

Effectiveness of Intelligent Tutoring Systems, Adaptive Learning, and Artificial Intelligence-Based Technologies in Mathematics Learning: An Umbrella Review of Systematic Reviews and Meta-Analyses

Ferdian Yunizar*, Suprih Widodo¹, Ulva Elviani²

¹ Information systems and technology education, [Universitas Pendidikan Indonesia], [Purwakarta, Indonesia]

² [Information systems and technology education, [Universitas Pendidikan Indonesia], [Purwakarta, Indonesia]

³ Information systems and technology education, [Universitas Pendidikan Indonesia], [Purwakarta, Indonesia]

* Corresponding author. Email: [ferdianyunizar24@upi.edu]; Tel: [+6285156258644]

Abstract

This umbrella review synthesizes evidence from systematic reviews and meta-analyses on the effectiveness of intelligent tutoring systems, adaptive learning platforms, and artificial intelligence-based technologies in mathematics learning. A semi-automated search was conducted on 18 December 2025 using an indexed academic search engine, complemented by forward-backward citation chasing. Eligible reviews were screened and appraised using a proxy version of AMSTAR 2 focused on critical domains. Across the included reviews, technology-enhanced and adaptive interventions generally showed small-to-moderate positive effects on mathematics achievement, with more consistent benefits when feedback was immediate, personalization was meaningful, and implementation was supported by teachers and well-aligned tasks. Evidence for non-cognitive outcomes (e.g., motivation, self-efficacy, engagement) was positive but less consistently reported. Heterogeneity across contexts and substantial overlap among primary studies remain major challenges. The review highlights implementation conditions, recurring moderators, and priority gaps to strengthen future research and reporting in mathematics education.

Keywords: umbrella review; intelligent tutoring systems; adaptive learning; artificial intelligence; mathematics education

INTRODUCTION

Digital transformation in education has accelerated the adoption of technologies that do not merely deliver content but also model student learning, diagnose misconceptions, and provide sustained problem-solving support. Within this landscape, intelligent tutoring systems are computer-based learning environments designed to provide step-by-step guidance, feedback, and adaptive scaffolding. Adaptive learning platforms and other artificial intelligence-based educational technologies extend these ideas by personalizing learning paths, recommending tasks, and analyzing learner data in real time, all of which are particularly relevant to mathematics learning where misconceptions and procedural fluency interact with conceptual understanding.

Despite rapid uptake, empirical findings about the effectiveness of these technologies in mathematics are not always consistent. Existing meta-analyses and systematic reviews vary in scope (grade levels, tools, outcomes) and methodological quality, creating fragmented messages for practitioners and policy makers. Therefore, this study conducts an umbrella review (an overview of reviews) to consolidate the evidence base, summarize effect directions and magnitudes where available, identify recurring moderators, and appraise the strength of the underlying review evidence.

Research questions

RQ1. What are the direction and magnitude of effects of intelligent tutoring systems, adaptive learning platforms, and artificial intelligence-based interventions on mathematics learning outcomes?

RQ2. Which non-cognitive outcomes (e.g., motivation, self-efficacy, engagement) are most consistently influenced?

RQ3. Which moderators are most frequently reported to explain variation (e.g., education level, learner characteristics, intervention duration, teacher involvement, platform design)?

RQ4. How strong is the methodological quality of the included reviews (using a proxy AMSTAR 2 approach), and how does quality relate to reported effects?

RQ5. What evidence gaps are most salient, and what should be prioritized in the future research agenda?

METHOD

Design and reporting guidance

This study is an umbrella review (overview of reviews) that synthesizes findings from systematic reviews and meta-analyses. Reporting was guided by PRISMA 2020 where applicable, with additional

considerations for overviews of reviews to ensure transparent screening, appraisal, and synthesis decisions.

Data sources and search strategy

The literature search was conducted on 18 December 2025 using an indexed academic search engine (Semantic Scholar) as a primary starting point, complemented by forward-backward citation chasing to improve coverage. Searches were restricted to 2010-2025 to capture contemporary intelligent tutoring and adaptive technologies. Search strings combined (a) technology terms (e.g., “intelligent tutoring system”, “adaptive learning”, “AI tutoring”) with (b) mathematics terms (e.g., “mathematics”, “numeracy”, “algebra”) and (c) review terms (e.g., “systematic review”, “meta-analysis”, “overview of reviews”).

Eligibility criteria

The unit of analysis was a secondary synthesis article (systematic review, meta-analysis, scoping review, evidence map, or other structured review) that examined intelligent tutoring systems, adaptive learning, or artificial intelligence-enabled learning technologies in mathematics education. Reviews were included if they (1) reported outcomes relevant to mathematics learning (achievement, performance, procedural/conceptual outcomes, or closely related affective/behavioral outcomes in mathematics contexts), (2) described a systematic search and selection process, and (3) provided sufficient reporting to extract technology category, context, and main findings. Non-mathematics-only reviews, narrative commentaries without transparent methods, and duplicate records were excluded.

Study selection and data extraction

Records were deduplicated using DOI matching; for records without DOI, titles were normalized and compared to identify duplicates. Screening proceeded in three stages: title/abstract screening, full-text eligibility assessment, and extraction. Extracted items included review type, time window, education level, country/context where available, technology category, primary outcome(s), effect size summaries (if meta-analysis), and moderator analyses.

Prioritization of reviews synthesized (n = 40)

To balance breadth and depth, the synthesis prioritized 40 reviews using rule-based criteria emphasizing recency, methodological rigor, and relevance to mathematics education. When multiple reviews covered the same intervention family or heavily overlapped in primary studies, the most comprehensive and best-reported review was prioritized to reduce redundancy in the narrative synthesis.

Quality appraisal and overlap considerations

Methodological quality was appraised using a proxy version of AMSTAR 2 focusing on critical domains: protocol/registration, comprehensiveness of search, appropriateness of risk-of-bias appraisal, handling of heterogeneity/synthesis, and publication bias assessment. Confidence ratings (high/moderate/low/critically low) were assigned based on the pattern of critical weaknesses. Overlap among primary studies across reviews was considered qualitatively; where complete lists of primary studies were unavailable, overlap metrics (e.g., CCA) could not be computed and were noted as a limitation.

Study flow (PRISMA-like)

The study flow is summarized using a PRISMA-like diagram tailored for overviews of reviews (identification, screening, eligibility, and inclusion) (see Figure 1).

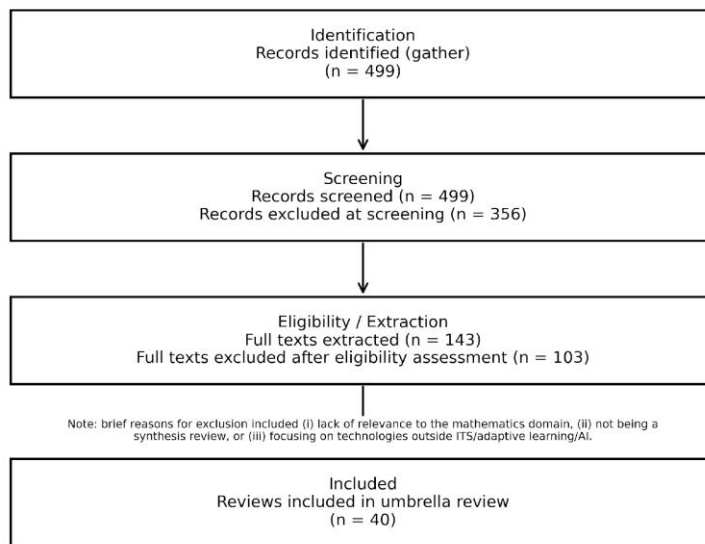


Figure 1. PRISMA-like study selection flow diagram.

RESULTS AND DISCUSSION

Evidence map and publication trend

Across the included sample, the review literature has grown rapidly since 2020, reflecting increased adoption of adaptive technologies and artificial intelligence-enabled tools in mathematics education. The evidence map indicates that intelligent tutoring systems and adaptive platforms dominate the synthesis literature, while studies explicitly evaluating newer generative or conversational systems remain limited. Figure 2 summarizes the technology-category evidence map, and Figure 3 shows the year-by-year publication trend.

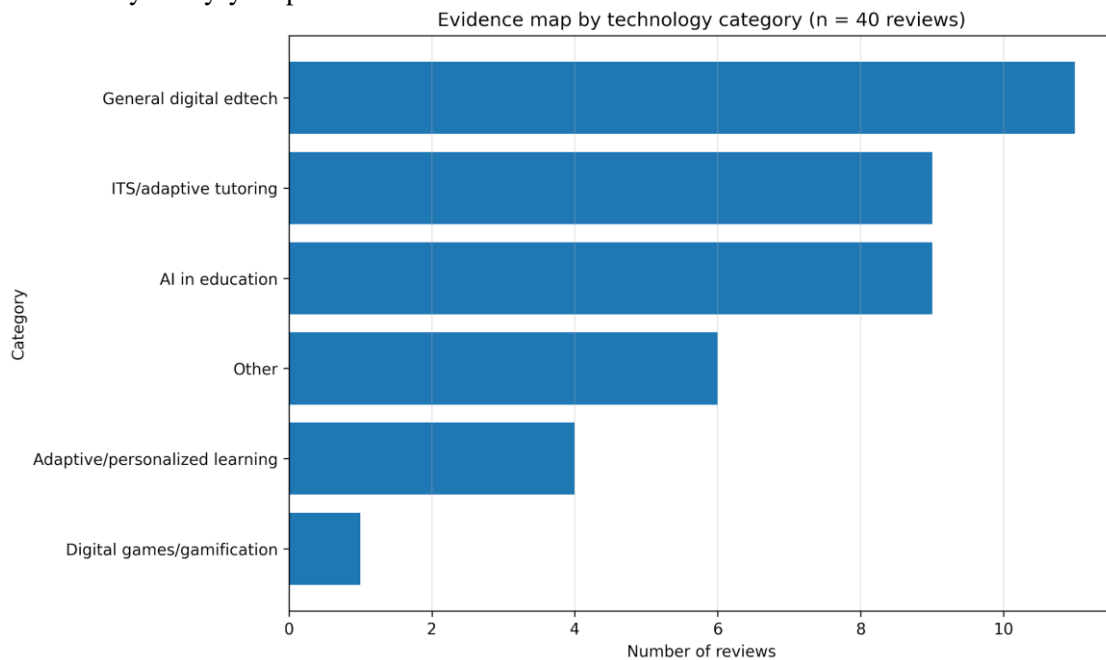


Figure 2. Evidence map by technology category across the 40 included reviews.

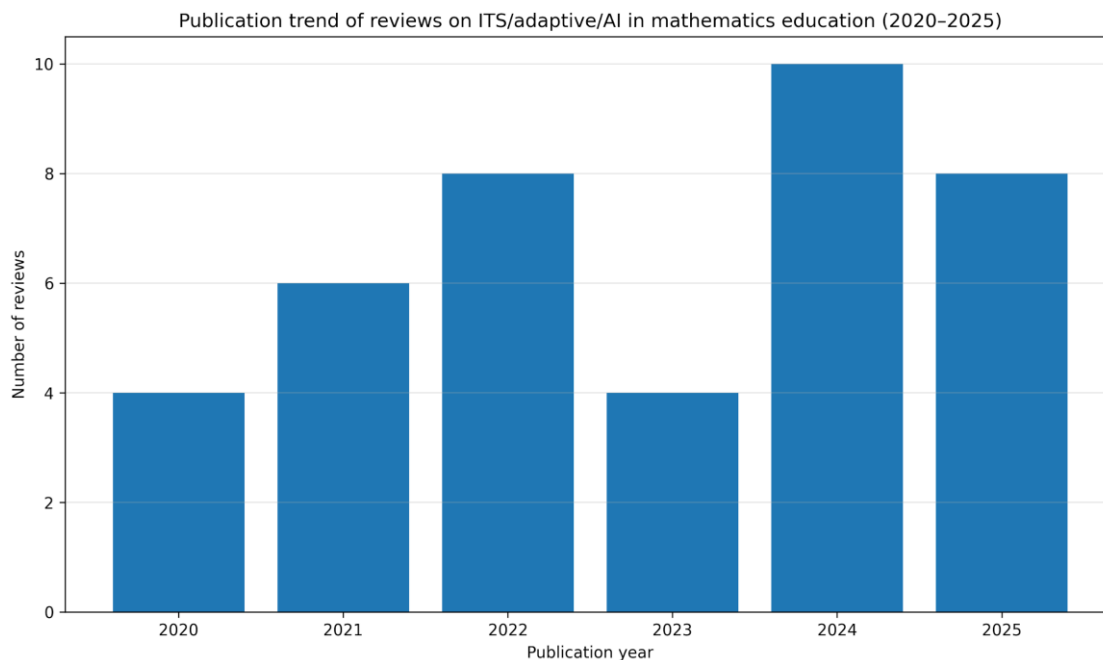


Figure 3. Publication trend of reviews on intelligent tutoring systems/adaptive learning/artificial intelligence in mathematics (2020–2025).

Characteristics of included reviews

Table 1 summarizes key characteristics of the 40 included reviews, including review type, education level, number of primary studies (if reported), DOI, and focal outcomes.

Table 1. Characteristics of Included Reviews (n = 40)

Author (year)	Review type	Education level	Number of primary studies	DOI	Main focus/outcomes
Sun (2021)	Meta-analysis	K-12, higher education	33	https://doi.org/10.1080/19477503.2021.1926194	ITS/adaptive tutoring; achievement/performance
Hahs-Vaughn (2024a)	Evidence gap map (systematic mapping)	Undergraduate mathematics	42	https://doi.org/10.1109/fie61694.2024.10893331	Adaptive/personalized learning; learning outcomes (general)
Olmo-Muñoz (2022)	Systematic review	Kindergarten, elementary school, secondary school, higher education, pre-service teachers, in-service teachers, students with special needs	174	https://doi.org/10.1007/s11858-022-01396-w	ITS/adaptive tutoring; achievement/performance, problem-solving
Hahs-Vaughn (2024b)	Protokol meta-analysis	Undergraduate		https://doi.org/10.23865/njsre.v2.6339	Adaptive/personalized learning; learning outcomes (general)
Son (2024)	Systematic review	Elementary school, middle school, high school, colleges, teacher education	58	https://doi.org/10.3390/computers13100270	ITS/adaptive tutoring; achievement/performance
Mredula (2024)	Systematic review	Not mentioned		https://doi.org/10.47392/irjaeh.2024.0268	AI in education; learning outcomes (general)
Maanu	Meta-analysis	Varied/not	7	https://doi.org/10.47191/ijmcr/v13i9.09	AI in education;

(2025)		specified				achievement/performance, motivation/engagement
Benavides-Varela (2020)	Meta-analysis	Primary school, Preschool	15	https://doi.org/10.1016/j.compedu.2020.103953		General digital edtech; achievement/performance
Hwang (2022)	Meta-analysis	Elementary	21	https://doi.org/10.3390/su142013185		AI in education; achievement/performance
Létourneau (2025)	Systematic review	K-12	28	https://doi.org/10.1038/s41539-025-00320-7		ITS/adaptive tutoring; achievement/performance
Hwang (2021)	Bibliometric + systematic review	Junior high school, elementary school, higher education	43	https://doi.org/10.3390/math9060584		AI in education; achievement/performance, motivation/engagement
Major (2021)	Meta-analysis	Primary and early secondary education (ages 6–15)	16	https://doi.org/10.1111/bjet.13116		Adaptive/personalized learning; learning outcomes (general)
Tamur (2023)	Meta-analysis	Varied/not specified	28	https://doi.org/10.22460/infinity.v12i1.p101-116		General digital edtech; achievement/performance
Huang (2025)	Meta-analysis	Varied/not specified	30	https://doi.org/10.4018/ijdet.368420		ITS/adaptive tutoring; achievement/performance, problem-solving
Feng (2021)	Systematic review	Varied/not specified	22	https://doi.org/10.1109/fie49875.2021.9637240		ITS/adaptive tutoring; problem-solving
Awang (2025)	Systematic review	Varied/not specified	32	https://doi.org/10.29333/iejme/16006		AI in education; achievement/performance
Borah (2024)	Systematic review	Not explicitly mentioned (broad scope likely including K-12 and higher education)	60	https://doi.org/10.71097/ijstat.v15.i4.1190		AI in education; achievement/performance, problem-solving
Panqueban (2024)	Systematic review	Not specified; no evidence at early childhood education level	29	https://doi.org/10.15359/ru.38-1.20		Other; learning outcomes (general)
Ran (2021)	Meta-analysis	K-12	31	https://doi.org/10.1177/0735633120952063		General digital edtech; achievement/performance, problem-solving
Wang (2024)	Meta-analysis	Not specified	45	https://doi.org/10.1177/07356331241240459		Adaptive/personalized learning; cognitive/metacognitive
Ran (2022)	Meta-analysis	K-12 classrooms	77	https://doi.org/10.1111/jcal.12611		General digital edtech; achievement/performance, problem-solving
Myers (2021)	Meta-analysis	Secondary school	45	https://doi.org/10.1111/ldrj.12244		Other; achievement/performance
Meylani (2025)	Systematic review	Not mentioned	25	https://doi.org/10.2478/dcse-2025-0014		AI in education; motivation/engagement
Hillmayr (2020)	Meta-analysis	Secondary school (grade levels 5–13)	92	https://doi.org/10.1016/j.compedu.2020.103897		General digital edtech; achievement/performance
Subramaniam (2022)	Systematic review	Not explicitly	34	https://doi.org/10.18844/cjes.v17i6.7494		Other; computational thinking

		mentioned (various levels implied)			
Liu (2025)	Systematic review	General educational settings (not specified)	86	https://doi.org/10.48550/arxiv.2503.09748	ITS/adaptive tutoring; motivation/engagement
Moore (2022)	Systematic review	K-12	12	https://doi.org/10.1080/15391523.2022.2148785	AI in education; motivation/engagement
Opesemowo (2024)	Systematic review	Elementary, secondary, university	10	https://doi.org/10.29333/ejmste/14762	AI in education; motivation/engagement
Williams (2022)	Meta-analysis	Varied/not specified	191	https://doi.org/10.1080/19345747.2021.2009072	Other; learning outcomes (general)
Knogler (2022)	Tinjauan meta-analysis	Secondary	41	https://doi.org/10.3389/fpsyg.2022.873995	Other; motivation/engagement
Lecca (2024)	Systematic review	Varied/not specified	55	https://doi.org/10.61707/j5c9ts33	Digital games/gamification; motivation/engagement
Rakes (2020)	Meta-analysis	Not explicitly mentioned	100	https://doi.org/10.1016/j.edurev.2020.100337	General digital edtech; achievement/performance
Kurniati (2023)	Systematic review	Not explicitly mentioned	77	https://doi.org/10.24252/10.24252/mapan.2023v11n2a12	General digital edtech; problem-solving, motivation/engagement
Zhumashev (2025)	Systematic review	General Education, Programming (likely K-12 and higher education)	26	https://doi.org/10.53894/ijirss.v8i6.10048	ITS/adaptive tutoring; learning outcomes (general)
Carvajal (2025)	Systematic review	Not mentioned		https://doi.org/10.60096/fundacionavanza/12642025	General digital edtech; learning outcomes (general)
Saat (2024)	Systematic review	Elementary school through college	13	https://doi.org/10.6007/ijarped/v13-i4/22956	General digital edtech; achievement/performance, problem-solving
(谢晨) (2020)	Meta-analysis	Elementary (Grades 1-6 in China or K-5 in the U.S.), Secondary (Grade 7-12 in China or 6-12 in the U.S.)		https://doi.org/10.1177/2096531120944410	General digital edtech; achievement/performance
Liu (2023)	Meta-analysis	Not mentioned	27	https://doi.org/10.1177/01626434231180579	Other; achievement/performance
Çavus (2022)	Meta-analysis	Various (not explicitly detailed in quotes)	98	https://doi.org/10.17275/per.22.45.9.2	General digital edtech; achievement/performance
Tlili (2023)	Meta-analysis	Primary education	14	https://doi.org/10.1109/icalt58122.2023.00020	ITS/adaptive tutoring; learning outcomes (general)

Synthesis of effectiveness evidence

To improve readability, quantitative estimates reported by meta-analyses are summarized in Table 2. Overall, the meta-analytic evidence generally indicates small-to-moderate positive effects of intelligent tutoring and adaptive learning technologies on mathematics achievement and performance, although effect sizes vary substantially by platform, learner population, and implementation conditions. Some reviews also report positive trends for non-cognitive outcomes such as motivation and engagement, but these outcomes are less consistently measured and reported across reviews.

Table 2. Summary of Quantitative Estimates Reported by Meta-Analyses (as extracted from the reviews)

Review (author, year)	k (primary studies, if reported)	Focus/outcomes (brief)	Metric & estimate (as reported)	Interpretation notes (based on reporting review)	DOI
Sun (2021)	33	ITS/adaptive tutoring; achievement/performance	Hedges' $g = 0.05$ (95% CI [-0.01, 0.20]) for ALEKS vs traditional instruction: $g = 0.43$ (95% CI [0.02, 0.83]) when ALEKS as a supplement	Overall effect is very small; stronger effects were reported under specific subgroup conditions (as noted in the review).	https://doi.org/10.1080/19477503.2021.1926194
Huang (2025)	30	ITS/adaptive tutoring; achievement/performance, problem-solving	$g = 0.86$ (overall)	A large positive effect was reported; model details/CI were not fully specified in the extracted summary.	https://doi.org/10.4018/ijdet.368420
Ran (2021b)	77	General digital edtech; achievement/performance, problem-solving	Average Hedges' $g = 0.23$	The review reports a small but statistically significant effect.	https://doi.org/10.1111/jcal.12611
Major (2021)	16	Adaptive/personalized learning; learning outcomes (general)	$ES = 0.18$ ($p = 0.001$)	The effect-size metric was not fully specified in the extracted summary; effects were small and significant.	https://doi.org/10.1111/bjet.13116
Liu (2023)	27	Other; achievement/performance	Tau-U = 0.98 (95% CI 0.92–1.00)	Tau-U is common for particular single-case designs and is not directly comparable to g/d .	https://doi.org/10.1177/01626434231180579
Knogler (2022)	41	Other; motivation/engagement	Range ES = 0.01 hingga 1.3; 80% ES ≥ 0.2 ; 54% ES ≥ 0.4	Indicates substantial variation; dominant outcomes were general affective/cognitive measures.	https://doi.org/10.3389/fpsyg.2022.873995
Tlili (2023)	14	ITS/adaptive tutoring; learning outcomes (general)	Not reported (quantitative estimate not available in the extracted summary).	Treat as qualitative/conceptual evidence only.	https://doi.org/10.1109/icalt58122.2023.00020
Wang (2024)	45	Adaptive/personalized learning; cognitive/metacognitive	$g = 0.70$	Moderate-to-large effects for cognitive outcomes compared with non-adaptive conditions.	https://doi.org/10.1177/07356331241240459
Williams (2022)	191	Other; learning outcomes (general)	Average effect size = 0.31 (for mathematics)	The metric was not specified in the extracted summary; effects were small-to-moderate.	https://doi.org/10.1080/19345747.2021.2009072
Ran (2021a)	31	General digital edtech; achievement/performance, problem-solving	$\bar{d} = 0.56$ (overall, low-	Moderate-to-large effects in certain subgroups;	https://doi.org/10.1177/0735633120952063

			performing students); problem-solving system $\bar{d} = 0.86$; tutorial $\bar{d} = 0.58$; computerized practice $\bar{d} = 0.23$	dependent on intervention type.	
Hillmayr (2020)	92	General digital edtech; achievement/performance	$g = 0.65$ ($p < .001$)	Moderate effect; heterogeneity and intervention variation should be noted.	https://doi.org/10.1016/j.compedu.2020.103897
Myers (2021)	45	Other; achievement/performance	$g = 0.52$	Moderate effect (the review summary describes it as 'moderately large').	https://doi.org/10.1111/ldrp.12244
Rakes (2020)	100	General digital edtech; achievement/performance	$\gamma_{Ach} = 0.113$ ($p = .004$); $\gamma_{Ormn} = 0.125$ ($p < .001$)	Gamma/coefficient metric; small effects but consistent for certain outcomes.	https://doi.org/10.1016/j.edurev.2020.100337
Benavides-Varela (2020)	15	General digital edtech; achievement/performance	Mean ES = 0.55	Moderate effect; types of digital interventions varied.	https://doi.org/10.1016/j.compedu.2020.103953
Maanu (2025)	7	AI in education; achievement/performance, motivation/engagement	$g = 0.603$ ($p < 0.001$); 95% CI 0.393–0.813)	Moderate-to-large effect; the number of primary studies was relatively small (low k).	https://doi.org/10.47191/ijmcr/v13i9.09

Note. k indicates the number of primary studies reported by each review (if available) and does not necessarily represent exclusive studies due to overlap across reviews. Estimates are presented as reported in the source reviews.

Consistent moderators

Across reviews, recurring moderators include education level, baseline achievement, intervention duration, the presence of meaningful personalization, and the quality of instructional feedback. Interventions tended to yield stronger effects when they provided immediate, informative feedback; adapted practice based on learner responses; and were integrated into classroom routines with teacher support. Conversely, weaker or null effects were more frequently reported when technology use was brief, poorly aligned with learning objectives, or implemented without adequate guidance.

Strength of evidence and methodological issues

Table 3 presents the methodological appraisal results based on a proxy AMSTAR 2 approach. Confidence in the evidence varies across reviews, with common weaknesses including limited protocol registration, incomplete reporting of search coverage, and inconsistent handling of publication bias. In addition, overlap among primary studies across reviews remains a key threat to interpretation and motivates cautious synthesis.

Table 3. Methodological Quality Appraisal (proxy AMSTAR 2) for the Included Reviews

Review	Protocol/registration	Comprehensive search	Risk of bias/quality appraisal	Heterogeneity/synthesis methods	Publication bias	Confidence (proxy)
Hahs-Vaughn (2024a)	Yes	No clear	Yes	Yes	NA	Moderate
Saat (2024)	No clear	Yes	Yes	Yes	NA	Moderate
Sun (2021)	No clear	No clear	Yes	Yes	No clear	Low (insufficient reporting)
Liu (2025)	No clear	Yes	Yes	Yes	NA	Moderate
Hahs-Vaughn (2024b)	Yes	NA	NA	NA	NA	NA (protokol)
(谢晨) (2020)	No clear	No clear	Yes	Yes	No clear	Low (insufficient reporting)

						reporting)
Hwang (2021)	No clear	No clear	Yes	Yes	NA	Moderate
Çavus (2022)	No clear	Yes	Yes	Yes	Yes	Moderate
Feng (2021)	No clear	No clear	Yes	Yes	NA	Moderate
Huang (2025)	No clear	Yes	Yes	Yes	No clear	Moderate
Ran (2022)	No clear	No clear	Yes	Yes	No clear	Low (insufficient reporting)
Olmo-Muñoz (2022)	No clear	No clear	Yes	Yes	NA	Moderate
Mredula (2024)	No clear	No clear	Yes	Yes	NA	Moderate
Major (2021)	No clear	Yes	Yes	Yes	No clear	Moderate
Liu (2023)	No clear	No clear	Yes	Yes	No clear	Low (insufficient reporting)
Knogler (2022)	No clear	No clear	Yes	Yes	Yes	Moderate
Tlili (2023)	No clear	No clear	Yes	Yes	No clear	Low (insufficient reporting)
Subramaniam (2022)	No clear	Yes	Yes	Yes	NA	Moderate
Kurniati (2023)	No clear	Yes	Yes	Yes	NA	Moderate
Meylani (2025)	No clear	No clear	Yes	Yes	NA	Moderate
Opesemowo (2024)	No clear	Yes	Yes	Yes	NA	Moderate
Wang (2024)	No clear	No clear	Yes	Yes	No clear	Low (insufficient reporting)
Zhumashev (2025)	No clear	Yes	Yes	Yes	NA	Moderate
Panqueban (2024)	No clear	No clear	Yes	Yes	NA	Moderate
Awang (2025)	No clear	No clear	Yes	Yes	NA	Moderate
Hwang (2022)	No clear	Yes	Yes	Yes	Yes	Moderate
Williams (2022)	No clear	Yes	Yes	Yes	No clear	Moderate
Létourneau (2025)	No clear	Yes	Yes	Yes	NA	Moderate
Son (2024)	No clear	No clear	Yes	Yes	NA	Moderate
Borah (2024)	No clear	No clear	Yes	Yes	NA	Moderate
Ran (2021)	No clear	No clear	Yes	Yes	No clear	Low (insufficient reporting)
Moore (2022)	No clear	Yes	Yes	Yes	NA	Moderate
Lecca (2024)	No clear	Yes	Yes	Yes	NA	Moderate
Hillmayr (2020)	No clear	Yes	Yes	Yes	Yes	High
Myers (2021)	No clear	No clear	Yes	Yes	No clear	Low (insufficient reporting)
Tamur (2023)	No clear	Yes	Yes	Yes	Yes	Moderate
Rakes (2020)	No clear	No clear	Yes	Yes	Yes	Moderate
Benavides-Varela (2020)	No clear	No clear	Yes	Yes	Yes	Moderate
Carvajal (2025)	No clear	Yes	No clear	No clear	NA	Low (insufficient reporting)
Maanu (2025)	No clear	No clear	Yes	Yes	Yes	Moderate

Implications for implementation in Indonesia

For Indonesian contexts, the synthesized evidence suggests that intelligent tutoring and adaptive platforms can be beneficial when they complement teacher instruction rather than replace it. Priority implementation conditions include (1) alignment of tasks with the curriculum, (2) teacher professional development on interpreting learning analytics and orchestrating blended learning, (3) equitable access to devices and connectivity, and (4) careful attention to data privacy and ethical use of learner data.

Pilot implementations should include monitoring of both learning gains and student experience to avoid widening existing learning gaps.

Research and reporting agenda

Future research should improve transparency and comparability by: (1) preregistering review protocols, (2) reporting complete search strings and selection decisions, (3) providing complete lists of included primary studies to enable overlap quantification (e.g., CCA), (4) consistently measuring both cognitive and non-cognitive outcomes, and (5) testing moderators that reflect real implementation constraints (teacher involvement, dosage, and platform features). In addition, as newer conversational and generative systems enter classrooms, evaluative studies should clearly define the intervention, data handling, and safeguards.

Limitations and contribution

This umbrella review is limited by incomplete reporting in some source reviews (e.g., missing lists of primary studies), which constrained overlap analysis and fully standardized quality appraisal. Nevertheless, the manuscript contributes a consolidated evidence map, a structured synthesis of reported effect sizes, and a pragmatic appraisal of review quality to support decision making in mathematics education.

CONCLUSION

Across the 40 included reviews, the synthesized evidence indicates that intelligent tutoring systems, adaptive learning platforms, and related AI-enabled technologies generally produce positive effects on mathematics achievement and performance, most often in the small-to-moderate range. Benefits are most consistently reported when personalization is instructionally meaningful, feedback is timely and informative, and technology use is embedded in classroom routines with active teacher support. In contrast, evidence for newer conversational or generative systems remains comparatively limited and heterogeneous, highlighting the need for cautious claims about broad effectiveness.

From an implementation perspective particularly for Indonesian school and higher-education settings, the review suggests prioritizing (1) curricular alignment of digital tasks, (2) teacher professional development focused on interpreting learning analytics and orchestrating blended learning, (3) equitable access to devices and connectivity, and (4) data governance that protects student privacy and supports ethical use of learner data. Pilot deployments should track both cognitive outcomes (achievement/problem solving) and non-cognitive outcomes (motivation/engagement), because the latter are reported less consistently across reviews yet are central to sustained use and learning gains.

Methodologically, the confidence of this umbrella synthesis is constrained by incomplete reporting in a subset of source reviews and by potential overlap of primary studies across reviews. Strengthening future evidence will require preregistered protocols, transparent and reproducible search and screening logs, standardized outcome definitions, and clearer reporting of moderators (e.g., dosage, learner level, platform features, and teacher involvement). Taken together, the findings support targeted, well-supported adoption of adaptive/ITS technologies in mathematics education while emphasizing rigorous evaluation and responsible implementation.

REFERENCES

- Awang, L. A., Yusop, F., & Danaee, M. (2025). Current practices and future direction of artificial intelligence in mathematics education: A systematic review. *International Electronic Journal of Mathematics Education*. <https://doi.org/10.29333/iejme/16006>
- Benavides-Varela, S., Zandonella Callegher, C., Fagiolini, B., Leo, I., Altoó, G., & Lucangeli, D. (2020). Effectiveness of digital-based interventions for children with mathematical learning difficulties: A meta-analysis. *Computers & Education*, 157, 103953. <https://doi.org/10.1016/j.compedu.2020.103953>
- Borah, P., & Borah, A. C. (2024). A review of use of artificial intelligence in teaching and learning of mathematics. *International Journal for Sciences and Technology*. <https://doi.org/10.71097/ijst.v15.i4.1190>
- del Olmo-Muñoz, J., González-Calero, J. A., Diago, P., Arnau, D., & Arevalillo-Herráez, M. (2022). Intelligent tutoring systems for word problem solving in COVID-19 days: Could they have been (part of) the solution? *ZDM—Mathematics Education*, 54, 1381–1397. <https://doi.org/10.1007/s11858-022-01396-w>
- Feng, S., Magana, A. J., & Kao, D. (2021). A systematic review of literature on the effectiveness of intelligent tutoring systems in STEM. In *Proceedings of the Frontiers in Education Conference*. <https://doi.org/10.1109/fie49875.2021.9637240>

- Gates, M., Gates, A., Pieper, D., Fernandes, R. M., Tricco, A. C., Brennan, S. E., Li, T., Pollock, M., Lunny, C., Sepúlveda, D., McKenzie, J. E., Scott, S. D., Robinson, K. A., Matthias, K., Bougioukas, K. I., Fusar-Poli, P., Whiting, P., Moss, S. J., & Hartling, L. (2022). Reporting guideline for overviews of reviews of healthcare interventions: Development of the PRIOR statement. *BMJ*, 378, e070849. <https://doi.org/10.1136/bmj-2022-070849>
- Hahs-Vaughn, D., Moskal, P. D., Teixeira, K., Muhs, T., Bishop, C., Farotimi, O., & Carassas, C. (2024a). WIP: An evidence gap map analysis of personalized adaptive learning in undergraduate mathematics. In *Proceedings of the Frontiers in Education Conference*. <https://doi.org/10.1109/fie61694.2024.10893331>
- Hahs-Vaughn, D., Moskal, P. D., Teixeira, K., Muhs, T., Danaee, M., Farotimi, O., & Carassas, C. (2024b). Protocol for meta-analyses examining personalized adaptive learning in undergraduate mathematics. *Nordic Journal of Systematic Reviews in Education*, 2. <https://doi.org/10.23865/njsre.v2.6339>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S., & Reiss, K. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Huang, X., Xu, W., & Liu, R. (2025). Effects of intelligent tutoring systems on educational outcomes: Evidence from a comprehensive analysis. *International Journal of Distance Education Technologies*. <https://doi.org/10.4018/ijdet.368420>
- 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), 584. <https://doi.org/10.3390/math9060584>
- Hwang, S. (2022). Examining the effects of artificial intelligence on elementary students' mathematics achievement: A meta-analysis. *Sustainability*, 14(20), 13185. <https://doi.org/10.3390/su142013185>
- Knogler, M., Hetmanek, A., & Seidel, T. (2022). Determining an evidence base for particular fields of educational practice: A systematic review of meta-analyses on effective mathematics and science teaching. *Frontiers in Psychology*, 13, 873995. <https://doi.org/10.3389/fpsyg.2022.873995>
- Kurniati, R., Sugiarto, S., & Lestari. (2023). Systematic literature review (SLR) of technology in mathematics learning during the last decade. *MaPan: Jurnal Matematika dan Pembelajaran*, 11(2), 386–402. <https://doi.org/10.24252/mapan.2023v11n2a12>
- Liu, D., Mao, Y., Cai, W., Lei, Q., Kang, R., & Zeng, Y. (2023). Meta-analysis of tablet-mediated interventions to teach mathematics for individuals with autism spectrum disorder and/or intellectual disability. *Journal of Special Education Technology*. <https://doi.org/10.1177/01626434231180579>
- Liu, V., Latif, E., & Zhai, X. (2025). Advancing education through tutoring systems: A systematic literature review. *arXiv*. <https://doi.org/10.48550/arXiv.2503.09748>
- Lozano Carvajal, I., & Pérez Marín, D. (2025). Pedagogical agents in advancing computer science and mathematics education: A systematic review of effective practices and outcomes. *Fundación Avanza*. <https://doi.org/10.60096/fundacionavanza/12642025>
- Lunny, C., Pieper, D., Thabet, P., & Kanji, S. (2021). Managing overlap of primary study results across systematic reviews: Practical considerations for authors of overviews of reviews. *BMC Medical Research Methodology*, 21(1), 140. <https://doi.org/10.1186/s12874-021-01269-y>
- Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education. *npj Science of Learning*. <https://doi.org/10.1038/s41539-025-00320-7>
- Maanu, V., Boateng, F., & Larbi, E. (2025). Effect of artificial intelligence on mathematics performance: A meta-analysis. *International Journal of Mathematics and Computer Research*. <https://doi.org/10.47191/ijmcr/v13i9.09>
- Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low- and middle-income countries: A meta-analysis. *British Journal of Educational Technology*, 52(5), 1935–1964. <https://doi.org/10.1111/bjet.13116>
- Meylani, R., & Kutluca, T. (2025). AI-powered discourse in mathematics education in support of SDG 4: A systematic review of contemporary research literature. *Discourse and Communication for Sustainable Education*, 16(2), 25–43. <https://doi.org/10.2478/dcse-2025-0014>
- Moore, R., Jiang, S., & Abramowitz, B. (2022). What would the matrix do?: A systematic review of K-12 AI learning contexts and learner-interface interactions. *Journal of Research on Technology in Education*. <https://doi.org/10.1080/15391523.2022.2148785>
- Mredula, K. P., Jonita, R., & Sajja, P. (2024). AI-based tools in mathematics education: A systematic review of characteristics, applications, and evaluation methods. *International Research Journal on Advanced Engineering Hub*, 2(7), 1958–1967. <https://doi.org/10.47392/irjaeh.2024.0268>

- Myers, J. A., Brownell, M. T., Griffin, C. C., Hughes, E. M., Witzel, B. S., Gage, N. A., Peyton, D., Acosta, K., & Wang, J. (2021). Mathematics interventions for adolescents with mathematics difficulties: A meta-analysis. *Learning Disabilities Research & Practice*, 36(2), 145–166. <https://doi.org/10.1111/ldrp.12244>
- Opesemowo, O., & Adewuyi, H. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(7), em2478. <https://doi.org/10.29333/ejmste/14762>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Panqueban, D., & Huincahue, J. (2024). Inteligencia artificial en educación matemática: Una revisión sistemática. *Uniciencia*, 38(1). <https://doi.org/10.15359/ru.38-1.20>
- Pieper, D., Antoine, S., Mathes, T., Neugebauer, E. A. M., & Eikermann, M. (2014). Systematic review finds overlapping reviews were not mentioned in every other overview. *Journal of Clinical Epidemiology*, 67(4), 368–375. <https://doi.org/10.1016/j.jclinepi.2013.11.007>
- Rakes, C. R., Ronau, R. N., Bush, S. B., Driskell, S. O., Niess, M., & Pugalee, D. (2020). Mathematics achievement and orientation: A systematic review and meta-analysis of education technology. *Educational Research Review*, 31, 100337. <https://doi.org/10.1016/j.edurev.2020.100337>
- Ran, H., Kasli, M., & Secada, W. G. (2021). A meta-analysis on computer technology intervention effects on mathematics achievement for low-performing students in K-12 classrooms. *Journal of Educational Computing Research*, 59(1), 119–153. <https://doi.org/10.1177/0735633120952063>
- Ran, H., Kim, N. J., & Secada, W. G. (2022). A meta-analysis on the effects of technology's functions and roles on students' mathematics achievement in K-12 classrooms. *Journal of Computer Assisted Learning*, 38(1), 258–284. <https://doi.org/10.1111/jcal.12611>
- Saat, N. A., Alias, A. F., & Saat, M. Z. (2024). Digital technology approach in mathematics education: A systematic review. *International Journal of Academic Research in Progressive Education and Development*, 13(4), 173–184. <https://doi.org/10.6007/ijarped/v13-i4/22956>
- Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358, j4008. <https://doi.org/10.1136/bmj.j4008>
- Son, T. (2024). Intelligent tutoring systems in mathematics education: A systematic literature review using the substitution, augmentation, modification, redefinition model. *Computers*, 13(10), 270. <https://doi.org/10.3390/computers13100270>
- Subramaniam, S., Maat, S., & Mahmud, M. S. (2022). Computational thinking in mathematics education: A systematic review. *Cypriot Journal of Educational Sciences*, 17(6). <https://doi.org/10.18844/cjes.v17i6.7494>
- Sun, S., Else-Quest, N. M., Hodges, L., French, A. M., & Dowling, R. (2021). The effects of ALEKS on mathematics learning in higher education: A meta-analysis. *Investigations in Mathematics Learning*, 13(3), 182–196. <https://doi.org/10.1080/19477503.2021.1926194>
- Tamur, M., Ndiung, S., Weinhandl, R., Wijaya, T., Jehadus, E., & Sennen, E. (2023). Meta-analysis of computer-based mathematics learning in the last decade Scopus database: Trends and implications. *Infinity Journal*, 12(1), 101–116. <https://doi.org/10.22460/infinity.v12i1.p101-116>
- Tlili, A., Salha, S., Wang, H., & Huang, R. (2023). Intelligent tutoring systems examined in social experiments—Is the magic gone? A meta-analysis. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies*. <https://doi.org/10.1109/icalt58122.2023.00020>
- Velásquez Lecca, S. M., Manco Chávez, J. A., Borja Torres, R. J., Huamán Malca, W. A., Candia Quispe, W. W., & Cortez Egusquiza, R. (2024). Math gamification and ICT for university learning: Systematic review article. *International Journal of Religion*, 5(1), 218–236. <https://doi.org/10.61707/j5c9ts33>
- Wang, X., Huang, R., Sommer, M., Pei, B., Shidfar, P., Rehman, M. S., Ritzhaupt, A. D., & Martin, F. (2024). The efficacy of artificial intelligence-enabled adaptive learning systems from 2010 to 2022 on learner outcomes: A meta-analysis. *Journal of Educational Computing Research*. <https://doi.org/10.1177/07356331241240459>
- Williams, R., Citkowicz, M., Miller, D. I., Lindsay, J., & Walters, K. (2022). Heterogeneity in mathematics intervention effects: Evidence from a meta-analysis of 191 randomized experiments. *Journal of Research on Educational Effectiveness*, 15(3), 584–634. <https://doi.org/10.1080/19345747.2021.2009072>

- Xie, C. (2021). What can China learn from evidence-based educational reform? A comparative review of educational technology programs' effects on mathematics achievement. *ECNU Review of Education*, 4(1), 65–83. <https://doi.org/10.1177/2096531120944410>
- Zhumashev, M., Barlybayev, A., Kintonova, A., Sabitov, A., & Kudubayeva, S. (2025). A systematic review (2010–2024) of smart textbooks and intelligent tutoring systems: Models, methods, subjects, and geographies. *International Journal of Innovative Research and Scientific Studies*, 8(6). <https://doi.org/10.53894/ijirss.v8i6.10048>
- Çavus, H., & Deniz, S. (2022). The effect of technology assisted teaching on success in mathematics and geometry: A meta-analysis study. *Participatory Educational Research*, 9(2), 140–160. <https://doi.org/10.17275/per.22.45.9.2>

APPENDIX

Appendix A. Search strategy and query strings

Source and parameters: the primary search was conducted on 18 December 2025 using Semantic Scholar. The time window was restricted to 2010-2025. Forward-backward citation chasing was performed for high-relevance reviews.

Query 1 (main): ("intelligent tutoring system" OR "intelligent tutoring systems" OR "adaptive learning" OR "adaptive tutoring" OR "AI tutoring" OR "artificial intelligence tutor" OR "math tutoring system") AND ("mathematics" OR "math" OR "numeracy" OR "algebra" OR "geometry") AND ("systematic review" OR "meta-analysis" OR "review" OR "overview of reviews").

Query 2 (school mathematics focus): ("intelligent tutoring system" OR "adaptive learning" OR "adaptive tutoring") AND ("mathematics education" OR "school mathematics" OR "K-12") AND ("systematic review" OR "meta-analysis").

Query 3 (tutoring-specific): ("intelligent tutoring system" OR "adaptive tutoring" OR "step-by-step feedback") AND ("mathematics") AND ("systematic review" OR "meta-analysis" OR "overview of reviews").

Appendix B. Deduplication and selection rules

Eligibility screening: full texts were examined to confirm that the study was a structured review and that the intervention category (intelligent tutoring/adaptive/artificial intelligence-enabled learning technology) was applied in mathematics learning contexts.

Deduplication: (1) remove exact duplicates using DOI matching; (2) for records without DOI, normalize titles (lowercase, remove punctuation, remove extra whitespace) and treat records as duplicates when normalized titles match; (3) when uncertain, retain the record and decide at the full-text stage.