated hints is associated with improved exam performance. *Physical Review Physics Education Research*, *21*(1), Article 010160. https://doi.org/10.1103/PhysRevPhysEducRes.21.010160

Lyu, B., Li, C., Li, H., Oh, H., Song, Y., Zhu, W., & Xing, W. (2025). Exploring the role of teachable AI agents' personality traits in shaping student interaction and learning in mathematics education. *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 887–894. https://doi.org/10.1145/3706468.3706532

Martin, C. (2026). Multimodal vocabulary instruction for English language learners: From interactive whiteboards to artificial intelligence. *European Journal of Education & Language Review*, *2*(1), Article 4. https://doi.org/10.20897/ejelr/18945

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

Mondal, S., Khatua, D., Mandal, S., Prasad, D. K., & Sekh, A. A. (2025). BMWP: The first Bengali math word problems dataset for operation prediction and solving. *Discover Artificial Intelligence*, *5*(1), Article 25. https://doi.org/10.1007/s44163-025-00243-7

Moratti, S., Hellstrand, I., Bruijning, N., & Søraa, R. A. (2026). Gender imaginaries in sci-fi and real-life AI and the ongoing relevance of second-wave feminism. *Feminist Encounters: A Journal of Critical Studies in Culture and Politics*, *10*(2), Article 9. https://doi.org/10.20897/femenc/18844

Nguyen, D. T., & Pham, Q. V. (2025). The evolving landscape of AI integration in mathematics education: A systematic review of trends (2015–2025). *Eurasia Journal of Mathematics, Science and Technology Education*, *21*(10), Article em2714. https://doi.org/10.29333/ejmste/17078

Oh, S. (2025). Integration of MATH41 and generative AI in pre-service mathematics teacher education: An empirical study on lesson design competency. *IEEE Access*, *13*, 128959–128973. https://doi.org/10.1109/ACCESS.2025.3586593

Olarewaju, A. O., Moshood, A., & Awoyemi, S. A. (2025). Pre-service teachers' awareness and utilization of AI tools in mathematics education: Evidence from Kwara State, Nigeria. *Al-Hikmah Journal of Education*, *12*(1), 45–60.

Otero, N., Druga, S., & Lan, A. (2025). A benchmark for math misconceptions: Bridging gaps in middle school algebra with AI-supported instruction. *Discover Education*, *4*(1), Article 277. https://doi.org/10.1007/s44217-025-00742-w

Oğuz Haçat, S., & Kepceoğlu, İ. (2025). Help of ChatGPT about math skills on the subjects of social studies. *The Journal of Educational Research*, *118*(2), 1–12. https://doi.org/10.1080/00220671.2025.2510389

Panqueban, D., & Huincahue, J. (2024). Inteligencia artificial en educación matemática: Una revisión sistemática [Artificial intelligence in mathematics education: A systematic review]. *Uniciencia*, *38*(1), 1–17. https://doi.org/10.15359/ru.38-1.20

Parra, V., Sureda, P., Corica, A., Schiaffino, S., & Godoy, D. (2024). Can generative AI solve geometry problems? Strengths and weaknesses of LLMs for geometric reasoning in Spanish. *International Journal of Interactive Multimedia and Artificial Intelligence*, *8*(5), 65. https://doi.org/10.9781/ijimai.2024.02.009

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

Song, Y., Kim, J., Xing, W., Liu, Z., Li, C., & Oh, H. (2025). Elementary school students' and teachers' perceptions toward creative mathematical writing with Generative AI. *Journal of Research on Technology in Education*, *57*(2), 1–23. https://doi.org/10.1080/15391523.2025.2455057

Song, Y., Li, C., Xing, W., Lyu, B., & Zhu, W. (2025). Investigating perceived fairness of AI prediction system for math learning: A mixed-methods study with college students. *The Internet and Higher Education*, *65*, Article 101000. https://doi.org/10.1016/j.iheduc.2025.101000

Spreitzer, C., Straser, O., Zehetmeier, S., & Maaß, K. (2024). Mathematical modelling abilities of artificial intelligence tools: The case of ChatGPT. *Education Sciences*, *14*(7), Article 698. https://doi.org/10.3390/educsci14070698

Sultan, Y., Dautova, G., & Dalle, J. (2025). Examining the Relationship Among Artificial Intelligence Literacy, Cultural Literacy, and Intercultural Communication Proficiency of Philology Students. *Journal of Ethnic and Cultural Studies*, *12*(5), 345–362. https://doi.org/10.29333/ejecs/2839

Sutomo, W. A. B., & Turmudi, T. (2025). Integration of artificial intelligence in mathematics learning: Systematic literature review. *SJME (Supremum Journal of Mathematics Education)*, *9*(1), 29–41. https://doi.org/10.35706/sjme.v9i1.100

Tashtoush, M. A., Qasimi, A. B., Shirawia, N. H., & Hussein, L. A. (2025). The efficacy of utilizing artificial intelligence techniques in developing critical thinking in mathematics among secondary school students and their attitudes toward it. *Iraqi Journal for Computer Science and Mathematics*, *6*(1), 102–115. https://doi.org/10.52866/2788-7421.1231

Tashtoush, M. A., Wardat, Y., Ali, R. A., & Saleh, S. (2024). Artificial intelligence in education: Mathematics teachers' perspectives, practices and challenges. *Iraqi Journal for Computer Science and Mathematics*, *5*(1), 45–58. https://doi.org/10.52866/ijcsm.2024.05.01.004

Thapa Magar, A., Shakya, A., Fancsali, S. E., Rus, V., Murphy, A., Ritter, S., & Venugopal, D. (2025). "Can a language model represent math strategies?": Learning math strategies from big data using BERT. *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 655–666. https://doi.org/10.1145/3706468.3706558

Topali, P., Coetsier, N., & Molenaar, I. (2025). Engaging math teachers as co-designers of a classroom analytics dashboard. *CEUR Workshop Proceedings*, *3995*, 112–119.

Udias, A., Alonso-Ayuso, A., Alfaro, C., Algar, M. J., Cuesta, M., Fernández-Isabel, A., Gómez, J., Lancho, C., Cano, E. L., Martín De Diego, I., & Ortega, F. (2024). ChatGPT's performance in university admissions tests in mathematics. *International Electronic Journal of Mathematics Education*, *19*(4), Article em0795. https://doi.org/10.29333/iejme/15517

Uğraş, H., Uğraş, M., Papadakis, S., & Kalogiannakis, M. (2024). ChatGPT-supported education in primary schools: The potential of ChatGPT for sustainable practices. *Sustainability*, *16*(22), 9855. https://doi.org/10.3390/su16229855

Wang, X., & Wei, Y. (2025). The influence of Gen-AI assisted learning on primary school students' math anxiety: An intervention study. *Applied Cognitive Psychology*, *39*(4), Article e70088. https://doi.org/10.1002/acp.70088

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

Yi, L., Liu, D., Jiang, T., & Xian, Y. (2025). The effectiveness of AI on K-12 students' mathematics learning: A systematic review and meta-analysis. *International Journal of Science and Mathematics Education*, *23*(4), 1105–1126. https://doi.org/10.1007/s10763-024-10499-7

Zacharis, G., & Papadakis, S. (2025). Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment. *Educational Process: International Journal*, *19*, Article e2025591. https://doi.org/10.22521/edupij.2025.19.591

Zhang, F., Li, C., Henkel, O., Xing, W., Baral, S., Heffernan, N., & Li, H. (2025). Math-LLMs: AI cyberinfrastructure with pre-trained transformers for math education. *International Journal of Artificial Intelligence in Education*, *35*(2), 509–532. https://doi.org/10.1007/s40593-024-00416-y

Şimşek, N. (2025). Integration of ChatGPT in mathematical story-focused 5E lesson planning: Teachers and pre-service teachers' interactions with ChatGPT. *Education and Information Technologies*, *30*(8), 11391–11462. https://doi.org/10.1007/s10639-024-13258-x

Appendix A

Table A1

Codification of papers included in the systematic review

Code	Year	Authors	Journal name	Title
S1	2025	Li, C., Xing, W., Song, Y., & Lyu, B.	Computers and Education: Artificial Intelligence	RICE AlgebraBot: Lessons learned from designing and developing responsible conversational AI using induction, concretization, and exemplification to support algebra learning
S2	2025	Chisholm, K., & Bergwall, E.	Computers and Education: Artificial Intelligence	"I just think it is the way of the future": Teachers' use of ChatGPT to develop motivationally-supportive math lessons
S3	2025	Oh, S.	IEEE Access	Integration of MATH41 and Generative AI in Pre-Service Mathematics Teacher Education: An Empirical Study on Lesson Design Competency
S4	2025	Şimşek, N.	Education and Information Technologies	Integration of ChatGPT in mathematical story-focused 5E lesson planning: Teachers and pre-service teachers' interactions with ChatGPT
S5	2025	Mondal, S., Khatua, D., Mandal, S., Prasad, D. K., & Sekh, A. A.	Discover Artificial Intelligence	BMWP: The first Bengali math word problems dataset for operation prediction and solving
S6	2025	Lyu, B., Li, C., Li, H., Oh, H., Song, Y., Zhu, W., & Xing, W.	Proceedings of the 15th International Learning Analytics and Knowledge Conference	Exploring the Role of Teachable AI Agents' Personality Traits in Shaping Student Interaction and Learning in Mathematics Education
S7	2025	Kim, Y. R., Park, M. S., & Joung, E.	School Science and Mathematics	Exploring the integration of artificial intelligence in math education: Preservice Teachers' experiences and reflections on problem-posing activities with ChatGPT
S8	2025	Thapa Magar, A., Shakya, A., Fancsali, S. E., Rus, V., Murphy, A., Ritter, S., & Venugopal, D.	Proceedings of the 15th International Learning Analytics and Knowledge Conference	"Can A Language Model Represent Math Strategies?": Learning Math Strategies from Big Data using BERT
S9	2025	Li, C., & Lyu, B.	British Journal of Educational Technology	Investigating the motivational and knowledge affordances of conversational AI using induction, concretization and exemplification in math learning
S10	2025	Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R.	Proceedings of the National Academy of Sciences	Generative AI without guardrails can harm learning: Evidence from high school mathematics
S11	2025	Oğuz Haçat, S., & Kepceoğlu, İ.	The Journal of Educational Research	Help of ChatGPT about math skills on the subjects of social studies
S12	2025	Topali, P., Coetsier, N., & Molenaar, I.	CEUR Workshop Proceedings	Engaging Math Teachers as Co-designers of a Classroom Analytics Dashboard
S13	2024	Alvarez, J. I.	Jurnal Ilmiah Ilmu	Evaluating the impact of ai-powered tutors mathgpt and flexo 2.0 in enhancing calculus learning
S14	2024	Udias, A., Alonso-Ayuso, A., Alfaro, C.,	International Electronic	ChatGPT's performance in university

		Algar, M. J., Cuesta, M., Fernández-Isabel, A., Gómez, J., Lancho, C., Cano, E. L., Martín De Diego, I., & Ortega, F.	Journal of Mathematics Education	admissions tests in mathematics
S15	2024	Parra, V., Sureda, P., Corica, A., Schiaffino, S., & Godoy, D.	International Journal of Interactive Multimedia and Artificial Intelligence	Can Generative AI Solve Geometry Problems? Strengths and Weaknesses of LLMs for Geometric Reasoning in Spanish
S16	2024	Lawrence, L., Echeverria, V., Yang, K., Aleven, V., & Rummel, N.	British Journal of Educational Technology	How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process
S17	2024	Spreitzer, C., Straser, O., Zehetmeier, S., & Maaß, K.	Education Sciences	Mathematical Modelling Abilities of Artificial Intelligence Tools: The Case of ChatGPT
S18	2024	Ventura, A. M. C.	International Journal of Information and Education Technology	Unlocking the Future of Learning: Assessing Students' Awareness and Usage of AI Tools
S19	2024	Henkel, O., Horne-Robinson, H., Kozhakhmetova, N., & Lee, A.	Artificial Intelligence in Education (Workshops & Tutorials)	Effective and Scalable Math Support: Experimental Evidence on the Impact of an AI-Math Tutor in Ghana
S20	2024	Alhazzani, N. S.	International Journal of Innovative Research and Scientific Studies	Enhancing mathematics teachers' pedagogical skills by using ChatGPT
S21	2023	Dahal, N., Lamichhane, B. R., Luitel, B. C., & Pant, B. P.	Proceedings of the 28th Asian Technology Conference in Mathematics	AI Chatbots as Math Algorithm Problem Solvers: A Critical Evaluation of Its Capabilities and Limitations
S22	2022	Jancarik, A., Novotna, J., & Michal, J.	European Conference on E-Learning	Artificial Intelligence Assistant for Mathematics Education
S23	2021	Lu, Y., Pian, Y., Chen, P., Meng, Q., & Cao, Y.	Proceedings of the AAAI Conference on Artificial Intelligence	RadarMath: An Intelligent Tutoring System for Math Education
G24	2025	Tashtoush, M. A., Qasimi, A. B., Shirawia, N. H., & Hussein, L. A.	Iraqi Journal for Computer Science and Mathematics	The Efficacy of Utilizing Artificial Intelligence Techniques in Developing Critical Thinking in Mathematics among Secondary School Students and their Attitudes Toward it
G25	2025	Gabriel, F., Kennedy, J., Marrone, R., & Leonard, S.	Npj Science of Learning	Pragmatic AI in education and its role in mathematics learning and teaching
G26	2025	Cho, M. K & Kim, S.	International Electronic Journal of Mathematics Education	Analyzing AI-based educational platforms for supporting personalized mathematics learning
G27	2025	Inweregbuh, Onyemauche Christopher & Ugwuanyi, Chika Charity	*African Journal of Science, Technology and Mathematics Education (AJSTME)*	Mathematics Teachers' Self-Efficacy in Utilizing Artificial Intelligence (AI) Tools as Correlate of their Professional Development in Mathematics
G28	2025	Olarewaju, Adijat Omoladun, Moshood, Abdullateef, & Awoyemi, Sunday Aremu	Al-Hikmah Journal of Education	Pre-Service Teachers' Awareness and Utilization of AI Tools in Mathematics Education: Evidence from Kwara State
G29	2025	Akosah, E. F.	International Journal of Technology in Education and Science	AI-Assisted Collaborative Learning in Mathematics Education: A Qualitative Approach
G30	2025	Garcia, K. F., Ong, A. K. S., Gumasing, Ma. J. J., & Delos Reyes, C. R. V.	Acta Psychologica	Engineering students' perceptions and actual use of AI-based math tools for solving mathematical problems
G31	2025	Lu, Y., Supriya, K., Shaked, S., Simmons, E. H., & Kusenko, A.	Physical Review Physics Education Research	Incentivizing supplemental math assignments and using AI-generated hints is associated with improved exam performance
G32	2025	Otero, N., Druga, S., & Lan, A.	Discover Education	A Benchmark for math misconceptions: Bridging gaps in middle school algebra with AI-supported instruction
G33	2025	Song, Y., Kim, J., Xing, W., Liu, Z., Li, C., & Oh, H.	Journal of Research on Technology in Education	Elementary school students' and teachers' perceptions toward creative mathematical writing with Generative AI
G34	2025	Song, Y., Li, C., Xing, W., Lyu, B., & Zhu, W.	The Internet and Higher Education	Investigating perceived fairness of AI prediction system for math learning: A mixed-methods study with college students
G35	2025	Wang, X., & Wei, Y.	Applied Cognitive Psychology	The Influence of Gen-AI Assisted Learning on Primary School Students' Math Anxiety: An Intervention Study
G36	2025	Zhang, F., Li, C., Henkel, O., Xing, W., Baral, S., Heffernan, N., & Li, H.	International Journal of Artificial Intelligence in Education	Math-LLMs: AI Cyberinfrastructure with Pre-trained Transformers for Math Education
G37	2024	Inoferio, H. V., Espartero, M., Asiri, M., Damin, M., & Chavez, J. V.	Environment and Social Psychology	Coping with math anxiety and lack of confidence through AI-assisted Learning
G38	2024	Li, C., Xing, W., & Leite, W.	Interactive Learning Environments	Using fair AI to predict students' math learning outcomes in an online platform
G39	2024	Schorcht, S., Buchholtz, N., &	Frontiers in Education	Prompt the problem – investigating the

		Baumanns, L.		mathematics educational quality of AI-supported problem solving by comparing prompt techniques
G40	2024	Tashtoush, M. A., Wardat, Y., Ali, R. A., & Saleh, S.	Iraqi Journal for Computer Science and Mathematics	Artificial Intelligence in Education: Mathematics Teachers' Perspectives, Practices and Challenges
G41	2024	Kanvaria, V. K.	Revista Electronica De Veterinaria	Artificial Intelligence In The Classroom: Experimental Research On Innovative Approaches To Mathematics Instruction
G42	2024	Egara, F. O., & Mosimege, M.	Education Sciences	Exploring the Integration of Artificial Intelligence-Based ChatGPT into Mathematics Instruction: Perceptions, Challenges, and Implications for Educators

Appendix B

Methodological Quality Appraisal of Included Studies (MMAT 2018)

Y = Yes; N = No; CT = Cannot tell from the verified methods information; N/A = MMAT not applicable to the evidence type. S1–S2 are the two MMAT screening questions; Q1–Q5 are the five design-specific criteria. CT was deliberately retained where the available report did not support a defensible Yes/No judgment. Technical/model-performance papers were not converted into pedagogical-effect evidence.

Interpretive note. MMAT is used here to prevent heterogeneous evidence from being treated as equivalent. A "Cannot tell" judgment is not converted to a negative score, and no overall percentage score is calculated. Studies outside MMAT's empirical design categories are marked N/A and contribute only to technology-capability or contextual mapping, not to conclusions about educational effectiveness.

Code	Study	MMAT design	S1	S2	Q1	Q2	Q3	Q4	Q5	Appraisal note	Use in synthesis
S1	Li et al. (2025), RICE AlgebraBot	Design/development + empirical evaluation	Y	Y	CT	CT	CT	CT	CT	Large authentic Math Nation dataset and development/evaluation studies are reported; accessible report confirms 2,097,139 reply pairs from 71,918 students/tutors. Full design-specific MMAT judgments require the complete empirical evaluation details.	Empirical synthesis; interpret according to design/quality.
S2	Chisholm & Bergwall (2025)	Teacher study; design not verified	CT	CT	CT	CT	CT	CT	CT	Appendix provides bibliographic identification, but the methods needed for criterion-level MMAT appraisal were not verified in the available source set.	Empirical synthesis; interpret according to design/quality.
S3	Oh (2025), MATH41 + GenAI	Empirical teacher-education study; design not fully verified	CT	CT	CT	CT	CT	CT	CT	Study is described as empirical in the title; criterion-level methods were not sufficiently available for defensible Yes/No judgments.	Empirical synthesis; interpret according to design/quality.
S4	Şimşek (2025)	Qualitative / multi-source case study	Y	Y	Y	Y	Y	Y	Y	Three teachers and three pre-service teachers; open-ended questionnaires, focus groups and prompt archives were analysed, with inter-rater reliability assessment reported. The qualitative approach is coherent with the research questions.	Empirical synthesis; interpret according to design/quality.
S5	Mondal et al. (2025), BMWP dataset	Technical dataset/benchmark	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Primarily a technical dataset/benchmark paper. MMAT is not the appropriate appraisal framework for this evidence type; retained only for technical-context synthesis, not pedagogical-effect claims.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.

Code	Study	MMAT design	S1	S2	Q1	Q2	Q3	Q4	Q5	Appraisal note	Use in synthesis
S6	Lyu et al. (2025), teachable AI agents	Experimental study; conference paper	Y	Y	CT	CT	CT	CT	CT	Random assignment to agents with different personality emphases is reported in accessible abstract-level evidence, but full RCT/non-randomized MMAT criteria were not all verifiable.	Empirical synthesis; interpret according to design/quality.
S7	Kim et al. (2025/2026), preservice teachers + ChatGPT	Empirical educational study; design not fully verified	Y	Y	CT	CT	CT	CT	CT	Study examines preservice teachers' AIM activities and error recognition/correction; accessible abstract confirms empirical educational evaluation but not enough detail for all MMAT criteria.	Empirical synthesis; interpret according to design/quality.
S8	Thapa Magar et al. (2025), BERT math strategies	Technical/modeling study	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Primary aim is representation/model performance rather than an empirical human educational study. MMAT not applied.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
S9	Li & Lyu (2025), conversational AI in algebra	Quantitative non-randomized / between-subjects pre-post	Y	Y	Y	CT	Y	Y	Y	Between-subjects pre/post experimental study with 151 analysed participants; eligibility/attention checks and intervention/comparison conditions are reported. Random allocation was not established in the accessible report, so non-randomized criteria are used conservatively.	Empirical synthesis; interpret according to design/quality.
S10	Bastani et al. (2025)	Randomized controlled trial	Y	Y	Y	CT	Y	Y	Y	Large-scale school RCT with random assignment to GPT Base, GPT Tutor, or control; preregistration, covariate balance, independent graders and a common rubric are reported. Blinding/completeness details are not fully established here.	Empirical synthesis; interpret according to design/quality.
S11	Oğuz Haçat & Kepceoğlu (2025)	Qualitative exploratory case study	Y	Y	Y	Y	Y	Y	Y	Four Grade 7 students; chat logs, written responses and semi-structured interviews were used. The case-study design and multiple qualitative data sources are appropriate to the exploratory question.	Empirical synthesis; interpret according to design/quality.
S12	Topali et al. (2025), dashboard co-design	Participatory design / preliminary qualitative work	Y	Y	CT	CT	CT	CT	CT	Preliminary human-centred co-design study with Dutch primary teachers. Because the accessible conference report is preliminary, several MMAT qualitative criteria remain Cannot tell.	Empirical synthesis; interpret according to design/quality.
S13	Alvarez (2024), MathGPT/Flexi 2.0	Educational intervention; design not verified	CT	CT	CT	CT	CT	CT	CT	The title indicates an impact evaluation in calculus, but the source set did not provide sufficient methods detail for criterion-level appraisal.	Empirical synthesis; interpret according to design/quality.

Code	Study	MMAT design	S1	S2	Q1	Q2	Q3	Q4	Q5	Appraisal note	Use in synthesis
S14	Udias et al. (2024), admissions-test performance	Technical performance evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Evaluates ChatGPT performance on mathematics admissions tests rather than effects on human learners. MMAT not applied; evidence treated as system-performance evidence only.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
S15	Parra et al. (2024), LLM geometry reasoning	Technical performance evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Evaluates LLM geometric reasoning. Not a human-participant empirical design covered by MMAT; used only for capability/limitation claims.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
S16	Lawrence et al. (2024), AI co-orchestration	Multiyear teacher-centred design study	Y	Y	CT	CT	CT	CT	CT	Teacher-centred multiyear design process is reported, but full qualitative/design appraisal details were not available in the reviewed source set.	Empirical synthesis; interpret according to design/quality.
S17	Spreitzer et al. (2024), ChatGPT modelling abilities	Technical performance evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Focuses on mathematical modelling performance of AI tools. MMAT not applied.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
S18	Ventura (2024), student awareness/use of AI	Quantitative descriptive; methods not verified	CT	CT	CT	CT	CT	CT	CT	Bibliographic/title information suggests an awareness/usage study, but sampling, measurement and response-rate details were not verified.	Empirical synthesis; interpret according to design/quality.
S19	Henkel et al. (2024), AI-math tutor in Ghana	Educational intervention; conference report	Y	Y	CT	CT	CT	CT	CT	Experimental evidence is reported in the title and proceedings citation, but the short conference format did not provide enough verified detail here for all intervention criteria.	Empirical synthesis; interpret according to design/quality.
S20	Alhazzani (2024), teacher pedagogical skills + ChatGPT	Empirical/conceptual status not fully verified	CT	CT	CT	CT	CT	CT	CT	Insufficient verified methods information for a defensible criterion-level MMAT appraisal.	Empirical synthesis; interpret according to design/quality.
S21	Dahal et al. (2023), AI chatbots as math solvers	Technical/critical evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Primarily evaluates chatbot mathematical problem-solving capability. MMAT not applied.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
S22	Jancarik et al. (2022), AI assistant	Design/development conference study	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Primarily a system/design contribution; no verified human empirical design suitable for MMAT was established.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
S23	Lu et al. (2021), RadarMath	System design/evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Primarily an intelligent tutoring system development/performance paper; MMAT not applied unless a separable human empirical component is documented.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
G24	Tashtoush et al. (2025), critical thinking	Educational intervention; methods not fully verified	CT	CT	CT	CT	CT	CT	CT	Title indicates efficacy evaluation among secondary students, but full design, allocation, confounding and outcome-completeness details were not verified.	Empirical synthesis; interpret according to design/quality.

Code	Study	MMAT design	S1	S2	Q1	Q2	Q3	Q4	Q5	Appraisal note	Use in synthesis
G25	Gabriel et al. (2025), pragmatic AI	Empirical/conceptual study; design not fully verified	CT	CT	CT	CT	CT	CT	CT	Insufficient verified methods detail for criterion-level MMAT appraisal.	Empirical synthesis; interpret according to design/quality.
G26	Cho & Kim (2025), AI platforms	Platform analysis / comparative evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Primary focus is analysis of AI-based educational platforms; no verified human empirical design suitable for MMAT was established.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
G27	Inweregbuh & Ugwuanyi (2025), teacher self-efficacy	Quantitative descriptive/correlational; methods not verified	CT	CT	CT	CT	CT	CT	CT	Title indicates a correlational teacher study, but sampling, instrument validity, response rate and analysis details were not verified.	Empirical synthesis; interpret according to design/quality.
G28	Olarewaju et al. (2025), preservice teacher awareness/use	Quantitative descriptive; methods not verified	CT	CT	CT	CT	CT	CT	CT	The title indicates a descriptive awareness/utilization study; criterion-level methods were not verified.	Empirical synthesis; interpret according to design/quality.
G29	Akosah (2025), collaborative learning	Qualitative phenomenological study	Y	Y	Y	Y	Y	Y	Y	Qualitative phenomenological design; purposive sample of 10 pre-service mathematics teachers; semi-structured interviews and thematic analysis are explicitly reported.	Empirical synthesis; interpret according to design/quality.
G30	Garcia et al. (2025), engineering students' AI-math tools	Quantitative descriptive/cross-sectional	Y	Y	Y	Y	Y	Y	Y	565 valid engineering-student respondents, purposive sampling, adapted questionnaire constructs and PLS-SEM are reported. Sampling is non-probability, so generalisability is limited, but the quantitative approach is coherent with the stated acceptance-model questions.	Empirical synthesis; interpret according to design/quality.
G31	Lu et al. (2025), AI-generated hints	Quantitative non-randomized intervention/observational	Y	Y	Y	CT	Y	Y	CT	Course-based intervention links supplemental assignments and AI-generated hints with exam performance. The paper uses association language; random allocation is not established, so confounding/complete outcome criteria are judged conservatively.	Empirical synthesis; interpret according to design/quality.
G32	Otero et al. (2025), misconception benchmark	Technical benchmark with educational framing	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Benchmark-oriented study; MMAT not applied to technical performance evidence.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
G33	Song et al. (2025), creative mathematical writing	Perception/educational study; methods not fully verified	CT	CT	CT	CT	CT	CT	CT	The title establishes student/teacher perception focus, but sampling, measurement and analysis details were not sufficiently verified.	Empirical synthesis; interpret according to design/quality.
G34	Song et al. (2025), perceived fairness	Mixed methods (explanatory sequential)	Y	Y	Y	Y	Y	Y	Y	428 college students in randomized between-subjects experiment plus purposively selected one-to-one interviews; regression/ANOVA	Empirical synthesis; interpret according to design/quality.

Code	Study	MMAT design	S1	S2	Q1	Q2	Q3	Q4	Q5	Appraisal note	Use in synthesis
										and thematic analysis are reported, with qualitative findings used to elaborate quantitative results.	
G35	Wang & Wei (2025), math anxiety	Mixed methods / quasi-experimental intervention	Y	Y	CT	CT	Y	Y	Y	Mixed quasi-experimental intervention with pre/post quantitative comparisons and semi-structured interviews is reported. Integration is evident, but full confounding and sampling details were not all verified.	Empirical synthesis; interpret according to design/quality.
G36	Zhang et al. (2025), Math-LLMs	Technical infrastructure/model study	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Primarily AI cyberinfrastructure/model development and performance. MMAT not applied to technical evidence.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
G37	Inoferio et al. (2024), math anxiety/confidence	Empirical educational study; design not verified	CT	CT	CT	CT	CT	CT	CT	Outcome focus is educational/affective, but criterion-level methods were not sufficiently verified.	Empirical synthesis; interpret according to design/quality.
G38	Li, Xing & Leite (2024), fair AI prediction	Technical predictive-model evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Develops and compares Fair-LR with other prediction models using fairness/accuracy metrics. This is algorithmic evaluation rather than a human educational-effect study; MMAT not applied.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
G39	Schorcht et al. (2024), prompt techniques	Comparative AI-output/education-quality evaluation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Primary unit is AI-supported problem-solving/prompt quality rather than a human-participant empirical study suitable for MMAT.	Context/technology capability only; excluded from claims of learner/teacher effectiveness.
G40	Tashtoush et al. (2024), teacher perspectives	Quantitative descriptive survey	Y	Y	CT	Y	Y	CT	Y	Survey of 580 mathematics teachers across public/private schools and three Abu Dhabi educational regions is reported. Full sampling representativeness, nonresponse and instrument details require methods-level verification.	Empirical synthesis; interpret according to design/quality.
G41	Kanvaria (2024), AI classroom experiment	Educational experiment; methods not fully verified	CT	CT	CT	CT	CT	CT	CT	The title identifies experimental research, but allocation, baseline comparability, outcome completeness and measurement details were not verified.	Empirical synthesis; interpret according to design/quality.
G42	Egara & Mosimege (2024), ChatGPT instruction	Mixed methods (sequential exploratory)	Y	Y	Y	Y	Y	Y	Y	Sequential mixed-methods design: survey of 80 mathematics teachers selected with stratified random sampling followed by in-depth interviews with five ChatGPT-familiar teachers. Quantitative and qualitative strands address the same integration question.	Empirical synthesis; interpret according to design/quality.