MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2026, 22(12), em2967
ISSN:1305-8223 (online)
OPEN ACCESS Literature Review https://doi.org/10.29333/ejmste/19955

Evaluating the effect of metacognitive instruction on mathematics achievement: A meta-analysis

Riyan Hidayat [1], Zetriuslita [2*], Sudirman [3], Yousef Wardat [4]

[1] Faculty of Education, Languages, Psychology & Music, SEGi University, Selangor, MALAYSIA
[2] Department of Mathematics Education, Universitas Islam Riau, Pekanbaru, INDONESIA
[3] Department of Mathematics education, Universitas Terbuka, Jakarta, INDONESIA
[4] Department of Curriculum and Instruction, Yarmouk University, Irbid, JORDAN

Received 03 August 2026 ▪ Accepted 07 November 2026

Abstract

This meta-analysis examines the effects of metacognitive instruction on mathematics achievement by investigating several potential moderator variables. Consistent with the preferred reporting items for systematic reviews and meta-analyses guidelines, a systematic search across Web of Science, Scopus, and ERIC produced 34 eligible studies. Following the screening and quality assessment using the Joanna Briggs Institute checklist, a total of 34 studies involving 13,805 participants were included in the analysis. Effect sizes were calculated in Jamovi using a random-effects model, reporting standardized mean differences (Hedges's g) with 95% confidence intervals (CIs). The findings indicate a large and statistically significant positive effect of metacognitive instruction on mathematics achievement (g = .87, 95% CI [.72, 1.02]). Among the moderators examined, educational level ($p > .05$), student characteristics ($p > .05$), question type ($p > .05$), mathematics content area ($p > .05$), instructional medium ($p > .05$), and metacognitive strategy type ($p > .05$) showed no statistically significant moderating effects. However, sample size ($p < .05$) and intervention duration $p < .05$) exhibited significant moderation effects.

Keywords: active learning environment, mathematics achievement, mathematical skill, metacognitive instruction, problem-solving

INTRODUCTION

There is continued scholarly interest in the role of metacognitive instruction in improving learning outcomes, alongside on-going debate on the effectiveness of metacognitive instruction in diverse educational contexts. There is growing evidence that it works, but researchers are still trying to figure out for whom and under what conditions these interventions are most helpful. de Boer et al. (2018) conducted a large meta-analysis and found metacognitive strategy instruction to have significant medium-sized effects on students' long-term academic performance. Since then, these positive results have been replicated in different disciplinary contexts, such as English language learning (Bozorgian & Shamsi, 2025), mathematics education, and other subject areas (Hidayat et al., 2025). Encouragingly, these instructional approaches seem to benefit even low-achieving students (Zohar & Ben-David, 2008), suggesting that metacognitive strategies can act as an equalizing force in heterogeneous classrooms. Essentially, metacognition allows students to self-evaluate their performance critically, identify and solve learning obstacles, monitor their understanding, utilize information appropriately, set achievable goals, and review their academic progress in a systematic manner (Qudratuddarsi et al., 2022). This ability to self-regulate is widely seen as core to lifelong learning and academic independence. However, the practical delivery of metacognitive instruction within real-world classroom settings is by no means straightforward. More recently, Fono and Zohar (2025) point to existing issues of lack of time, rigid curricula, poor teacher preparation, and difficulty in assessing metacognitive development adequately. These obstacles underscore the gap between theoretical promise and pedagogical reality, reminding

© 2026 by the authors; licensee Modestum. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/).
✉ riyan160989@gmail.com ✉ zetriuslita@edu.uir.ac.id **(*Correspondence)** ✉ sudirman.official@ecampus.ut.ac.id
✉ yousef_alwardat@yahoo.com

Contribution to the literature

- This study synthesizes existing research on the effects of metacognitive instruction on mathematics achievement, incorporating data published up to 2025 across primary, secondary, and higher education levels.
- We identified that educational level, student characteristics, question type, mathematics content area, instructional medium, and type of metacognitive strategy did not significantly moderate the effect of metacognitive instruction on mathematics achievement.
- The review identified that sample size and intervention duration significantly moderated the effects of metacognitive instruction on mathematics achievement.

educators that effective integration demands careful planning, sustained institutional support, and adaptive instructional design tailored to specific learner populations and subject matter demands.

In the context of standard mathematics education, Schneider and Artelt (2010) argue that the cognitive processes alone do not explain mathematical performance, as metacognitive functions, such as monitoring, strategic planning and self-regulation, are essential in the way students' approach and solve mathematical tasks. Importantly, the benefits of metacognitive instructional strategies are not restricted to high achieving students, as both typically developing learners and those with poorer mathematical skills have been shown to gain significantly from such interventions. Further, metacognition is not a static trait, but a teachable set of skills, as Shilo et al. (2019) emphasize, and well-designed instructional frameworks can successfully develop these abilities by engaging students in metacognitive mathematical conversations and providing explicit instruction on the procedural and rational dimensions of metacognitive strategies. This means, for example, within the framework of mathematical modelling, a systematic application of metacognitive strategies throughout the problem-solving cycle. Planning strategies guide learners from initial simplification to interpretation; monitoring strategies ensure awareness throughout the modelling process; regulation strategies control the transition from mathematization to formal mathematical work; and evaluation strategies are used at crucial stages such as simplification, mathematization and validation (Mansilla et al., 2025).

Previous meta-analyses have mainly focused on specific aspects of metacognition, including learning strategies related to monitoring accuracy (Gutierrez de Blume, 2022), computer-based prompts for self-regulated learning (SRL) (Guo, 2022), and broader metacognitive instruction to improve general academic achievement (de Boer et al., 2018). These studies all consistently show positive impacts on mathematics performance, but vary considerably in the timeframes, inclusion parameters, effect size (ES) metrics, and moderators evaluated. Importantly, few individual studies have comprehensively investigated how intervention duration, sample size, student demographics (e.g., learning disabilities and baseline achievement), task type, type of strategy, and delivery medium interact exclusively within mathematics achievement. To address this gap, our study evaluates metacognitive teaching in mathematics by systematically testing these moderating factors including duration, sample size, educational level, student characteristics, question and content types, strategy style, and instructional format.

Research Questions

1. What is effect of metacognitive instruction on mathematics achievement?
2. Do moderator variables, including intervention duration, sample size, educational level, characteristic of students, type of question, mathematics content area, type of metacognitive strategy, instructional medium, significantly influence the effects of metacognitive instruction on mathematics achievement?

LITERATURE REVIEW

Metacognitive Instruction

Metacognition refers to an individual's awareness of their own thinking processes, as well as the conscious monitoring and regulation of those processes. Metacognition has been defined by Flavell (1979) as "one's knowledge concerning one's own cognitive processes and products or anything related to them", and "active monitoring and consequent regulation and orchestration of these processes" (p. 232). This conceptualization is a base level thinking about thinking and is distinct from cognition itself. Later students elaborated the concept into two big factors. Brown (1987) had distinguished between metacognitive knowledge (the knowledge of what people know about their cognition, tasks and strategies) and metacognitive regulation (the executive functions of planning, monitoring and evaluating cognitive processes). Schraw and Moshman (1995) also expanded the definition of metacognitive knowledge, which consists of declarative knowledge, procedural knowledge, conditional knowledge, and regulation, which refers to planning

(setting goals and choosing strategies), monitoring (keeping track of their comprehension and progress), and evaluating (judging the effectiveness of the strategies and results).

Metacognition serves as a central mechanism within this cycle, enabling students to regulate their cognition, motivation, and behavior during mathematical learning. (Schoenfeld, 1985). If the pupils are not able to control their thinking processes, they tend to use algorithms mechanically or give up on a problem too soon when they encounter problems. These processes are designed to be explicit and trainable in the context of metacognitive instruction. The most significant is IMPROVE (Mevarech & Kramarski, 1997). Introducing new concepts, metacognitive questioning, practicing, reviewing and reducing difficulties, obtaining mastery, verification, and enrichment. The key to the method is the systematic application of four categories of self-addressed questions: Comprehension questions (e.g., "What is the problem about?"), connections to other work done (e.g., "How does this problem relate to other problems you've already completed?"), and questions that require students to take a position (e.g., "what should be done?"), and what were the reflection questions in the problem? (e.g., "does the solution make sense?") what have you learned?").

Subsequently, Kramarski and Mevarech (2003) showed that the interaction of IMPROVE and cooperative learning led to greater effects on mathematical reasoning than either method. Other methods of instruction involve think-aloud modelling by the teacher, self-explanation prompts, reflective journaling, concept mapping and digital scaffolds to guide planning and monitoring in problem-solving. SRL theory is closely related to metacognitive instruction. Zimmerman (2008) proposed that SRL consists of three stages: forethought (goal setting and strategic planning), performance (self-control and self-observation), and self-reflection (self-evaluation and adaptation). Metacognition is a major mechanism in this cycle which helps students control their cognition, motivation and behavior in mathematical learning.

The Effect of Metacognitive Instruction on Mathematics Achievement

A substantial body of experimental and quasi-experimental evidence demonstrates the effectiveness of metacognitive instruction in improving mathematics achievement. Results have been found to be positively related to the level of education, the mathematical content of tasks, and the type of students. In his early research, Cardelle-Elawar (1995) found that metacognitive instruction was effective in enhancing problem-solving skills of low achieving children in grade 3-grade 8. The students who were trained in asking themselves questions related to the problem, their strategies, and reasonableness of their solutions, performed better than control groups. Likewise, Mevarech and Kramarski (1997) discovered that there was a marked improvement in mathematics achievement in the IMPROVE method in ungraded classrooms with no tracking system. Kramarski et al. (2002) expanded on these results by applying the metacognitive instruction to authentic, real world mathematical tasks, showing that metacognitive instruction boosted both standard problems solving and the ability to solve complex, open-ended problems.

Montague et al. (2011) provided strong evidence for students with learning disabilities. Their cognitive and metacognitive strategy instruction program significantly improved mathematical problem-solving performance among middle-school students with learning disabilities compared with typical instruction. Effects were observed not only on immediate post-tests but also on maintenance measures, suggesting that students internalized the strategies. The mechanisms underlying these effects are well documented. Metacognitive instruction helps students:

(1) select and flexibly apply appropriate strategies,

(2) detect and correct errors earlier in the solution process,

(3) persist longer when facing difficulties,

(4) transfer knowledge and strategies to novel problems, and

(5) develop greater awareness of their own thinking processes.

Effects appear particularly beneficial for lower-achieving students and those with learning difficulties, who are less likely to spontaneously engage in effective metacognitive regulation (Lucangeli et al., 2019). Nevertheless, positive outcomes have also been reported for typically achieving and higher-achieving students, indicating that metacognitive instruction can benefit learners across the achievement spectrum.

Existing Meta-Analytic of the Effect of Metacognitive Instruction on Mathematics Achievement

Several meta-analytics have synthesized the growing body of evidence, confirmed the overall effectiveness of metacognitive approaches while also revealed important limitations. Hidayat et al. (2025) conducted one of the recent syntheses. Analyzing 43 studies involving 13,924 participants (experimental group n = 7,061; control group n = 6,863), they reported a large overall effect of metacognitive instruction on mathematics achievement ($g = 1.11$, 95% confidence interval [CI] [.85, 1.36]). Large effects were also found for metacognitive skills ($g = 1.18$) and other learning outcomes ($g = 1.27$). The authors employed a random-effects model (REM) and followed preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines. However, they noted that insufficient data prevented a thorough

examination of student characteristics (e.g., comparisons among typically achieving students, low and high achievers, and students with learning disabilities) as potential moderators.

Sercenia and Prudente (2023) concentrated on studies between 2015 and 2022. 23 studies were included from a pool of 2,341 records. They used an REM to calculate a large overall ES (g = 1.358, 95% CI [.99, 1.72]). The results of the moderator analyses showed significant differences depending on the mathematics content area; the differences depending on educational level and targeted learning outcomes were not significant. Their findings are limited in generalizability due to the relatively brief publication period and limited number of moderators.

There are also some other relevant syntheses, such as the correlational relationship between metacognition and mathematics performance, which Muncer et al. (2022) investigated in adolescents aged 11-16, and found moderate positive correlation (r = .37, 95% CI [.29, .44]). A similar correlation was reported by Xie et al. (2024) (r = .32). Lee et al (2018) focused their research specifically on algebraic thinking and there was a positive outcome from metacognitive training. Wider studies of strategy instruction have found medium-to-large effects, which usually appear to be persistent (immediate post-test g ≈ .50; delayed post-test g ≈ .63) (de Boer et al., 2018).

Potential Moderators of the Effect of Metacognitive Instruction

The effect of metacognitive instruction may vary depending on characteristics of the learners, the instructional intervention, the mathematical tasks and content, and the methodological features of the studies. This assumption is consistent with metacognitive theory, which conceptualizes metacognition as involving knowledge about cognition and the regulation of cognitive processes through activities such as planning, monitoring, and evaluation (Schraw & Moshman, 1995). In mathematics learning, these regulatory processes are particularly important because students need to select appropriate strategies, monitor their progress, identify errors, and evaluate the reasonableness of their solutions (Schoenfeld, 1985). These moderators were selected and organized into four dimensions:

(1) learner-related factors,

(2) intervention-related factors,

(3) task and content-related factors, and

(4) study-design factors.

This framework provides a systematic rationale for examining why the effects of metacognitive instruction may differ across studies rather than treating the moderators as an arbitrary collection of variables.

Learner-related factors reflect differences in students' cognitive development, prior knowledge, metacognitive knowledge, and strategy use. Such differences may influence how students respond to explicit metacognitive instruction (Flavell, 1979; Schraw & Moshman, 1995). At the educational level, students at primary, secondary, and higher education levels may differ in cognitive development, mathematical knowledge, and capacity for independent regulation. Younger learners may require greater modelling and scaffolding, whereas older students may apply metacognitive strategies more independently. Educational level was therefore included as a potential moderator. Previous research has examined age and educational stage in relation to metacognition and mathematics achievement (Xie et al., 2024), while Sercenia and Prudente (2023) examined educational level in metacognitive-based mathematics interventions. Finally, Students also differ in achievement level, learning needs, and existing metacognitive abilities. Lower-achieving students and those with learning difficulties may benefit from explicit instruction because they may use effective strategies less spontaneously (Montague et al., 2011). In contrast, students with well-developed strategies may have less room for additional gains. Student characteristics were therefore included to examine whether learner profiles influence intervention effects.

Intervention-related factors concern how metacognitive instruction is implemented. Metacognitive regulation involves interconnected processes that may require explicit instruction and repeated practice (Schraw & Moshman, 1995). Longer interventions provide more opportunities to practice and internalize metacognitive strategies. However, brief but intensive interventions may also produce meaningful effects. Intervention duration was therefore examined to determine whether instructional exposure contributes to differences in ESs. Metacognitive instruction may involve planning, monitoring, evaluation, self-questioning, self-explanation, think-aloud procedures, or combinations of these strategies. Because these approaches target different aspects of metacognitive regulation, their effectiveness may vary (Schraw & Moshman, 1995). For example, the IMPROVE approach integrates metacognitive questioning into mathematical problem-solving (Mevarech & Kramarski, 1997). Strategy type was therefore included as a moderator. Finally, face-to-face, digital, and hybrid environments differ in opportunities for teacher modelling, feedback, prompting, and scaffolding. These differences may influence how students develop and apply metacognitive strategies. Instructional medium was therefore included as a potential moderator.

In term of task- and content-related factors, metacognitive regulation is particularly relevant to mathematical tasks requiring strategic decision-making, monitoring, and evaluation (Schoenfeld, 1985). Routine procedural questions generally involve established procedures, whereas complex problems, word

problems, and modelling tasks require greater planning, strategy selection, monitoring, and evaluation. Research has demonstrated the application of metacognitive instruction to both standard and complex mathematical problems (Kramarski et al., 2002). Question type was therefore included to examine whether task demands influence intervention effects. Different domains, including arithmetic, algebra, geometry, and problem-solving, involve different conceptual and strategic demands. Previous meta-analytic evidence identified mathematics content area as a potential source of variation in intervention effects (Sercenia & Prudente, 2023), while Xie et al. (2024) reported significant moderation by mathematical domain. Thus, content area was included as a moderator.

The fourth-dimension concerns methodological characteristics of the primary studies. Although such characteristics do not represent theoretical features of metacognitive instruction itself, they may contribute to variation in observed ESs and should therefore be considered when interpreting meta-analytic heterogeneity. Sample size can affect the precision and stability of estimated ESs. Studies with smaller samples may produce less precise estimates and greater sampling variability, whereas larger samples generally provide more precise estimates. For example, Turmuzi and Lu'luilmaknun (2023) conducted a meta-analysis of mathematics learning and found that studies with larger sample sizes, particularly those exceeding 30 participants, tended to report greater ESs than studies with fewer than 30 participants. Accordingly, sample size was included as a study-design moderator to examine whether differences in study size are associated with variation in the reported effects of metacognitive instruction.

The moderating effects of intervention duration, sample size, mathematical task type (procedural versus problem-solving versus modelling), type of metacognitive strategy (planning, monitoring, or both), and media type (traditional face-to-face versus digital/hybrid) were not examined systematically. Theoretically relevant and practically important, these are design and contextual factors relevant for the effective implementation of metacognitive instruction. Furthermore, in recent years, a dynamic integration of digital and blended learning environments has occurred, and a new synthesis is needed that explicitly relates technological to traditional learning. The present meta-analysis attempts to fill these gaps by systematically searching the literature comprehensively and comprehensively, applying a strict inclusion criterion, and analyzing more theoretically oriented moderators. Doing so, it seeks to offer a more sophisticated understanding of the circumstances that support the effectiveness of metacognitive instruction on mathematics achievement the most.

METHODOLOGY

Research Design

This meta-analysis investigated the effects of metacognitive instruction on mathematics achievement and was conducted in accordance with the PRISMA guidelines (Moher et al., 2015). PRISMA offers a standardized method to enhance the transparent and comprehensive reporting for systematic reviews and meta-analyses.

Meta-analysis is a systematic statistical method for synthesizing results of previous studies. It enables investigators to estimate overall ESs, assess heterogeneity across studies, and gain a more complete understanding of research outcomes. Meta-analysis, by pooling evidence from various studies, allows for more reliable conclusions regarding the efficacy of interventions or relationships among variables. The review was conducted in five steps: identification, screening, eligibility, applying inclusion/exclusion criteria and data analysis. Each stage is described in the following sections.

A methodical search of three sources was used to find the studies that were part of this meta-analysis during the identification phase. From the oldest records available until December 2025, the ERIC, Web of Science (WoS), and Scopus databases were examined for pertinent material. In order to facilitate an efficient retrieval procedure across several databases, keywords were carefully chosen to reflect the study focus on metacognitive instruction and mathematical achievement.

Boolean operators, such as "AND" and "OR" were used to create thorough search plans and increase the literature search's precision. ("Metacognit*") AND ("mathematics education" OR "Math OR "Arithmetic" OR "Statistics" OR "Numeracy") was the search string utilised throughout the databases. The database searches generated a substantial number of records, with 2,665 articles retrieved from Scopus, 125 from ERIC, and 1,933 from WoS, resulting in a total of 4,723 records requiring further screening (see **Figure 1**).

During the screening phase, a two-stage screening process was carried out, starting with three independent reviewers looking through the titles and abstracts of retrieved research. The predefined eligibility criteria, which included

(1) a pre-/post-test design,

(2) the inclusion of at least one measure of mathematics achievement,

(3) participation of at least one group receiving metacognitive instruction, and

(4) the existence of a control group, were evaluated for each abstract.

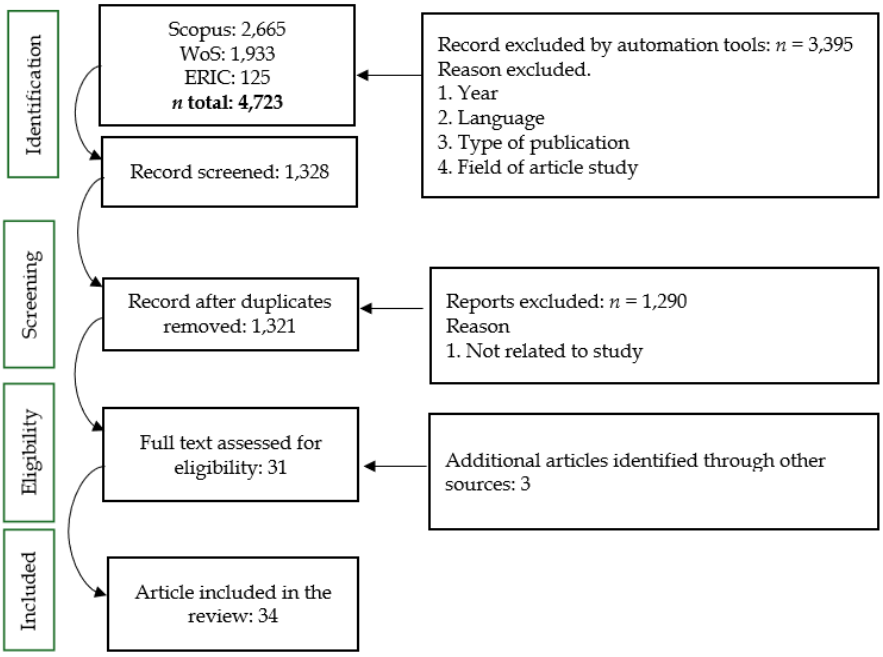


Figure 1. PRISMA flow diagram for the study selection process (adapted from Moher et al., 2015)

Studies that did not fit the predetermined standards for title relevancy, publication year, document format, language, accessibility, and research topic were eliminated. Studies examining the effect of metacognitive instruction on mathematical achievement were the focus of the review. To guarantee adequate empirical evidence for trustworthy meta-analytic findings, publications from 2000 to 2025 were taken into consideration. Book chapters were not included; only academic articles published in English-language academic journals and conference proceedings were. All included papers had to be available in full text, which produced a varied collection of research from several nations. First, we conducted automation-assisted screening using the filtering functions available within the three bibliographic databases to exclude records based on four predefined criteria: publication year, language, publication type, and field of study. A total of 3,395 records were excluded using these database filters across the three databases, while the remaining records proceeded to subsequent title/abstract and full-text screening. After duplicates were removed, a total of 1,321 records were retained for further evaluation.

In order to find studies that satisfied the requirements for inclusion in the meta-analysis, the remaining 1,321 records were methodically evaluated during the eligibility phase. After the first screening procedure, the reviewers conducted a thorough assessment in the third screening stage to ascertain whether the studies met the predetermined eligibility conditions. Articles were distributed independently among the reviewers for evaluation of title relevancy and alignment with the research objectives in order to reduce potential bias. A total of 1,290 records were excluded for further evaluation due to not related study. Next, three additional articles were identified through Google Scholar that met eligible criteria. Finally, thirty-four full-text papers were evaluated further after irrelevant research were eliminated and additional articles. The reliability and validity of the meta-analysis findings were strengthened by this stringent selection process, which made sure that only studies that satisfied the predetermined criteria and made significant contributions were included. Studies that met the eligibility requirements had to include enough statistical data, such as sample size, mean scores, and standard deviations (SDs), to allow for the computation of ESs. The inclusion and exclusion criteria used in the study selection procedure are compiled in **Table 1**.

Data Extraction and Quality Assessment

Two major tasks were engaged in the evaluation process: gathering pertinent information from scholarly sources and assessing the included research' methodological quality. A third reviewer was consulted in order to settle any disputes that arose during the procedure. Authorship, publication year, study location, mathematics accomplishment focus, educational level, types of mathematical problems, and student characteristics were among the study variables that were extracted. Furthermore, statistical data pertaining to mathematics achievement was gathered, such as result means and variability measures. The Joanna Briggs Institute (JBI) checklist was used to evaluate the methodological quality of the chosen research (Barker et al., 2024). Temporal sequence, confounding variables, intervention fidelity, outcome measurement, participant retention, and the validity of statistical analyses are among the possible causes of bias that this assessment tool assesses.

Table 1. Criteria for study inclusion and exclusion

Criteria	Inclusion	Exclusion
Article details and content	Relevant title and meets the eligibility criteria	Titles that were irrelevant or did not meet the eligibility criteria
Time frame	Publications from 2000 to 2025	Publications outside the specified timeframe
Publication type	Only articles published in peer-reviewed academic journals and conference proceedings were included in the review	Review articles, editorials, and non-empirical publications
Language criteria	English-language publications	Non-English languages
Subject domain	Mathematical domains	Subjects outside mathematics
Resource availability	Full-text publications	Accessing the full articles may require payment or a subscription
Study methodology	Non-equivalent control group pre-/post-test design	Pre-/post-test single-group design

Table 2. Coding scheme for moderating effect

Moderating effect	Coding
School level	Pre-school Primary school Secondary school Higher education
Students' characteristic	Normal Learning disabilities Low achievers High achievers
Intervention duration	≤ 4 weeks 5-8 weeks > 8 weeks
Sample size	< 50 50-100 > 100
Mathematics content area	Word problems/problem-solving Problem-solving Geometry Arithmetic/number sense Algebra
Type of metacognitive strategy	Combinations thereof Planning strategies Monitoring strategies
Instructional medium	Traditional face to face Digital/hybrid/blended learning
Types of tasks	Problem-solving task Procedural task Word problem task Modelling task

Studies were classified as low quality (≤ 3 points), moderate quality (4-5 points), or high quality (6-9 points) according to the criteria put forward by Stojanović et al. (2016). Low quality studies were not included in additional analysis. Next, the coding procedures for the eight moderators across the included studies are presented in **Table 2**.

To explore possible sources of between-study heterogeneity, moderator variables were extracted through a standardized coding framework adapted from prior meta-analyses (Donker et al., 2014; Sercenia & Prudente, 2023; Xie et al., 2024) (**Table 2**). Eight main moderating dimensions were operationalized across educational contexts, instructional features, and methodological variables: educational stage was coded according to school level (pre-school, primary school, secondary school, and higher education); learner profiles were categorized by students' characteristics (normal, learning disabilities, low achievers, and high achievers); intervention dosage was structured by intervention duration (≤ 4 weeks, 5-8 weeks, and > 8 weeks); and study scale was classified according to sample size (< 50, 50-100, and > 100). In addition, curriculum focus was defined by mathematics content area (word problems/problem-solving, problem-solving, geometry, arithmetic/number sense, and algebra), strategic focus by type of metacognitive strategy (planning strategies, monitoring strategies, or combinations of these), delivery approach by instructional medium (traditional face to face vs. digital/hybrid/blended learning), and assessment target by types of tasks (problem-solving, procedural, word problem, and modelling tasks).

Computation of the ES

In the current meta-analysis, some studies reported multiple sets of statistical data, including means, SDs, and sample sizes (n), which allowed more than one relevant ES to be derived from a single study. These ESs were analyzed separately to capture variations across the different conditions and variables examined within the same study (Arifin et al., 2025). Consequently, although the analysis included data from 34 studies, the number of entries increased to 79 to ensure that all relevant effect-size information was appropriately incorporated into the meta-analysis. Two distinct analyses looking at academic performance and other learning outcomes were part of the meta-analytics, which were carried out using the Jamovi software package. Standardized mean differences (Hedges' g) with 95% CIs were used to compute ESs, which were then interpreted using Cohen's (2013) criteria: .2-.5 for small, .5-.8 for moderate, and > .8 for large. An REM was used because of the anticipated variation in participant characteristics and assessment techniques between studies. Between-study heterogeneity was evaluated

Table 3. ES estimates for metacognitive instruction on mathematics achievement

Variables	k	g	Standard error	Z	p	CI lower bound	CI upper bound
Mathematics achievement	79	.87	.0746	11.7	< .001	.72	1.02

using the Q statistic, I^2, and τ^2 estimates, complemented by subgroup analyses. Sub-group analyses based on educational level, student characteristics, question type, mathematics content area, instructional medium, metacognitive technique type, intervention length, and sample size were carried out to investigate possible sources of variability. To guarantee statistical reliability, subgroup analyses were only carried out when each category had at least three ESs. Lastly, publication bias was assessed to determine whether the findings may have been influenced by the selective publication of studies reporting statistically significant results (Nakagawa et al., 2023). The assessment included funnel plots, Egger's test, fail-safe N and the Trim-and-Fill method.

RESULTS

Study Characteristics

The 34 studies included in this review are summarized in **Appendix A**. Secondary school settings represented the largest proportion (13, 39.4%), followed by primary schools (12, 36.4%). One study (3.0%) focused on preschoolers, while three (9.1%) were conducted in higher education. Four studies (12.1%) involved general secondary education or multiple educational levels. Most interventions involved students with normal academic achievement (27, 81.8%). Other participants included students with learning difficulties (3, 9.1%), low achievers (2, 6.1%), and autism spectrum disorder (1, 3.0%). Two studies (6.1%) also included high achievers as a comparison group. Intervention duration varied considerably. Eleven studies (33.3%) lasted five to eight weeks, nine (27.3%) lasted four weeks or less, and seven (21.2%) exceeded eight weeks. Duration was unspecified in six studies (18.2%). Regarding sample size, 11 studies (33.3%) included 50-100 participants, 10 (30.3%) exceeded 100 participants, five (15.2%) included fewer than 50, and seven (21.2%) reported multiple groups.

Word problems and problem-solving were the most common mathematics content areas (13, 39.4%), followed by arithmetic and number sense (8, 24.2%), algebra (6, 18.2%), and geometry (3, 9.1%). Most studies (30, 90.9%) used multiple metacognitive strategies, while planning and monitoring were each examined in one study (3.0%). Instructional approaches were relatively balanced, with 17 studies (51.5%) using face-to-face instruction and 16 (48.5%) using digital, hybrid, or blended approaches. Procedural tasks were most common (11, 33.3%), followed by problem-solving (9, 27.3%) and word problems (7, 21.2%). Modelling tasks appeared in two studies (6.1%), while one combined procedural and problem-solving tasks. Finally, all 34 studies (100%) were rated as having a high level of evidence according to the JBI criteria, indicating that the findings are generally reliable and robust for informing practice and future research directions (see **Appendix A**). The overall sample comprised 13,805 participants, including 7,003 in the experimental group and 6,802 in the control group.

Effect of Metacognitive Instruction on Mathematics Achievement

An REM was used to account for between-study variability in ES estimates in accordance with accepted meta-analytic practices. The impact of metacognitive education on pupils' mathematical success was measured using Hedges's g, a standardized ES index. When compared to students in the control conditions, the pooled results, which are shown in **Table 3**, show that students who received metacognitive education had statistically significant gains in their mathematical ability.

The overall effect estimates of metacognitive instruction on mathematics achievement are shown in **Table 3**. Using Hedges's g as the standardized ES measure and an REM, the study showed a significant positive impact on mathematics achievement (g = .87, 95% CI [.72, 1.02]) (Cohen, 2013). This result was statistically significant (p <.001), suggesting that students' performance in mathematics is reliably improved by metacognitive teaching. When taken as a whole, the results show that metacognitive instruction significantly improves mathematics achievement in a variety of educational contexts. **Figure 2** shows the related forest plot with the effect estimates.

Publication Bias of Mathematics Achievement

There were no notable studies or outliers found in terms of mathematics achievement (**Table 4**). Both Egger's regression test (p < .0001) and the rank correlation test (p < .0001) revealed significant funnel plot asymmetry, indicating the potential existence of publishing bias. In addition, the trim-and-fill analysis did not impute any missing studies. Despite significant asymmetry tests, trim and fill number of studies procedure yielded an adjustment of .000 missing studies. This demonstrates that the detected asymmetry is not driven by missing "file-drawer" studies but rather reflects structural heterogeneity or small-study effects. Again, these can be explained by methodological variations, different study characteristics, or the actual effects of the phenomenon under study rather than as a

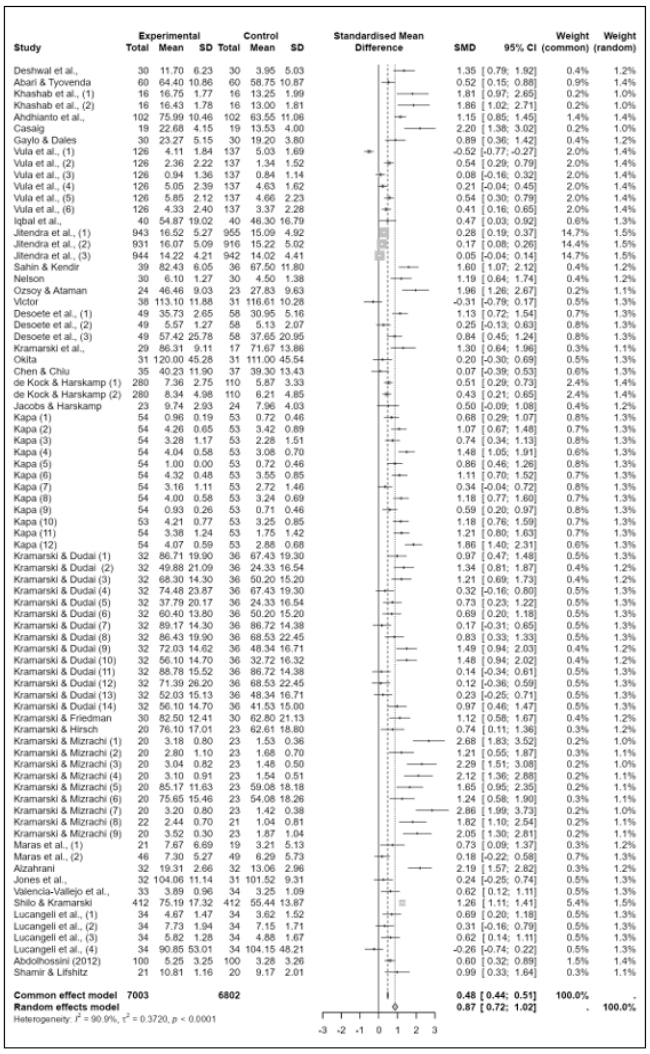


Figure 2. Forest plot of metacognitive instruction's effect on achievement (generated by the authors)

result of selection bias (Neethirajan et al., 2005). Taken together, these results indicate that publication bias is likely to have substantially affected the meta-analysis findings. As a result, care should be taken while interpreting the results. However, a visual analysis of the mathematics achievement funnel plot (**Figure 3**) revealed a fairly symmetrical distribution of ESs around the pooled estimate, with no obvious signs of significant publication bias or small-study effects (Hidayat et al., 2025).

Analysis of Heterogeneity

Cochran's Q test and the I^2 statistic were used to assess heterogeneity among the included studies. To determine if the observed variations in ESs were larger

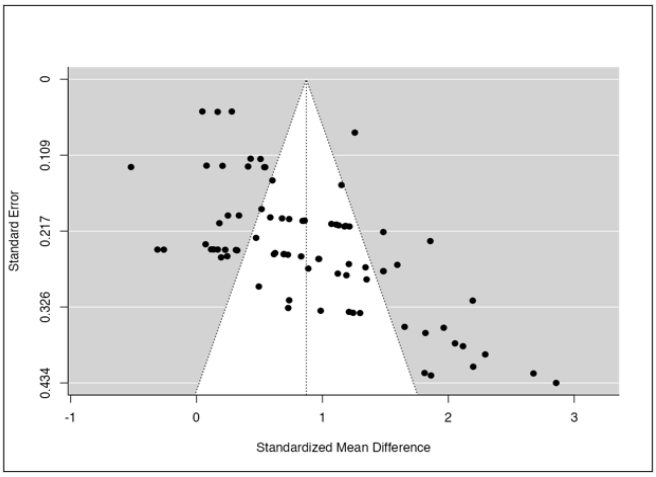


Figure 3. A funnel plot in academic achievement (generated by the authors)

Table 5. Heterogeneity statistics

Variables	Tau²	I²	df	Q	p
Mathematics achievement	.372	90. 9%	78	852.52	< .0001

than those predicted by sampling error alone, Cochran's Q test was used. A significant Q test result ($p < .10$) indicates the presence of between-study heterogeneity and supports the use of an REM, while a non-significant result ($p > .10$) indicates low heterogeneity and suggests that a fixed-effects model may be appropriate (Zhang et al., 2019). The limited maximum likelihood approach was used to estimate the between-study variance (tau^2). The Q statistic and I^2 value were used in addition to tau^2 to evaluate heterogeneity. When heterogeneity was identified (i.e., $tau^2 > 0$), regardless of the significance level of the Q test, a prediction interval was calculated to estimate the range of plausible true ESs. The heterogeneity results, including tau^2, I^2, df, Q statistics, and associated p values, are reported in **Table 5**.

As shown in **Table 5**, the heterogeneity analyses indicated considerable variability across all three outcome domains. Specifically, for mathematics achievement, Cochran's Q test demonstrated significant heterogeneity among the true ESs, $Q(78) = 852.52$, $p < .0001$, with $tau^2 = .372$ and $I^2 = 90.9\%$. Therefore, an REM was employed to obtain a more reliable and generalizable estimate of the overall ES.

Moderator Analysis of ESs

To further examine the variability in ESs, this study analyzed the moderating effects of eight study characteristics: educational level, student characteristics,

Table 4. Publication bias analysis for mathematics achievement

Variables	Test name	Value	p
Mathematics achievement	Fail-safe N	31,157.000	< .001
	Begg and Mazumdar's rank correlation	.431	< .001
	Egger's regression	7.398	< .001
	Trim and fill number of studies	.000	

Table 6. Moderators of metacognitive instruction and achievement

Variables	Moderator	k	g	Intergroup effect (p)
Mathematics achievement	Educational level	79		.528
	Students' characteristic	79		.354
	Type of question	79		.073
	Mathematics content area	79		.279
	Instructional medium	79		.435
	Type of metacognitive strategy	78		.698
	Intervention duration			.002
	< 4 weeks	22	1.29	< .05
	5-8 weeks	36	.817	< .05
	> 8 weeks	18	.550	< .05
	Sample size			<.001
	< 50	15	1.76	< .05
	50-100	37	.621	< .05
	> 100	25	.776	< .05

question type, mathematics content area, instructional medium, metacognitive strategy type, intervention duration, and sample size. The findings from the moderator analyses are summarized in **Table 6**.

Table 6 presents the moderator analyses examining whether various study characteristics influenced the relationship between metacognitive instruction and academic achievement. The findings indicate that most of the proposed moderators were not statistically significant. Specifically, educational level ($p > .05$), student characteristics ($p > .05$), question type ($p > .05$), mathematics content area ($p > .05$), instructional medium ($p > .05$), and type of metacognitive strategy ($p > .05$) did not significantly moderate the effect of metacognitive instruction on academic achievement. In contrast, intervention duration significantly moderated the effect ($p < .05$). Interventions lasting less than 4 weeks ($g = 1.29$, $p < .05$) and 5-8 weeks ($g = .817$, $p < .05$) produced the largest positive effects, whereas interventions exceeding 8 weeks also showed a significant but comparatively smaller effect ($g = .550$, $p < .05$). Similarly, sample size was a significant moderator ($p < .05$). Studies with fewer than 50 participants reported the largest effect ($g = 1.76$, $p < .05$), while studies with sample sizes of 50-100 ($g = .621$, $p < .05$) and more than 100 participants ($g = .776$, $p < .05$) demonstrated significant but more moderate effects.

DISCUSSION

With respect to publication bias, both Begg and Mazumdar's rank correlation test and Egger's regression test showed significant funnel plot asymmetry ($p < .001$). Even though funnel plot asymmetry is often cited as evidence of publication bias, methodological literature stresses that asymmetry frequently reflects wider 'small-study effects', such as genuine clinical or educational heterogeneity, differences in intervention intensity, or varying study design quality across sample sizes, rather than simple selective non-publication (Sterne et al., 2011). In support of this distinction, the trim-and-fill procedure neither detected nor imputed missing studies on the counter-factual side, and the Fail-Safe N showed that more than 31,000 unpublished studies with null effects would be needed to reduce our findings to non-significance. Nonetheless, since smaller studies produced systematically larger effect estimates, the pooled ES should be interpreted with awareness of possible small-study inflation.

This meta-analysis synthesized 79 ESs drawn from 34 primary studies to evaluate the impact of metacognitive instruction on mathematics achievement. The pooled estimate ($g = .87$, 95% CI [.72, 1.02]) indicates a large and statistically robust effect, corroborating a substantial body of prior evidence that metacognitive instruction can meaningfully improve students' mathematical performance (Hidayat et al., 2018; Montague et al., 2011). This magnitude, however, sits between the two most directly comparable recent syntheses: it is smaller than the $g = 1.11$ reported by Hidayat et al. (2025) across 43 studies, and considerably smaller than the $g = 1.358$ reported by Sercenia and Prudente (2023) across 23 studies published between 2015 and 2022. Rather than treating this divergence as noise, it is instructive to consider its sources. Sercenia and Prudente's (2023) narrower publication window and smaller study pool likely captured a less heterogeneous, and possibly less representative, slice of the literature, whereas the present review's more exhaustive search (through December 2025) and stricter methodological inclusion criteria (non-equivalent control-group pre-/post-test designs only) plausibly diluted the pooled estimate toward a more conservative, but arguably more defensible, figure.

From an educational perspective, the magnitude of the pooled effect is important because it indicates that metacognitive instruction is not merely associated with statistically detectable changes in mathematics achievement but can be associated with differences in performance that are potentially meaningful for classroom learning. Metacognitive instruction encourages students to plan how they will approach a

mathematical problem, monitor their understanding and progress while solving it, and evaluate the appropriateness of their strategies and answers. Thus, the educational value of the intervention extends beyond improvement on a single mathematics assessment: it concerns the development of students' capacity to approach unfamiliar problems more deliberately, identify errors, select appropriate strategies, and regulate their own learning. Nevertheless, the standardized ES should not be interpreted as a direct number of additional marks or grade levels gained, because the practical meaning of g = .87 depends on the assessment, student population, instructional context, and duration of implementation. The educational significance is therefore best understood as evidence that metacognitive instruction has the potential to produce substantively meaningful improvements in mathematical learning when translated into effective classroom practice, rather than as a guarantee of a fixed amount of achievement gain for every student.

The magnitude of heterogeneity observed here (I^2 = 90. 9%, Q(78) = 852.52, $p < .0001$) warrants explicit interpretive attention rather than being treated merely as a statistical footnote. As Borenstein (2022) cautions, a high I^2 value does not by itself quantify how much the true effect varies in absolute terms; it is the accompanying 95% prediction interval that does so, and here that interval spans from -.338 to 2.082. This range indicates that, in some plausible future implementations, metacognitive instruction may confer negligible or even reversed benefit relative to control conditions. Educationally, this means that the pooled effect should not be interpreted as evidence that metacognitive instruction will produce equally substantial achievement gains in every classroom. Rather, the wide prediction interval suggests that implementation conditions matter: the instructional approach, quality of implementation, characteristics of learners, mathematical tasks, and characteristics of the comparison condition may influence whether students experience meaningful benefits. The "average" benefit reported in **Table 2** therefore masks considerable context-dependency that practitioners should not overlook. For educators, the implication is that adopting metacognitive instruction should involve attention not only to whether the strategy is used, but also to how consistently and purposefully it is embedded within mathematics teaching.

Turning to the moderator analyses, the absence of statistically significant moderation by educational level, student characteristics, question type, mathematics content area, instructional medium, and metacognitive strategy type indicates that systematic differences in ESs were not detected across these categories within the available evidence. Taken together with theoretical accounts of metacognition as a domain-general, teachable capacity (Panadero, 2017; Zimmerman, 2008), these findings are consistent with the possibility that metacognitive instruction may be relevant across different educational levels and mathematical contexts; however, they do not establish that the effects are equivalent across these categories or confirm the generalizability of the intervention. The educational significance of these findings is therefore that the current evidence does not identify a specific educational level, student characteristic, mathematical topic, question type, instructional medium, or metacognitive strategy type for which the effects systematically differ. However, the interpretation should remain cautious because nonsignificant moderator results may reflect limited statistical power, particularly where some subgroups contain relatively few studies, as well as substantial between-study heterogeneity. Further research with larger and more balanced subgroup samples is needed to determine whether meaningful differences exist across these educational and instructional contexts.

The null result for instructional medium deserves particular comment: it suggests that digitally mediated metacognitive instruction is neither more nor less effective than traditional face-to-face delivery, a pattern broadly consistent with meta-analytic evidence that technology-enhanced mathematics instruction yields only modest average gains over conventional instruction (Cheung & Slavin, 2013) and that technology's benefits depend more on its specific instructional function than on its mere presence (Ran et al., 2022). Educationally, this suggests that schools do not necessarily need to adopt digital delivery solely on the assumption that technology itself will enhance the effectiveness of metacognitive instruction. Instead, the pedagogical function of the technology and the quality of the metacognitive prompts or activities appear more important considerations. At the same time, this null finding should be interpreted cautiously: 90.9% of included studies employed combined rather than isolated metacognitive strategies, sharply restricting the between-category variance available to detect a genuine moderating effect. The absence of significance may therefore reflect insufficient statistical power as readily as genuine equivalence.

In contrast, intervention duration and sample size emerged as significant moderators, and both patterns run counter to intuitive expectation. Interventions shorter than four weeks produced the largest effect (g = 1.29), while those exceeding eight weeks, though still significant, yielded a comparatively modest effect (g = .550). This inverse dose-duration pattern echoes findings from mathematics intervention research more broadly: Nelson and McMaster (2019) similarly reported that interventions exceeding eight weeks produced smaller effects than shorter ones, while dose-response modelling of intensive mathematics interventions for at-risk learners has shown that once basic hypotheses of linear dosage benefit are tested against moderators, the

apparent advantage of additional instructional time can diminish or disappear entirely (Miller et al., 2025). Other mathematics-intervention meta-analyses have found dosage effects that are inconsistent or non-significant once study-level confounds are considered (Rojo et al., 2024), suggesting that duration rarely operates as an isolated causal ingredient.

The educational significance of the duration finding is therefore not that shorter interventions are inherently preferable, but that extending an intervention does not necessarily translate into proportionally greater mathematical achievement. A short, focused intervention may be sufficient to establish basic routines such as planning, monitoring, and evaluating, after which the value of additional instructional time may depend on how those routines are reinforced and integrated into regular mathematics teaching. Several explanations merit consideration for the present pattern: short interventions may benefit from heightened novelty and engagement, tighter experimental control and fidelity, or a ceiling effect whereby extended exposure yields diminishing marginal returns once foundational metacognitive routines have been established. For teachers and program designers, this finding highlights the importance of instructional efficiency: the objective should be sustained and meaningful use of metacognitive practices rather than simply increasing the number of weeks allocated to a program. Future primary research should report intervention dosage with sufficient granularity (total contact hours, frequency of sessions, and opportunities for practice, rather than calendar duration alone) to permit more precise dose-response modelling of the kind now standard in mathematics intervention research.

The sample-size moderation pattern is more troubling from a methodological standpoint. Studies with fewer than 50 participants yielded the largest effect ($g = 1.76$), while larger studies produced smaller, though still significant, effects. This is consistent with a small-study effect, which may arise from publication bias, selective reporting, sampling variability, or differences in methodological rigor across studies (Turner et al., 2013). This pattern directly reinforces the publication-bias indicators already detected in the present review: Egger's regression test returned significant asymmetry ($p < .0001$), and this warning should not be discounted merely because visual inspection of the funnel plot appeared broadly symmetrical. Formal statistical tests may detect asymmetry that is not readily apparent visually, particularly when interpreting a body of evidence with substantial heterogeneity. Educational meta-analyses have also been shown to neglect, or inadequately correct for, such bias (Ropovik et al., 2021).

The educational significance of this finding is particularly important for interpreting the magnitude of benefit claimed for classroom practice. If smaller studies systematically produce larger estimates, educators and policymakers should be cautious about assuming that the largest reported effects will be replicated when metacognitive instruction is implemented across larger and more diverse student populations. Consequently, the overall pooled effect reported here should be interpreted as evidence of meaningful potential rather than as a precise estimate of the achievement gain that schools should expect. In practical terms, the evidence supports the educational relevance of metacognitive instruction while simultaneously indicating that expectations should be calibrated to the more variable effects observed across studies. Replication through adequately powered, methodologically rigorous trials is therefore essential before strong claims can be made about the size of achievement gains likely to occur on scale.

Theoretically, these findings extend support to SRL frameworks that position metacognitive regulation as a generalizable, trainable mechanism rather than a domain- or context-specific skill (Schneider & Artelt, 2010). Educationally, this theoretical interpretation implies that metacognitive instruction can be viewed not simply as an additional mathematics activity, but as a way of strengthening how students engage with mathematical learning itself. Students who learn to plan their approaches, monitor their reasoning, recognize errors, and evaluate the effectiveness of their strategies may become better equipped to manage unfamiliar and increasingly complex mathematical tasks. However, the substantial heterogeneity observed in the present review indicates that the transfer of these potential benefits into classroom achievement is not automatic.

Practically, the results suggest that metacognitive instruction can be a meaningful component of mathematics teaching, particularly when metacognitive processes are explicitly connected to mathematical problem-solving and students are given opportunities to apply, monitor, and evaluate these processes. The educational significance of the present findings therefore lies not only in the size of the pooled statistical effect, but in what the intervention enables students to do differently as learners: approach problems strategically, monitor their own understanding, identify and correct errors, and evaluate the effectiveness of their mathematical reasoning. The findings on intervention duration further suggest that educational value should not be equated with greater program length; rather, program designers should prioritize purposeful practice, instructional fidelity, and integration with ordinary mathematics instruction. At the same time, the substantial heterogeneity and evidence of small-study effects warrant caution in translating the pooled estimate directly into expected classroom gains. Overall, the evidence indicates that metacognitive instruction has meaningful educational potential, but its value is best realized when implementation focuses on the quality and relevance of metacognitive practice rather than on

the mere presence, duration, or technological format of the intervention.

CONCLUSION AND IMPLICATION

Synthesizing 34 JBI-rated studies, this meta-analysis found a large and statistically significant positive effect of metacognitive instruction on students' mathematics achievement. The evidence indicates that interventions combining multiple metacognitive strategies, particularly planning, monitoring, and evaluation, were commonly implemented across the included studies and were associated with positive mathematics outcomes. Moderator analyses did not detect statistically significant differences in the effects of metacognitive instruction according to educational level, student academic baseline, mathematical topic, task type, or delivery medium. However, these findings should not be interpreted as evidence that the effectiveness of metacognitive instruction is equivalent across these categories. The absence of statistically significant moderation indicates that systematic differences were not detected within the available evidence, but differences may remain undetected because of limited numbers of studies or participants within some subgroups and the statistical power of the moderator analyses. Therefore, the generalizability of the effects across educational levels, mathematical content, instructional media, and other categories should be interpreted cautiously. Intervention duration and sample size significantly moderated the observed effects, with interventions lasting less than eight weeks tending to produce larger effects, whereas interventions exceeding eight weeks continued to show positive but comparatively smaller effects. Given the substantial heterogeneity and evidence of small-study effects, the findings should therefore be interpreted as indicating variability in the outcomes of metacognitive instruction rather than uniform effectiveness across educational settings.

The findings have implications for research, teaching, learning, and curriculum design. Future research should address gaps in the evidence base, particularly among underrepresented populations such as preschool and higher-education students and investigate potential explanations for the smaller effects observed in longer interventions, including intervention fidelity, implementation characteristics, and sustainability of metacognitive practices over time. For teachers, the findings support the incorporation of explicit metacognitive practices, such as planning, monitoring, and evaluating, into mathematics instruction, while recognizing that implementation may need to be adapted to students' needs, instructional contexts, and mathematical tasks. For students, opportunities to develop metacognitive processes may support more deliberate approaches to mathematical problem-solving, although the extent of benefit may vary according to learner and instructional characteristics. For curriculum developers, the findings support the inclusion of metacognitive prompts and self-reflection activities in mathematics materials and teacher guidance, while allowing flexibility for adaptation across contexts. The evidence supports metacognitive instruction as a promising approach to improving mathematics achievement, but the substantial heterogeneity, small-study effects, and limited representation of some educational populations underscore the need for cautious interpretation and further high-quality research to determine the conditions under which metacognitive instruction is most beneficial.

Limitations and Recommendations

Although this meta-analysis provides evidence that metacognitive instruction positively influences mathematics achievement, several limitations should be acknowledged. First, substantial heterogeneity was observed across the included studies (I^2 = 93.49%), indicating considerable variation in intervention characteristics, participant populations, instructional contexts, and outcome measures. Although an REM was employed to account for this variability, the high heterogeneity limits the precision and generalizability of the pooled ES. Second, statistical tests (Begg and Mazumdar's rank correlation and Egger's regression) suggested the presence of publication bias, implying that studies reporting significant positive findings may have been more likely to be published than those with null or negative results. Third, several moderator categories contained relatively few studies, particularly those involving preschool learners, higher education students, planning-only strategies, monitoring-only strategies, and modelling tasks, reducing the statistical power to detect meaningful subgroup differences. Fourth, only English-language publications from peer-reviewed journals and conference proceedings were included, potentially introducing language and publication bias while excluding relevant evidence from other sources. Finally, most primary studies reported short-term post-test outcomes, providing limited evidence regarding the long-term sustainability of metacognitive instruction and its impact on students' mathematical achievement over extended periods.

Based on these findings, several recommendations are provided for educational research and practice. Researchers should prioritize methodologically rigorous designs with larger sample sizes to mitigate small study bias and improve statistical power. Future studies ought to focus on underrepresented educational contexts, specifically pre-school and higher education environments, to evaluate age dependent dynamics of metacognitive instruction. To better understand dosage effects, researchers should systematically report granular implementation data, such as total contact hours and implementation fidelity, rather than relying

solely on calendar weeks. Longitudinal studies are also needed to explore why longer interventions show reduced ESs and to investigate strategies for sustaining long term student engagement. For practical application, educators and curriculum designers should feel confident embedding explicit multi strategy metacognitive tools, such as structured planning, monitoring, and evaluation prompts, into mathematics instruction. Because instructional medium was not a significant moderator, teachers can seamlessly integrate these metacognitive scaffolds across traditional face to face, digital, and blended learning environments. Finally, schools should implement metacognitive instruction focused 4 to 8 week learning cycles to maximize achievement gains efficiently.

Author contributions: RH: conceptualization, data curation, writing – original draft, writing – review & editing; **Z & YW:** data curation, writing – original draft, writing – review & editing; **S:** writing – original draft, writing – review & editing. All authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: This study is a literature review. Thus, it does not require any ethical approval.

AI statement: AI was used to improve the manuscript by checking its coherence, grammar, and overall readability, thereby enhancing the clarity and structure of the text.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Arifin, Z., Saputro, S., & Kamari, A. (2025). The effect of inquiry-based learning on students' critical thinking skills in science education: A systematic review and meta-analysis. *Eurasia Journal of Mathematics, Science and Technology Education, 21*(3), Article em2592. https://doi.org/10.29333/ejmste/15988

Barker, T. H., Habibi, N., Aromataris, E., Stone, J. C., Leonardi-Bee, J., Sears, K., Hasanoff, S., Klugar, M., Tufanaru, C., Moola, S., & Munn, Z. (2024). The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. *JBI Evidence Synthesis, 22*(3), 378-388. https://doi.org/10.11124/jbies-23-00268

Borenstein, M. (2022). In a meta-analysis, the I-squared statistic does not tell us how much the effect size varies. *Journal of Clinical Epidemiology, 152*, 281-284. https://doi.org/10.1016/j.jclinepi.2022.10.003

Bozorgian, H., & Shamsi, E. (2025). A review of research on metacognitive instruction for listening development. *International Journal of Listening, 39*(1), 1-16. https://doi.org/10.1080/10904018.2023.2197008

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert, & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Lawrence Erlbaum. https://doi.org/10.4324/9781003758167-5

Cardelle-Elawar, M. (1995). Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education, 11*(1), 81-95. https://doi.org/10.1016/0742-051X(94)00019-3

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88-113. https://doi.org/10.1016/j.edurev.2013.01.001

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

de Boer, H., Donker, A. S., Kostons, D. D., & van der Werf, G. P. (2018). Long-term effects of metacognitive strategy instruction on student academic performance: A meta-analysis. *Educational Research Review, 24*, 98-115. https://doi.org/10.1016/j.edurev.2018.03.002

Donker, A. S., de Boer, H., Kostons, D., Dignath van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review, 11*, 1-26. https://doi.org/10.1016/j.edurev.2013.11.002

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

Fono, D., & Zohar, A. (2025). Metacognitive instruction: Central aspects of pre-service and novice in-service teachers' knowledge and professional development. *Professional Development in Education, 51*(8), 1766-1789. https://doi.org/10.1080/19415257.2024.2409779

Guo, L. (2022). Using metacognitive prompts to enhance self-regulated learning and learning outcomes: A meta-analysis of experimental studies in computer-based learning environments. *Journal of Computer Assisted Learning, 38*(3), 811-832. https://doi.org/10.1111/jcal.12650

Gutierrez de Blume, A. P. G. (2022). Calibrating calibration: A meta-analysis of learning strategy instruction interventions to improve metacognitive monitoring accuracy. *Journal of Educational Psychology, 114*(4), 681-700. https://doi.org/10.1037/edu0000674

Hidayat, R., Mohd Saad, M. R., & Wewe, M. (2025). A meta-analysis of the effect of metacognitive instruction on mathematics achievement. *Cogent Education, 12*(1), Article 2517510. https://doi.org/10.1080/2331186X.2025.2517510

Hidayat, R., Syed Zamri, S. N. A., & Zulnaidi, H. (2018). Does mastery of goal components mediate the relationship between metacognition and mathematical modelling competency? *Educational Sciences: Theory and Practice, 18*(3), 579-604. https://doi.org/10.12738/estp.2018.3.0108

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281-310. https://doi.org/10.3102/00028312040001281

Kramarski, B., Mevarech, Z. R., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. *Educational Studies in Mathematics, 49*(2), 225-250. https://doi.org/10.1023/A:1016282811724

Lee, Y., Capraro, M. M., Capraro, R. M., & Bicer, A. (2018). A meta-analysis: Improvement of students' algebraic reasoning through metacognitive training. *International Education Studies, 11*(10), 42-49. https://doi.org/10.5539/ies.v11n10p42

Lucangeli, D., Fastame, M. C., Pedron, M., Porru, A., Duca, V., Hitchcott, P. K., & Penna, M. P. (2019). Metacognition and errors: The impact of self-regulatory trainings in children with specific learning disabilities. *ZDM Mathematics Education, 51*, 577-585. https://doi.org/10.1007/s11858-019-01044-w

Mansilla, N. C., Aravena Díaz, M. D., & Berres, S. (2025). Metacognitive strategies in mathematical modelling: A case study with engineering students. *International Journal of Mathematical Education in Science and Technology, 56*(11), 2212-2235. https://doi.org/10.1080/0020739X.2024.2404421

Mevarech, Z. R., & Kramarski, B. (1997). Improve: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34*(2), 365-394. https://doi.org/10.3102/00028312034002365

Miller, A. H., Espinas, D. R., McNeish, D., & Barnes, M. A. (2025). Dosage response in intensive mathematics interventions for early elementary students with or at-risk for mathematics learning disability. *Educational Psychology Review, 37*(4), Article 91. https://doi.org/10.1007/s10648-025-10070-y

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*, Article 1. https://doi.org/10.1186/2046-4053-4-1

Montague, M., Enders, C., & Dietz, S. (2011). Effects of cognitive strategy instruction on math problem solving of middle school students with learning disabilities. *Learning Disability Quarterly, 34*(4), 262-272. https://doi.org/10.1177/0731948711421762

Muncer, G., Higham, P. A., Gosling, C. J., Cortese, S., Wood-Downie, H., & Hadwin, J. A. (2022). A meta-analysis investigating the association between metacognition and math performance in adolescence. *Educational Psychology Review, 34*, 301-334. https://doi.org/10.1007/s10648-021-09620-x

Nakagawa, S., Yang, Y., Macartney, E. L., Spake, R., & Lagisz, M. (2023). Quantitative evidence synthesis: A practical guide on meta-analysis, metaregression, and publication bias tests for environmental sciences. *Environmental Evidence, 12*, Article 8. https://doi.org/10.1186/s13750-023-00301-6

Neethirajan, S., Tuteja, S. K., Huang, S.-T., & Kelton, D. (2005). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Biosensors and bioelectronics (vol. 98)*. Wiley. https://doi.org/10.1002/0470870168

Nelson, G., & McMaster, K. L. (2019). The effects of early numeracy interventions for students in preschool and early elementary: A meta-analysis. *Journal of Educational Psychology, 111*(6), 1001-1022. https://doi.org/10.1037/edu0000334

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.00422

Qudratuddarsi, H., Hidayat, R., Shah, R. L. Z. b. R. M., Nasir, N., Imami, M. K. W., & Nor, R. b. M. (2022). Rasch validation of instrument measuring gen-Z science, technology, engineering, and mathematics (STEM) application in teaching during the pandemic. *International Journal of Learning, Teaching and Educational Research, 21*(6), 104-121. https://doi.org/10.26803/ijlter.21.6.7

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

Rojo, M., Gersib, J., Powell, S. R., Shen, Z., King, S. G., Akther, S. S., Arsenault, T. L., Bos, S. E., Lariviere, D. O., & Lin, X. (2024). A meta-analysis of mathematics interventions: Examining the impacts of intervention characteristics. *Educational Psychology Review, 36*(1), Article 9. https://doi.org/10.1007/s10648-023-09843-0

Ropovik, I., Adamkovic, M., & Greger, D. (2021). Neglect of publication bias compromises meta-analytic of educational research. *PLoS ONE, 16*(6), Article Article e0252415. https://doi.org/10.1371/journal.pone.0252415

Schneider, W., & Artelt, C. (2010). Metacognition and mathematics education. *ZDM Mathematics Education, 42*, 149-161. https://doi.org/10.1007/s11858-010-0240-2

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351-371. https://doi.org/10.1007/BF02212307

Sercenia, J. C., & Prudente, M. S. (2023). Effectiveness of the metacognitive-based pedagogical intervention on mathematics achievement: A meta-analysis. *International Journal of Instruction, 16*(4), 561-578. https://doi.org/10.29333/iji.2023.16432a

Shilo, A., & Kramarski, B. (2018). Mathematical-metacognitive discourse: How can it be developed among teachers and their students? Empirical evidence from a videotaped lesson and two case studies. *ZDM Mathematics Education, 51*, 625-640. https://doi.org/10.1007/s11858-018-01016-6

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rucker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*, Article d4002. https://doi.org/10.1136/bmj.d4002

Stojanović, E., Ristić, V., McMaster, D. T., & Milanović, Z. (2016). Effect of plyometric training on vertical jump performance in female athletes: A systematic review and meta-analysis. *Sports Medicine, 47*(5), 975-986. https://doi.org/10.1007/s40279-016-0634-6

Turmuzi, M., & Lu'luilmaknun, U. (2023). The impact of online learning on the mathematics learning process in Indonesia: A meta-analysis. *Journal of Technology and Science Education, 13*(3), 694-717. https://doi.org/10.3926/jotse.2138

Turner, R. M., Bird, S. M., & Higgins, J. P. T. (2013). The impact of study size on meta-analyses: Examination of underpowered studies in Cochrane reviews. *PLoS ONE, 8*(3), Article e59202. https://doi.org/10.1371/journal.pone.0059202

Xie, Y., Zeng, F., & Yang, Y. (2024). A meta-analysis of the relationship between metacognition and academic achievement in mathematics: From preschool to university. *Acta Psychologica, 249*, Article 104486. https://doi.org/10.1016/j.actpsy.2024.104486

Zhang, Q. I., Hu, J., Wei, L., Cao, R. I., Ma, R., Song, H., & Jin, Y. I. (2019). Effects of traditional Chinese exercise on cognitive and psychological outcomes in older adults with mild cognitive impairment: A systematic review and meta-analysis. *Medicine, 98*(7), Article e14581. https://doi.org/10.1097/md.0000000000014581

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183. https://doi.org/10.3102/0002831207312909

Zohar, A., & Ben-David, A., 2008. Explict teaching of meta-strategic knowledge in authentic classroom settings. *Metacognition Learning, 3*(1), 59-82. https://doi.org/10.1007/s11409-007-9019-4

APPENDIX A

Table A1. Study characteristics

No	Reference	Country	School level	Students' characteristic	Intervention duration	Sample size	Mathematics content area	Type of metacognitive strategy	Instructional medium	Types of tasks	Outcomes	JBI level
1	Deshwal et al. (2022)	India	Secondary school	Normal	Not specified	50-100	Problem-solving	Combinations thereof	Traditional face to face	Procedural task	Achievement: [EG: 11.70±6.23], N=30; CG [3.95±5.03], N=30	High
2	Abari and Tyovenda (2021)	Nigeria	Secondary school	Normal	Not specified	> 100	Geometry	Combinations thereof	Traditional face to face	Procedural task	Achievement: [EG: 64.40±10.86], N=60; CG [58.75±10.87], N=60	High
3	Khashab et al. (2021)	Iran	Primary school	Learning disabilities	≤ 4 weeks	< 50	Word problems/ problem-solving	Combinations thereof	Traditional face to face	Problem-solving task	Achievement: [EG: 16.75±1.77], N=16; CG [13.25±1.99], N=16 Achievement: [EG: 16.43±1.78], N=16; CG [13.00±1.81], N=16	High
4	Ahdhianto et al. (2019)	Indonesia	Primary school	Normal	> 8 weeks	> 100	Arithmetic/ number sense	Combinations thereof	Traditional face to face	Problem-solving task	Achievement: [EG: 75.99±10.46], N=102; CG [63.55±11.06], N=102	High
5	Casaig (2019)	Philippines	Higher education	Normal	5-8 weeks	< 50	Algebra	Combinations thereof	Traditional face to face	Problem-solving task	Achievement: [EG: 22.68±4.15], N=19; CG [13.53±4.00], N=19	High
6	Gaylo and Dales (2017)	Philippines	Secondary school	Normal	5-8 weeks	50-100	Algebra	Combinations thereof	Traditional face to face	Procedural task	Achievement: [EG: 23.27±5.15], N=30; CG [19.20±3.80], N=30	High
7	Vula et al. (2017)	Portugal	Primary school	Normal	Not specified	> 100	Word problems/ problem-solving	Combinations thereof	Traditional face to face	Word problem task	Achievement: [EG: 4.11±1.84], N=126; CG [5.03±1.69], N=137 Achievement: [EG: 2.36±2.22], N=126; CG [1.34±1.52], N=137 Achievement: [EG: 0.94±1.36], N=126; CG [0.84±1.14], N=137 Achievement: [EG: 5.05±2.39], N=126; CG [4.63±1.62], N=137 Achievement: [EG: 5.85±2.12], N=126; CG [4.66±2.23], N=137 Achievement: [EG: 4.33±2.40], N=126; CG [3.37±2.28], N=137	High
8	Iqbal et al. (2016)	Pakistan	Secondary school	Normal	> 8 weeks	50-100	Word problems/ problem-solving	Combinations thereof	Traditional face to face	Word problem task	Achievement: [EG: 54.87±19.02], N=40; CG [46.30±16.79], N=40	High
9	Jitendra et al. (2015)	USA	Secondary school	Normal	5-8 weeks	> 100	Word problems/ problem-solving	Combinations thereof	Traditional face to face	Procedural task	Achievement: [EG: 16.52±5.27], N=943; CG [15.09±4.92], N=955 Achievement: [EG: 16.07±5.09], N=931; CG [15.22±5.02], N=916	High
10	Sahin and Kendir (2013)	Türkiye	Primary school	Normal	5-8 weeks	50-100	Geometry	Combinations thereof	Traditional face to face	Procedural task	Achievement: [EG: 82.43±6.05], N=39; CG [67.50±11.80], N=36	High
11	Nelson (2012)	USA	Secondary school	Normal	> 8 weeks	50-100	Algebra	Combinations thereof	Traditional face to face	Problem-solving task	Achievement: [EG: 6.10±1.27], N=30; CG [4.50±1.38], N=30	High
12	Özsoy and Ataman (2009)	Türkiye	Primary school	Normal	> 8 weeks	< 50	Word problems/ problem-solving	Combinations thereof	Traditional face to face	Problem-solving task	Achievement: [EG: 46.46±9.03], N=24; CG [27.83±9.63], N=23	High
13	Victor (2004)	USA	Primary school	Normal	≤ 4 weeks	50-100	Arithmetic/ number sense	Combinations thereof	Traditional face to face	Procedural task	Achievement: [EG: 113.10±11.88], N=38; CG [116.61±10.28], N=31	High
14	Desoete et al. (2003)	Belgium	Primary school	Normal	≤ 4 weeks	> 100	Word problems/ problem-solving	Planning strategies	Traditional face to face	Problem-solving task	Achievement: [EG: 35.73±2.65], N=49; CG [30.95±5.16], N=58 Achievement: [EG: 5.57±1.27], N=49; CG [5.13±2.07], N=58 Achievement: [EG: 57.42±25.78], N=49;	High

Table A1 (Continued).

No	Reference	Country	School level	Students' characteristic	Intervention duration	Sample size	Mathematics content area	Type of metacognitive strategy	Instructional medium	Types of tasks	Outcomes	JBI level
											CG [37.65±20.95], N=58	
15	Kramarski et al. (2002)	Israel	Secondary school	Low achievers	5-8 weeks	50-100	Word problems/ problem-solving	Combinations thereof	Traditional face to face	Word problem task	Achievement: [EG: 64.10±16.58], N=31; CG [41.16±11.32], N=14	High
				High achievers						Word problem task	Achievement: [EG: 86.31±9.11], N=29; CG [71.67±13.86], N=17	
16	Okita (2014)	United States	Secondary school	Normal	≤ 4 weeks	50-100	Arithmetic/ number sense	Monitoring strategies	Digital/ hybrid/ blended learning	Problem-solving task	Achievement: [EG: 120±45.28], N=31; CG [111±45.54], N=31	High
17	Shamir and Lifshitz (2013)	Israel	Pre-school	Learning disabilities	≤ 4 weeks	50-100	Arithmetic/ number sense	Combinations thereof	Digital/ hybrid/ blended learning	Procedural task	Achievement: [EG: 10.81±1.16], N=21; CG [9.17±2.01], N=20	High
18	Chen and Chiu (2015)	Taiwan	Primary school	Normal	5-8 weeks	50-100	Geometry	Combinations thereof	Digital/ hybrid/ blended learning	Procedural task	Achievement: [EG: 40.23±11.90], N=35; CG [39.30±13.43], N=37	High
19	de Kock and Harskamp (2014)	Netherlands	Primary school	Normal	> 8 weeks	> 100	Word problems/ problem-solving	Combinations thereof	Digital/ hybrid/ blended learning	Word problem task	Achievement: [EG: 7.36±2.75, N=280; CG [5.87±3.33], N=110 Achievement: [EG: 8.34±4.98, N=280; CG [6.21±4.85], N=110	High
20	Jacobs and Harskamp (2009)	Netherlands	Primary school	Normal	≤ 4 weeks	< 50	Word problems/ problem-solving	Combinations thereof	Digital/ hybrid/ blended learning	Word problem task	Achievement: [EG: 9.74±2.93, N=23; CG [7.96±4.03], N=24	High
21	Kapa (2015)	Israel	Secondary school	Normal	5-8 weeks	> 100	Word problems/ problem-solving	Combinations thereof	Digital/ hybrid/ blended learning	Word problem task	Achievement: [EG: 0.96±0.19, N=54; CG [0.72±0.46], N=53 Achievement: [EG: 4.26±0.65, N=54; CG [3.42±0.89], N=53 Achievement: [EG: 3.28±1.17, N=54; CG [2.28±1.51], N=53 Achievement: [EG: 4.04±0.58, N=54; CG [3.08±0.70], N=53	High
				Low achievers							Achievement: [EG: 1.00±0.00, N=54; CG [0.72±0.46], N=53 Achievement: [EG: 4.32±0.48, N=54; CG [3.55±0.85], N=53 Achievement: [EG: 3.16±1.11, N=54; CG [2.72±1.46], N=53 Achievement: [EG: 4.00±0.58, N=54; CG [3.24±0.69], N=53	
				High achievers							Achievement: [EG: 0.93±0.26, N=54; CG [0.71±0.46], N=53 Achievement: [EG: 4.21±0.77, N=54; CG [3.25±0.85], N=53 Achievement: [EG: 3.38±1.24, N=54; CG [1.75±1.42], N=53 Achievement: [EG: 4.07±0.59, N=54; CG [2.88±0.68], N=53	
22	Kramarski and Dudai (2009)	Israel	Secondary school	Normal	5-8 weeks	50-100	Algebra	Combinations thereof	Digital/ hybrid/ blended learning	Problem-solving task	Achievement: [EG: 86.71±19.90, N=32; CG [67.43±19.30], N=36 Achievement: [EG: 49.88±21.09, N=32; CG [24.33±16.54], N=36 Achievement: [EG: 68.3±14.3, N=32; CG [50.2±15.2], N=36	High

Table A1 (Continued).

No	Reference	Country	School level	Students' characteristic	Intervention duration	Sample size	Mathematics content area	Type of metacognitive strategy	Instructional medium	Types of tasks	Outcomes	JBI level
											Achievement: [EG: 74.48±23.87, N=32; CG [67.43±19.30], N=36 Achievement: [EG: 37.79±20.17, N=32; CG [24.33±16.54], N=36 Achievement: [EG: 60.4±13.8, N=32; CG [50.2±15.2], N=36 Achievement: [EG: 89.17±14.30, N=32; CG [86.72±14.38], N=36 Achievement: [EG: 86.43±19.90, N=32; CG [68.53±22.45], N=36 Achievement: [EG: 72.03±14.62, N=32; CG [48.34±16.71], N=36 Achievement: [EG: 56.10±14.70, N=32; CG [32.72±16.32], N=36 Achievement: [EG: 88.78±15.52, N=32; CG [86.72±14.38], N=36 Achievement: [EG: 71.39±26.20, N=32; CG [68.53±22.45], N=36 Achievement: [EG: 52.03±15.13, N=32; CG [48.34±16.71], N=36 Achievement: [EG: 56.10±14.70, N=32; CG [41.53±15.00], N=36	High
23	Kramarski and Friedman (2014)	Israel	Secondary school	Normal	≤ 4 weeks	50-100	Algebra	Combinations thereof	Digital/ hybrid/ blended learning	Problem-solving task	Achievement: [EG: 82.5±12.41, N=30; CG [62.8±21.13], N=30	High
24	Kramarski and Hirsch (2003)	Israel	Secondary school	Normal	> 8 weeks	< 50	Algebra	Combinations thereof	Digital/ hybrid/ blended learning	Procedural task	Achievement: [EG: 76.10±17.01, N=20; CG [62.61±18.80], N=23	High
25	Kramarski and Mizrachi (2006)	Israel	Secondary school	Normal	≤ 4 weeks	< 50	Word problems/ problem-solving	Combinations thereof	Digital/ hybrid/ blended learning	Modelling task	Achievement: [EG: 3.18±0.80, N=20; CG [1.53±0.36], N=23 Achievement: [EG: 2.80±1.10, N=20; CG [1.68±0.70], N=23 Achievement: [EG: 3.04±0.82, N=20; CG [1.48±0.50], N=23 Achievement: [EG: 3.10±0.91, N=20; CG [1.54±0.51], N=23	High
26	Kramarski and Mizrachi (2006)	Israel	Secondary school	Normal	≤ 4 weeks	50-100	Word problems/ problem-solving	Combinations thereof	Digital/ hybrid/ blended learning	Procedural task	Achievement: [EG: 85.17±11.63, N=20; CG [59.08±18.18], N=23 Achievement: [EG: 75.65±15.46, N=20; CG [54.08±18.26], N=23	High
27	Kramarski and Mizrachi (2006)	Israel	Secondary school	Normal	≤ 4 weeks	< 50	Word problems/ problem-solving	Combinations thereof	Digital/ hybrid/ blended learning	Modelling task	Achievement: [EG: 3.20±0.80, N=20; CG [1.42±0.38], N=23 Achievement: [EG: 2.44±0.70, N=22; CG [1.04±0.81], N=21	High
28	Maras et al. (2017)	UK	Secondary school	Autism spectrum disorder Normal	≤ 4 weeks	> 100	Arithmetic/ number sense	Combinations thereof	Digital/ hybrid/ blended learning	Procedural task	Achievement: [EG: 7.67±6.69, N=21; CG [3.21±5.13], N=19 Achievement: [EG: 7.30±5.27, N=46; CG [6.29±5.73], N=49	High
29	Alzahrani (2022)	Saudi Arabia	Higher education	Normal	≤ 4 weeks	50–100	Arithmetic/ number sense	Combinations thereof	Traditional face to face	Procedural task	Achievement: [EG: 19.31±2.66, N=32; CG [13.06±2.96], N=32	High

Table A1 (Continued).

No	Reference	Country	School level	Students' characteristic	Intervention duration	Sample size	Mathematics content area	Type of metacognitive strategy	Instructional medium	Types of tasks	Outcomes	JBI level
30	Jones et al. (2018)	UK	Primary school	Normal	5-8 weeks	50-100	Word problems/ problem-solving	Combinations thereof	Digital/ hybrid/ blended learning	Procedural task	Achievement: [EG: 104.06±11.14, N=32; CG [101.52±9.31], N=31	High
31	Valencia-Vallejo et al. (2019)	Colombia	Higher education	Normal	5-8 weeks	50-100	Arithmetic/ number sense	Combinations thereof	Digital/ hybrid/ blended learning	Procedural task	Achievement: [EG: 3.89±0.96, N=33; CG [3.25±1.09], N=34	High
32	Shilo and Kramarski (2018)	Israel	Primary school	Normal	> 8 weeks	> 100	Arithmetic/ number sense	Combinations thereof	Traditional face to face	Procedural task	Achievement: [EG: 75.19±17.32, N=412; CG [55.44±13.87], N=412	High
33	Lucangeli et al. (2019)	Italy	Primary and secondary school	Learning disabilities	> 8 weeks	50-100	Arithmetic/ number sense	Combinations thereof	Digital/ hybrid/ blended learning	Procedural task	Achievement: [EG: 4.67±1.472, N=34; CG [3.62±1.518], N=34 Achievement: [EG: 7.73±1.941, N=34; CG [7.15±1.708], N=34 Achievement: [EG: 5.82±1.286, N=34; CG [4.88±1.665], N=34 Achievement: [EG: 90.85±53.013, N=34; CG [104.15±48.215], N=34	High
34	Abdolhossini (2012)	Iran	Secondary education	Normal	Not specified	> 100	Not specified	Combinations thereof	Traditional face to face	Procedural task	Achievement: [EG: 5.245±3.249, N=100; CG [3.282±3.260], N=100	High

Table A2. Search terms used in the study

Variables	Search Terms Used in the Study	Boolean operators used
Metacognitive	("Metacognit*")	"AND" and "OR"
Mathematics achievement	("mathematics education" OR "Math OR "Arithmetic" OR "Statistics" OR "Numeracy")	
(("Metacognit*") AND ("mathematics education" OR "Math OR "Arithmetic" OR "Statistics" OR "Numeracy"))		

https://www.ejmste.com