Beyond Boundaries: How Mathematics Education Research Frames Generative AI Use Through ZPA–ZFM

Sitti Maesuri Patahuddin
University of Canberra
Sitti.Patahuddin@canberra.edu.au

Siti Rokhmah
Perempuan Pendidik Numerasi
sitirokhmah2020@gmail.com

Agustin Ernawati
Perempuan Pendidik Numerasi
agustin.rnwt@gmail.com

Susilahudin Putrawangsa
Universitas Islam Negeri Mataram
putrawangsa@uinmataram.ac.id

This paper investigates how generative artificial intelligence (GenAI) use is explicitly promoted in mathematics education research and how these promoted actions are constrained by reported challenges and ethical concerns. Using the Zone of Promoted Action (ZPA) as an analytic lens, we systematically examine empirical studies to identify promoted practices, sources of promotion, and their structural boundaries. Findings indicate that GenAI is most commonly promoted for instructional planning, feedback, and assessment support, while concerns related to reliability, training, curriculum alignment, and ethics delimit the scope of these promoted uses.

Introduction

Generative artificial intelligence (GenAI) tools have rapidly entered educational discourse and practice, with mathematics education emerging as a prominent area of exploration. Recent advances in GenAI have enabled automated generation of explanations, feedback, lesson plans, assessment tasks, and worked examples, raising both opportunities and concerns for the teaching and learning of mathematics (e.g., Canonigo, 2024; Davis, 2025; Henkel et al., 2025; Kim et al., 2025; Taani & Alabidi, 2024). In response, mathematics education researchers have begun to examine how these tools might support instructional practice, student learning, and teacher professional development (Biton & Segal, 2025; Kim et al., 2025; Pepin et al., 2025; Walkington, 2025).

To date, much of the empirical research on GenAI in mathematics education has focused on adoption, perceptions, and impact. Studies commonly investigate teachers' attitudes towards GenAI, their willingness to integrate such tools into practice, or perceived benefits and risks associated with their use (Busuttil & Calleja, 2025; Wahba et al., 2024, Zhuang & Zhang, 2025). Others examine student engagement, learning outcomes, or performance when GenAI-supported tools are introduced into instructional settings (Fardian et al., 2025; Ramprakash et al., 2024; Segal & Biton, 2024; Torres-Peña et al., 2024). While this body of work has contributed valuable insights into how GenAI is taken up and experienced, it often positions GenAI use as an individual choice or pedagogical intervention, foregrounding questions of effectiveness, acceptance, or readiness.

Existing GenAI-in-mathematics-education studies report uses and outcomes but rarely interrogate how the literature itself constructs and legitimises "appropriate" use through designs, pedagogical framings, PD models, and AI-mediated prompts. Crucially, challenges—accuracy, transparency, privacy, integrity, curriculum alignment—are treated as afterthoughts rather than structural conditions shaping what gets promoted (e.g., Gurl et al., 2025; Karaman & Goksu, 2024). This study is novel in shifting the unit of analysis from "reported practice" to the field's promotional logics, and in jointly analysing promotion and constraint as co-configuring legitimate GenAI practice. Methodologically, we conduct a PRISMA-guided systematic review (Page et al., 2021), develop and iteratively refine a coding scheme distinguishing promoted actions and reported constraints, and map co-occurrence patterns to surface configuration logics. We ask: *What forms of GenAI use are explicitly promoted in*

(2026). In N. Berger, M. Thompson, J. Ley & C. Wilsmore (Eds.), *Beyond boundaries: Strengthening futures in mathematics education. Proceedings of the 48th annual conference of the Mathematics Education Research Group of Australasia, Sydney* (pp. 305–312). MERGA.

mathematics education research, and how are these constrained or related by reported challenges, limitations, and ethical concerns?

Conceptual Framework: ZPA and Structural Boundaries

This study adopts the Zone of Promoted Action (ZPA) as its primary conceptual lens for examining how GenAI use is positioned within mathematics education research. Building on Goos et al. (2003), ZPA refers to the deliberate shaping of learning environments to create opportunities for particular forms of action. In classroom contexts, such shaping occurs through pedagogical decisions that prompt exploration, encourage mathematical reasoning, and extend student activity. ZPA thus captures how teaching involves not only enabling action, but also structuring the space within which action is possible.

In technology-rich mathematics learning environments, pedagogical shaping is increasingly mediated by digital tools. From this perspective, features such as prompt design, automated feedback, and AI-generated scaffolds can be understood as digital forms of ZPA, as they foreground particular approaches, representations, or problem-solving pathways when learners engage with GenAI (Biton et al., 2025; Slobodsky & Durcheva, 2025). ZPA therefore captures how particular actions are *invited and foregrounded* through design features and pedagogical moves.

Goos (2005) further conceptualises ZPA as the actions, strategies, and practices promoted by experts, mentors, and professional programs. For pre-service teachers, ZPA may be shaped through coursework, professional development, or supervision that *models, exemplifies, and invites uptake of* specific pedagogical approaches. This highlights that ZPA is not solely produced through individual teaching decisions but is socially and institutionally constituted through *endorsed professional discourses and exemplars* that *foreground* what is valued as effective practice.

Extending this notion, the present study examines ZPA as it is evidenced within empirical studies of mathematics education. ZPA is operationalised as the practices, strategies, or uses of GenAI that are explicitly reported as being encouraged, modelled, or required within the study contexts, such as professional development activities, teacher education programs, classroom interventions, or AI-mediated learning environments. In this sense, ZPA refers to promoted actions within the empirical settings described by the studies, rather than to individual teacher choice or inferred practice.

To maintain analytic precision, the ZPA is defined narrowly to include only practices explicitly promoted in empirical studies—such as stated recommendations, guided or modelled uses, or GenAI presented as an intended component of instruction, assessment, or professional learning. No inferences are made beyond what is reported. The identified ZPA reflects promotion occurring across the system, including PD initiatives, teacher-educator guidance, institutional expectations, and built-in AI supports within GenAI tools.

This study's central contribution is to analyse promoted actions (ZPA) *together with* reported challenges, limitations, and ethical concerns as *boundary-setting conditions* that shape promotion. Analytically, this study distinguishes *promotive moves* (what is encouraged or modelled: ZPA) from *boundary conditions* (what is permitted/feasible: Zone of Free Movement (ZFM)), while recognising they co-occur in study descriptions. It treats these *boundary conditions* as structural, boundary-setting conditions that determine how far—and in what ways—GenAI use is promoted in mathematics education research. Thus, promoted actions and boundary conditions are analytically inseparable in reporting—for example, *promoting* GenAI for planning (ZPA) while *restricting* assessment use due to integrity concerns (ZFM).

Methodology

This study employed a systematic literature review to examine how GenAI use is explicitly promoted in mathematics education research and how these promoted actions are constrained by reported challenges, limitations, and ethical concerns. The review followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA 2020) guidelines (Page et al., 2021) to ensure transparency and rigour in study identification, screening, and reporting. Searches were conducted across ERIC, Scopus, and Web of Science using Boolean combinations of GenAI-related and mathematics-education keywords. The search was limited to peer-reviewed journal articles published between 2018 and 2025, and reference lists of included studies were also checked for additional sources. Studies were included if they were written in English, explicitly focused on mathematics education, examined GenAI as a central analytic object, and employed empirical qualitative, quantitative, or mixed-methods designs involving students, in-service or pre-service teachers, or teacher educators. Conceptual papers, broader STEM studies without a clear mathematics focus, and studies mentioning AI only peripherally were excluded. A total of 344 records were imported into Covidence; after removing duplicates, 328 titles and abstracts were screened followed by full-text screening of 91 potentially relevant studies. The screening of each abstract and full-text were conducted by two reviewers. Disagreements between reviewers were resolved through discussion or independently by the third reviewer, resulting in a final corpus of 43 empirical studies.

Data were extracted using a structured coding template informed by a ZPD–ZFM–ZPA[1 3] framework (Goos et al., 2003; Goos, 2005), with this paper focusing specifically on ZPA and ZFM. ZPA was operationalised as practices, behaviours, or uses of GenAI that were explicitly encouraged or promoted within the study, including through professional development, teacher educator guidance, institutional or curricular directives, or AI-embedded supports. This study analytically distinguishes ZPA—promotive moves that encourage or model particular actions—from ZFM—boundary conditions that define what is permitted or feasible—while recognising that, in empirical reports, these often co-occur and are difficult to disentangle because promoted actions are typically presented alongside their enabling and constraining conditions. To maintain analytic rigour, only information explicitly stated in the studies was coded. No inferences were made beyond the text. When a promoted action, challenge, limitation, or ethical concern was not explicitly reported, it was coded as N/A. Each coded item was accompanied by a reference to its location in the article (e.g., section heading and paragraph or page number).

After completing data extraction in Covidence, we exported the outputs to an Excel matrix for subsequent coding. An initial attempt to analyse the exported fields using freely available institutional and one of the most popular GenAI tools produced inconsistent and sometimes inaccurate classifications; therefore, we adopted manual coding. Two researchers independently coded all data extraction, compared code applications, and resolved discrepancies through discussion to ensure interpretive consistency and analytic trustworthiness. Each code was operationalised using binary indicators (1 = present; 0 = absent). Coding focused on four domains: promoted GenAI use, mediations of promoted GenAI use, reported challenges/limitations, and reported ethical concerns. In total, 40 codes were developed, organised into four major themes and 18 sub-themes (listed in the Supplementary Materials). We also coded contextual descriptors (i.e., year of publication, country, population, mathematics content/strand).

[3] ZPD (zone of proximal development) refers to the range of learning possible with support; ZFM (zone of free movement) refers to the contextual constraints and possibilities shaping action; and ZPA (zone of promoted action) refers to the actions and practices encouraged within that setting. Together, these constructs frame how learning and participation are enabled, constrained, and directed.

To examine how promoted forms of GenAI use relate to reported challenges and ethical concerns, we conducted a co-occurrence analysis of the binary coding matrix. Pairwise associations were estimated using Jaccard similarity to capture shared presences across the 43 studies. A weighted, undirected network (Jaccard ≥ 0.20) was then constructed to represent the strongest associations and was complemented by a Jaccard heatmap to show the broader similarity structure (Podani & Schmera, 2011). All computations and visualisations were performed in Python using standard scientific libraries.

Findings and Discussion

Research Demographic Background

The publication-year distribution reveals a recent and concentrated evidence base: 97.6% of studies were published in 2024–2025, with only one study from 2023 (2.3%). The reviewed literature focuses mainly on GenAI use by teachers (60.5%), students (34.9%), or both teachers and students (4.7%) in mathematics education, largely within school and teacher-education settings. This recency and concentration suggest a field still in active formation, where conceptual frameworks, research designs, and pedagogical claims remain fluid.

Geographical coverage was uneven. The largest share of studies originated from North America (20.9%), followed by the Middle East (16.3%), Europe (11.6%), and Southeast Asia (11.6%). Representation was markedly lower for East Asia (7.0%) and Sub-Saharan Africa (2.3%), and nearly one-third of studies (30.2%) did not specify a national context. When reclassified using a Global North–Global South framing, 44% were situated in the Global North and 26% in the Global South. This imbalance mirrors long-standing structural asymmetries in educational technology research and constrains both contextual generalisability and theorisation sensitive to sociocultural and infrastructural variation.

Participant profiles were similarly concentrated. Pre-service teachers ($n = 21$; 48.8%) and students ($n = 15$; 34.9%) accounted for 83.7% of the sample, whereas in-service teachers and teacher educators were each examined in only two studies (4.7%). This distribution reflects a dominance of early-stage exploratory work situated in controlled learning environments, rather than sustained investigations of classroom enactment or professional practice. Content coverage followed a parallel pattern: mathematics was often addressed generically (23.1%), with fewer studies engaging specific domains such as Algebra (17.3%), Number Sense (17.3%), or Geometry/Measurement (15.4%). Advanced mathematics appeared infrequently (11.5%), while Calculus and Statistics/Probability were least represented (7.7% each). This pattern suggests that current evidence is drawn largely from foundational content areas, limiting how far claims about GenAI can yet extended across the breadth of school mathematics.

Promoted GenAI Practices: Constraints and Ethical Concerns

Next, we address the research question by synthesising what mathematics education research explicitly promotes educators and learners to do with GenAI, and how these advocated uses are shaped by reported challenges, limitations, and ethical concerns. We first characterise the dominant forms of promoted GenAI use across the included studies. We then examine how these promotions are mediated by constraints—such as reliability, pedagogical fit, access and workload—and by ethical considerations including academic integrity, responsibility and oversight, privacy, and fairness. Rather than treating promotion and constraint as separate themes, we interpret their co-occurrence as indicative of the conditions under which GenAI is positioned as educationally legitimate, feasible, and defensible. The synthesis therefore clarifies not only what is being advocated, but also under what boundaries and with what cautions these advocated practices are framed.

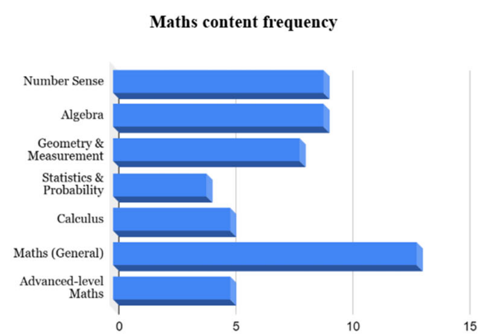
Across the four analytic domains (illustrated in Figure 1), GenAI use in mathematics education is situated within a cautiously delimited zone of promoted action. Promotion concentrates most strongly on explanation and conceptual support (A2; 28.9%), followed by critical, reflective, and responsible use (A5; 22.5%) and classroom pedagogy and mathematical interaction (A3; 17.6%), with comparatively limited attention to assessment and feedback (A4; 5.6%). These promoted uses are typically enacted through guided practice or structured activities (B3; 38.6%) and explicit professional development (B1; 37.1%), while system-level scaffolds—such as fixed prompts or built-in AI supports—remain marginal. This action space is shaped by constraints primarily related to accuracy and reliability (C1; 41.7%), followed by social and pedagogical concerns (C2; 27.0%), technical constraints (C3; 15.7%), and capability and knowledge gaps (C4; 15.7%). Ethical considerations add another layer of limitation, with strong emphasis on academic integrity and oversight (D1; 58.3%), and less frequent attention to bias and equity (D2; 25.0%) or privacy and security (D3; 16.7%), reflecting the continued influence of governance and institutional risk management.

Figure 1

The ZPA and ZFM analytic domains and mathematics content focus

Domains	%
A: Promoted forms of GenAI Use	
A1. Instructional design and task development	13.4
A2. Explanation and Conceptual Support	28.9
A3. Classroom pedagogy and mathematical interaction	17.6
A4. Assessment, feedback, and diagnostic Support	5.6
A5. Critical, Reflective, and Responsible Use	22.5
A6. Teacher Education and Professional Practice	12
B: Mediations of promoted GenAI Use	
B1. Explicit professional development or workshops	37.1
B2. Modelling of GenAI use by instructors or leaders	10
B3. Guided practice or structured activities	38.6
B4. Provided fixed prompts or Prescribed interaction protocols	4.3
B5. Built-in GenAI supports	10
C: Reported challenges and limitations	
C1. Accuracy and Reliability Issues	41.7
C2. Social and Pedagogical Concerns	27
C3. Technical and Practical Constraints	15.7
C4. Capability and Knowledge Gaps	15.7

Domains	%
D: Ethical concerns and risks	
D1. Academic Integrity, Responsibility and Oversight	58.3
D2. Bias, Fairness, and Equity	25
D3. Privacy or Security Concerns	16.7

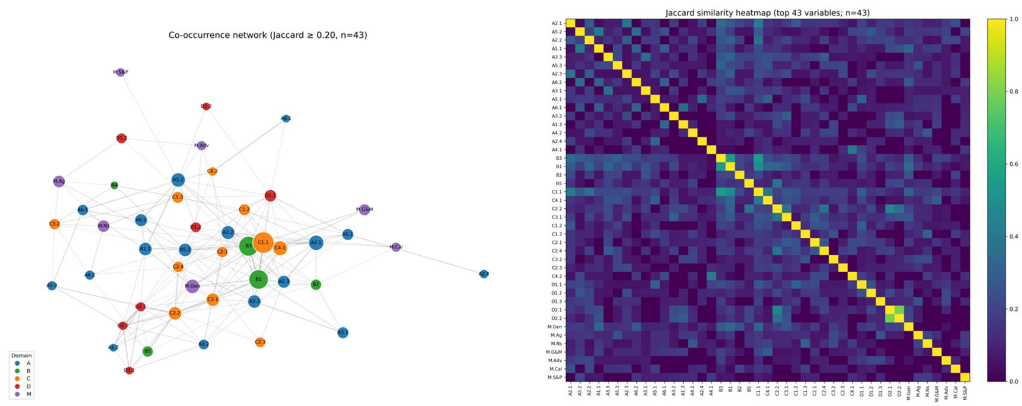


The co-occurrence network (Jaccard ≥ .20, n = 43) accompanied by the heatmap (see Figure 2) further clarifies how these patterns operate in combination. The field is organised around a practice-centred explanatory core, where the A2 uses—explanations, worked solutions, alternative strategies, and sense-making—form dense ties with B1 (professional development) and B3 (guided practice). These connections sit alongside two recurring constraints: C1.1 (GenAI output unreliability) and C4.1 (AI literacy limitations). Together, they indicate that explanatory uses are framed not as direct affordances but as practices requiring structured human mediation and ongoing verification. A3.3 (interactive questioning) aligns with this core, whereas A4 (assessment and diagnosis) and most D-domain concerns remain peripheral. Mathematics-content codes are also weakly integrated: only Maths (General) shows moderate connectivity, while specific strands such as Algebra, Geometry, Number, Advanced, and Statistics sit at the edges. Overall, the evidence base privileges pedagogical enactment over disciplinary specificity, and teacher- or program-led scaffolding over systemic or ethical analyses.

Taken together, these results provide a coherent picture of how GenAI is currently positioned within mathematics education research: promoted in carefully bounded ways, mediated through structured support, and accompanied by recurring concerns about accuracy, capability, and ethical oversight. The co-occurrence patterns, in particular, highlight the centrality of explanation-focused uses and the foundational role of professional development and guided practice in shaping how these uses are enacted. At the same time, several areas—including assessment, domain-specific mathematical content, and system-level ethical considerations—remain relatively underexplored. These patterns point toward important directions for future inquiry, which we elaborate in the following discussion.

Figure 2

Promoted uses and mediating constraints co-occur



Note. Code labels are defined in the Supplementary Materials; the full code list provides complete descriptions of each code.

1. Moving beyond scaffolded use towards classroom enactment

The co-occurrence network shows that promoted uses are tightly coupled with professional development (B1) and guided practice (B3). While this indicates responsible, well-scaffolded experimentation, it also reveals a field that is heavily anchored in controlled learning contexts rather than authentic classroom settings. There is a need for design-based, classroom-embedded, and longitudinal studies that trace how teachers' judgement, students' reasoning, and patterns of feedback evolve as GenAI becomes entangled in everyday teaching. Such studies would clarify not only what GenAI can do, but how teachers actually take it up within real pedagogical and institutional constraints.

2. Bridging reliability concerns with capability building

Across the literature, C1.1 (unreliability) and C4.1 (AI literacy gaps) consistently co-occur with explanatory uses (A2). This pairing suggests an opportunity for research focused on mechanisms of verification, correction, and sense-making. Rather than treating inaccuracy as an inherent limitation, future studies could investigate instructional routines—such as cross-checking, re-prompting, or pair-reasoning—that strengthen students' and teachers' capacity to evaluate GenAI outputs (Park & Choo, 2025). This direction aligns with broader trends in mathematics education emphasising critical engagement with tools rather than simple adoption (Cullen & Hertel, 2023).

3. Deepening the mathematical substance of GenAI research

The predominance of generic mathematics codes and the marginality of Algebra, Geometry, Number, and Statistics indicate that current studies rarely examine how GenAI interacts with the structure of mathematical ideas. This gap matters, because explanation quality, conceptual accuracy, and error-patterns differ substantially across domains. Future work would benefit from content-specific task design, enabling closer analysis of how GenAI handles symbolic manipulation, geometric reasoning, or stochastic representations, and how teachers and students evaluate these outputs. Such work would also support stronger claims about the disciplinary integrity of GenAI-mediated reasoning.4. Integrating ethical and equity considerations into pedagogical inquiry

Ethical concerns, especially academic integrity (D1), appear in the corpus but remain structurally peripheral in the network, with equity (D2) and privacy (D3) even less connected. Given the differential access to GenAI, linguistic inequalities, and the risks associated with data exposure, future work should integrate system-level ethics directly into pedagogical research designs. Studies in multilingual and under-resourced contexts are particularly needed to understand how GenAI behaviour, institutional policy, and infrastructural realities shape pedagogical possibilities.

5. Strengthening methodological transparency and reporting

Finally, the patterns observed in the network highlight the need for stronger methodological transparency. Future empirical work should report the prompting strategies, model versions, verification procedures, and AI-literacy baselines used in studies, as these factors directly shape the validity and replicability of findings. Pairing co-occurrence or descriptive analyses with mechanism-focused designs (e.g., mediation analysis, process tracing, or microgenetic methods) would help identify why certain GenAI uses are effective and how mediations mitigate or exacerbate risks.

In conclusion, the review highlights a field in transition—experimenting with GenAI's potential while negotiating its constraints. Future research must move beyond early trials to examine how GenAI interacts with mathematical thinking, classroom practice, and the diverse realities of learners and teachers. Although policy is not this paper's primary focus, the findings may also inform mathematics education policy by identifying conditions for the thoughtful, responsible, pedagogically meaningful integration of emerging technologies.

Acknowledgement

Generative AI tools were used to support language editing and improve clarity of expression in this paper. All substantive decisions regarding the study design, analysis, interpretation, and final wording were made by the authors, who reviewed and verified all content for accuracy and originality.

References

Biton, Y., & Segal, R. (2025). Learning to craft and critically evaluate prompts: The role of Generative AI (ChatGPT) in enhancing pre-service mathematics teachers' TPACK and problem-posing skills. *International Journal of Education in Mathematics, Science and Technology, 13*(1), 202-223. https://doi.org/10.46328/ijemst.4654

Biton, Y., Segal, R., & Alush, K. (2025). How utilizing Generative AI when addressing pedagogical and mathematical events contributes to mathematics teacher educators' TPACK (Technological Pedagogical Content Knowledge). *International Journal of Education in Mathematics, Science and Technology, 13*(4), 895-913. https://files.eric.ed.gov/fulltext/EJ1475751.pdf

Busuttil, L., & Calleja, J. (2025). Teachers' beliefs and practices about the potential of ChatGPT in teaching Mathematics in secondary schools. *Digital Experiences in Mathematics Education, 11*(1*)*, 140-166. https://doi.org/10.1007/s40751-024-00168-3

Canonigo, A. M. (2024). Levering AI to enhance students' conceptual understanding and confidence in mathematics. *Journal of computer assisted learning, 40*(6), 3215-3229. https://doi.org/10.1111/jcal.13065

Cullen, C. J., & Hertel, J. T. (2023). Improving our criticism of technology in mathematics education. *Mathematics Teacher: Learning and Teaching PK-12*, *116*(5), 314-321. https://doi.org/10.5951/mtlt.2023.0004

Davis, J. D. (2025). Evolving techniques and emerging schemes: Prospective teachers' transformation of ChatGPT. *International Journal of Science and Mathematics Education, 23*, 2181–2198. https://doi.org/10.1007/s10763-024-10533-8

Fardian, D., Suryadi, D., Prabawanto, S., & Jupri, A. (2025). Integrating Chat-GPT in the classroom: A study on linear algebra learning in higher education. *International Journal of Information and Education Technology, 15*(4). https://doi.org/10.18178/ijiet.2025.15.4.2279

Gurl, T. J., Markinson, M. P., & Artzt, A. F. (2025). Using ChatGPT as a lesson planning assistant with preservice secondary mathematics teachers. *Digital Experiences in Mathematics Education, 11*(1), 114-139. https://doi.org/10.1007/s40751-024-00162-9

Goos, M. (2005). A sociocultural analysis of the development of pre-service and beginning teachers' pedagogical identities as users of technology. *Journal of Mathematics Teacher Education, 8*(1), 35–59. https://doi.org/10.1007/s10857-005-0457-0

Goos, M., Galbraith, P., Renshaw, P., & Geiger, V. (2003). Perspectives on technology-mediated learning in secondary school mathematics classrooms. *Journal of Mathematical Behavior, 22*(1), 73–89. https://doi.org/10.1016/S0732-3123(03)00005-1

Henkel, O., Horne-Robinson, H., Dyshel, M., Thompson, G., Abboud, R., Ch, N. A. N., ... & Vanacore, K. (2025). Learning to love LLMs for answer interpretation: Chain-of-thought prompting and the AMMORE dataset. *Journal of Learning Analytics, 12*(1), 50–64. https://doi.org/10.18608/jla.2025.8621

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.* https://doi.org/10.1111/ssm.18336

Karaman, M. R., & Goksu, I. (2024). Are lesson plans created by ChatGPT more effective? An experimental study. *International Journal of Technology in Education, 7*(1), 107-127. https://doi.org/10.46328/ijte.607

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *bmj*, 372. https://doi.org/10.1136/bmj.n71

Park, J., & Choo, S. (2025). Generative AI prompt engineering for educators: Practical strategies. *Journal of Special Education Technology*, *40*(3), 411-417. https://doi.org/10.1177/01626434241298954

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

Podani, J., & Schmera, D. (2011). A new conceptual and methodological framework for exploring and explaining pattern in presence – absence data. *Oikos*, *120*(11), 1625-1638. https://doi.org/https://doi.org/10.1111/j.1600-0706.2011.19451.x

Ramprakash, B., Nithyakala, G., Bhumika, K., & Avanthika, S. (2024). Comparing traditional instructional methods to ChatGPT: A comprehensive analysis. *Journal of Engineering Education Transformations*, (37), 612-620. https://journaleet.in/index.php/jeet/article/view/2477/2297

Segal, R., & Biton, Y. (2024). The Contribution That utilizing Generative AI for problem posing makes to pre-service high school mathematics teachers' TPACK. *International Journal of Education in Mathematics, Science and Technology, 12*(6), 1559-1582. https://files.eric.ed.gov/fulltext/EJ1465780.pdf

Slobodsky, P., & Durcheva, M. (2025). Using ChatGPT errors to encourage student exploration and self-learning in mathematics. *International Journal of Mathematical Education in Science and Technology,* 1-25. https://doi.org/10.1080/0020739X.2025.2505200

Taani, O., & Alabidi, S. (2024). ChatGPT in education: Benefits and challenges of ChatGPT for mathematics and science teaching practices. *International Journal of Mathematical Education in Science and Technology,* 1748-1777. https://doi.org/10.1080/0020739X.2024.2357341

Torres-Peña, R. C., Peña-González, D., Chacuto-López, E., Ariza, E. A., & Vergara, D. (2024). Updating calculus teaching with AI: A classroom experience. *Education Sciences, 14*(9), 1019. https://doi.org/10.3390/educsci14091019

Wahba, F., Ajlouni, A. O., & Abumosa, M. A. (2024). The impact of ChatGPT-based learning statistics on undergraduates' statistical reasoning and attitudes toward statistics. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(7), em2468. https://doi.org/10.29333/ejmste/14726

Walkington, C. (2025). The implications of generative artificial intelligence for mathematics education. *School Science and Mathematics*. https://doi.org/10.1111/ssm.18356

Zhuang, Y., & Zhang, S. (2025). Pre-service mathematics teachers' perceptions of using GenAI for practicing teacher questioning: A semester-long study. *EURASIA Journal of Mathematics, Science and Technology Education, 21*(9), em2689. https://doi.org/10.29333/ejmste/16764

Note: Due to the MERGA page limit, we do not include the full list of the reviewed 43 empirical studies in the paper. The complete list is available in the Supplementary Materials.