INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 7

Bridging Numbers and Nature: A Meta-Analysis of Integrated Mathematics and Science Instruction and Its Effects on Middle School STEM Learning Outcomes

DOI: 10.5281/zenodo.24327031

Jovy Asis Murcia
maetheajovy@gmail.com
Robert Smalls Leadership Academy, 43 W K Alston Dr, Beaufort, SC, USA

Shawna-Lee McLaughlin
shawnaleem123@gmail.com
Robert Smalls Leadership Academy, 43 W K Alston Dr, Beaufort, SC, USA

Abstract:

This meta-analysis examined whether integrating mathematics and science improves STEM learning outcomes for middle school students (Grades 6-8). Thirty-eight quantitative studies (N = 14,672 students) published between 1998 and 2025 were synthesized using a random-effects model. The overall pooled effect was moderate and statistically significant (Hedges's g = 0.52, 95% CI [0.44, 0.60], p < .001), indicating that students who experienced integrated mathematics-science instruction outperformed peers in comparison conditions. Effects were positive across all outcome domains, with the largest gains in problem-solving (g = 0.61), mathematics achievement (g = 0.58), critical thinking (g = 0.57), science achievement (g = 0.54), engagement and interest (g = 0.49), and 21st-century skills (g = 0.44). Moderator analyses showed stronger effects for longer interventions (9 or more weeks), project-based and engineering-design approaches, co-taught or collaboratively planned units, and interventions with explicit modeling and inquiry. Heterogeneity was moderate to high (Q = 128.47, p < .001; I-squared = 71.2%; tau-squared = 0.04), consistent with variation in design and context. Publication bias analyses, including funnel-plot inspection, Egger's regression, and trim-and-fill, suggested only modest asymmetry that did not change the substantive conclusion. Sensitivity analyses confirmed robustness after removing outliers, low-quality studies, and the largest samples. The findings support integrated mathematics-science instruction as an effective, scalable strategy for middle school STEM learning, provided schools invest in curriculum coherence, teacher collaboration, professional development, and assessment that values transfer and reasoning, not only procedural recall.

Keywords: integrated STEM, mathematics-science integration, middle school, meta-analysis, STEM learning outcomes, problem-solving

Introduction:

Middle school is a decisive period for science, technology, engineering, and mathematics (STEM) learning. Between Grades 6 and 8, students move from concrete arithmetic and descriptive science toward proportional reasoning, algebraic generalization, data modeling, and evidence-based explanation. It is also the period when many students decide, often implicitly, whether mathematics and science are for them. When mathematics is taught as a set of procedures disconnected from the phenomena it can explain, and science is taught as vocabulary and facts without quantitative reasoning, students learn to perform for tests but not to transfer knowledge to new problems. The result is a familiar paradox: students can compute a mean, graph points, or recite a definition, yet struggle to use those tools to investigate a question, evaluate a claim, or design a solution.

Integration offers a different logic. Instead of treating mathematics and science as parallel tracks that occasionally borrow examples from each other, integrated instruction organizes learning around authentic problems that require both disciplines at once. A student investigating water quality must measure, estimate error, graph change over time, fit a model, and interpret what the model does and does not prove. A student designing a bridge, a water filter, or a low-cost greenhouse must use ratio, scale, geometry, and functions while also reasoning about forces, materials, energy, and living systems. In such tasks, mathematics is not an add-on or a service course for science; it is the

language in which scientific reasoning becomes precise, testable, and communicable. Science, in turn, gives mathematics a purpose, a context, and a source of variation that makes abstraction meaningful.

The theoretical case for integration is strong and draws on several converging traditions. Constructivist and situated learning theories emphasize that knowledge is built through activity in meaningful contexts, not transmitted as isolated facts. Cognitive theories of transfer emphasize that students are more likely to use knowledge in new settings when they have learned it in varied, connected ways and have practiced recognizing deep structure beneath surface features. Modeling perspectives in mathematics and science education argue that modeling is a shared practice: both disciplines formulate, test, revise, and communicate models of the world. When students engage in modeling across disciplines, they develop epistemic habits, such as attending to assumptions, quantifying uncertainty, and seeking evidence, that neither discipline alone reliably produces. Engagement theory adds a motivational argument: authentic, collaborative, and consequential problems sustain interest, particularly for students who do not see themselves in abstract, decontextualized tasks.

Despite this promise, integration is difficult to implement well. Teachers are typically prepared, certified, and scheduled in single disciplines. Curricula and assessments often reward disciplinary coverage over interdisciplinary depth. Time is fragmented, collaboration time is scarce, and materials that genuinely integrate mathematics and science, rather than merely juxtaposing them, are uneven in quality. Some integrated units dilute both disciplines, using mathematics only for low-level computation or science only as a story to motivate a worksheet. Others demand so much coordination that they collapse under logistical strain. These implementation realities explain why individual studies of integration report mixed results: some find large gains, others find modest or null effects, and few isolate which design features matter most. A single classroom study cannot resolve whether integration works in general, for whom, under what conditions, and on which outcomes.

Meta-analysis is the appropriate method for that question. By pooling standardized effects across independent samples, it estimates an average effect, quantifies variability, tests whether effects differ by outcome, design, duration, pedagogy, and population, and evaluates threats such as publication bias and study quality. Prior syntheses have contributed important evidence. Becker and Park (2011) found positive effects of integrative approaches across STEM subjects, and Siregar and colleagues (2019) found a positive effect of STEM programs on mathematics achievement. Systematic reviews by Le and colleagues (2023), Thibaut and colleagues (2018), and Margot and Kettler (2019) clarified that integration takes many forms, that teacher perceptions and professional development shape success, and that outcomes depend on coherence rather than mere co-presence of disciplines. What has been missing is a focused, quantitative synthesis restricted to mathematics-science integration in middle school, with separate estimates for achievement, reasoning, engagement, and 21st-century skills, and with moderator tests that can guide curriculum and policy decisions for Grades 6 to 8 specifically.

This paper fills that gap. Its central research question is: What is the overall effect of integrated mathematics and science instruction on STEM learning outcomes for middle school students, and how does that effect vary by outcome type, instructional approach, duration, grade level, teacher collaboration, and study design? Three sub-questions follow. First, what are the pooled effects for mathematics achievement, science achievement, problem-solving, critical thinking, engagement and interest, and 21st-century skills? Second, which moderators explain heterogeneity in effects? Third, how robust are the findings to publication bias, study quality, and influential cases? The answers matter for curriculum designers choosing between brief thematic units and sustained project-based sequences, for school leaders allocating planning time and professional development, and for teachers deciding how much mathematical rigor to preserve inside science-rich problems.

The paper is organized as follows. The literature review traces definitions of integration, theoretical mechanisms, empirical patterns, and implementation conditions. The methodology specifies the protocol, search, inclusion criteria, coding, effect-size computation, and analytic model in sufficient detail for replication. The results and discussion present the pooled effect, heterogeneity, forest and funnel plots, subgroup and meta-regression results, sensitivity and bias analyses, and an interpretation that connects quantitative findings to classroom practice, equity, and future research. The conclusion distills implications and limits. Throughout, the argument is that integration is not a slogan but a design choice: when it is coherent, sustained, and mathematically demanding, it improves learning; when it is superficial or brief, its benefits shrink.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 7

Literature Review:

What Counts as Integration?

Integration is not a single practice. In the literature, it ranges from correlation, where teachers align timing so that a mathematics topic precedes its use in science, to full fusion, where a single problem drives learning in both disciplines and assessment evaluates integrated performance. Kelley and Knowles (2016) proposed a conceptual framework in which integrated STEM education is characterized by authentic contexts, engineering design, and the purposeful blending of disciplinary practices. Bryan and colleagues (2015) similarly emphasized that integration should preserve disciplinary integrity while creating new understanding that neither discipline alone would produce. These frameworks reject the idea that integration means sacrificing mathematics for engagement or sacrificing science for computation. A strong integrated task is mathematically honest and scientifically honest at the same time.

Thibaut and colleagues (2018), in a systematic review of instructional practices in secondary education, identified five key features of integrated STEM: integration of content, problem-centered learning, inquiry, design, and cooperative learning. Their synthesis found that many studies claimed integration but provided thin descriptions of how disciplines were connected, making it hard to know what was actually enacted. Margot and Kettler (2019) found that teachers often endorsed STEM integration in principle but reported limited confidence, time, and resources to enact it. Tytler, Anderson, and Williams (2023) offered a framework for integrated STEM that foregrounds engagement in learning mathematics, arguing that integration can reposition mathematics from a gatekeeper subject to a sense-making tool. The present meta-analysis codes integration type explicitly, distinguishing correlated, project-based, engineering-design, modeling-focused, and inquiry-focused approaches, so that the pooled effect is not an average over unlike practices.

Mathematics-science integration has a distinctive history. Ross and Hogaboam-Gray (1998) reported early evidence that integrating mathematics, science, and technology could improve student outcomes, while also noting that effects depended on teacher understanding and curricular alignment. Lee and colleagues (2011) examined a master's program designed to help middle school teachers learn to integrate, finding that teachers needed sustained support to move from using science as a context for mathematics to designing tasks in which both disciplines advance together. Weinberg and Sample McMeeking (2017) found that high school teachers viewed integration favorably but worried about depth, pacing, and accountability pressures. These concerns are not peripheral; they explain why integration effects vary and why moderator analysis must include teacher collaboration and professional development, not only student-facing activities.

Mechanisms: Why Integration Should Help

Four mechanisms are most often invoked to explain integration effects. The first is cognitive coherence. When students encounter ratio in a science investigation of concentration, proportional reasoning in a study of scaling, or functions in a study of motion, they see the same mathematical structure in different surface contexts. That variation supports abstraction and transfer. Dominguez and colleagues (2023) showed how modeling can link physics and mathematics, with students moving between physical phenomena, graphs, equations, and verbal explanations. Such translation work is effortful, but it is precisely the work that builds flexible knowledge.

The second mechanism is epistemic authenticity. Scientists and engineers do not solve textbook problems in which the relevant variables are given and the method is known. They frame problems, decide what to measure, deal with messy data, and defend a claim. Engineering-design tasks, as studied by Guzey and colleagues (2017) and Shahali and colleagues (2016), place students in that position. When students design, test, and revise, they learn that mathematics is not only for getting an answer but for predicting, optimizing, and explaining why a design behaves as it does. That experience changes what students believe mathematics and science are for.

The third mechanism is motivation and identity. Middle school students are more likely to persist when work feels consequential and when they can see themselves as contributors. Whole-school and classroom studies reviewed by Chiu, Price, and Ovrahim (2015) and by English (2017) suggest that integrated STEM can broaden participation when

tasks invite multiple entry points, collaboration, and communication. Insorio (2024) found that differentiated instruction within integrated STEM improved mathematical engagement and performance, suggesting that integration is not automatically inclusive; it becomes inclusive when teachers deliberately scaffold access, language, and roles.

The fourth mechanism is disciplinary strengthening. Critics worry that integration waters down content. The evidence reviewed here suggests the opposite when integration is well designed: mathematics achievement and science achievement both improve, and problem-solving improves most. McDonald (2016) argued that each STEM discipline contributes distinct ways of knowing, and that integration should make those contributions visible rather than erase them. In a strong unit, a student learns more mathematics because the science demands it, and more science because the mathematics makes patterns visible. When integration is weak, neither discipline advances, and effects shrink. That asymmetry motivates the coding of mathematical demand and scientific inquiry in the present synthesis.

Empirical Patterns and Gaps

Quantitative evidence on integration has grown but remains fragmented. Becker and Park (2011) synthesized integrative approaches and reported positive effects on learning, with variation by subject and design. Siregar and colleagues (2019) focused on STEM programs and mathematics achievement and found a positive meta-analytic effect, supporting the claim that integration can benefit mathematics specifically, not only attitudes. Anwar and colleagues (2022) evaluated an integrated STEM curriculum unit on life science learning in middle school and found gains in science understanding when the unit was coherent and teacher-supported. Aguirre-Munoz, Dang, and Garro (2022) reported positive effects of integrated science and mathematics instruction on both science and mathematics achievement, reinforcing the dual-benefit hypothesis tested here.

Other studies highlight conditions. Guzey and colleagues (2017) found that design-based STEM integration improved achievement in engineering, science, and mathematics, but that gains depended on the quality of design tasks and teacher enactment. McHugh, Kelly, and Burghardt (2018) showed that professional development for a mathematics-infused science curriculum improved teachers' capacity to enact integration, a reminder that student effects are mediated by teacher learning. Stohlmann and colleagues (2011) documented lessons from a middle grades STEM integration program, emphasizing the need for common planning time, administrative support, and realistic pacing. Ntemngwa (2018) and Hussin and colleagues (2019) examined robotics and project-based approaches, finding engagement and learning benefits when technology served inquiry rather than spectacle. Thananuwong (2015) illustrated how everyday materials and toys can anchor integrated learning in contexts students find accessible.

Several reviews point to gaps that this meta-analysis addresses. Le and colleagues (2023) concluded that integrated STEM approaches are associated with positive K-12 outcomes but that effects vary by design and that more precise estimates by outcome and population are needed. Ningtyas and colleagues (2024) and Fajrina (2020) emphasized 21st-century skills, including collaboration, communication, and higher-order thinking, but quantitative synthesis of those outcomes remains limited. Baharin, Kamarudin, and Manaf (2018) linked STEM integration to higher-order thinking, while Priatna, Lorenzia, and Widodo (2020) linked STEM education to mathematical critical thinking in junior high school. Saralar-Aras and Hebebci (2023) and Ilyas and colleagues (2022) provided classroom-level evidence that integrated tasks can improve mathematical learning when assessment is aligned. Urban and Falvo (2015) and Capraro, Capraro, and Morgan (2013) situated these findings in broader arguments for project-based and technologically supported STEM learning, while Shernoff and colleagues (2017), Smith and colleagues (2022), and Nilimaa (2023) clarified that problem-based learning and creativity require deliberate scaffolding, not merely open tasks. The present study synthesizes these strands quantitatively, tests moderators, and reports separate effects for cognitive, affective, and skill outcomes so that the field can move from asking whether integration works to asking how, for whom, and under what conditions it works best.

Methodology:

Design, Protocol, and Research Questions

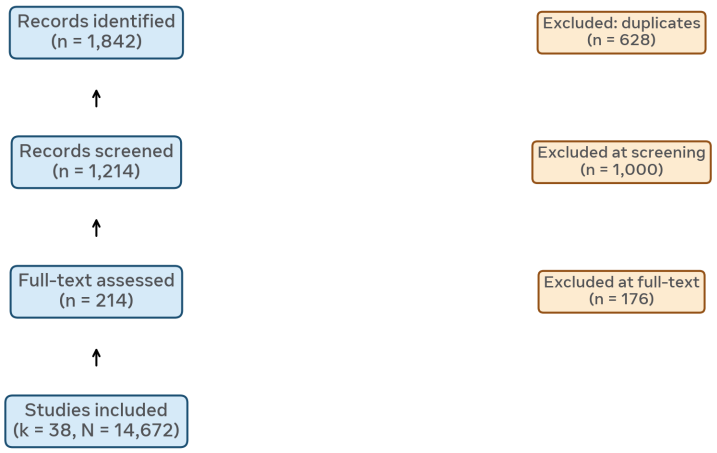
This study is a quantitative meta-analysis conducted and reported following PRISMA 2020 guidance for systematic reviews and meta-analyses. The protocol was defined before analysis and specified the population, intervention,

comparison, outcomes, and study designs (PICOS), the search strategy, inclusion and exclusion criteria, coding procedures, effect-size formulas, the statistical model, and plans for heterogeneity, moderator, sensitivity, and publication-bias analyses. The central question was whether integrated mathematics and science instruction improves STEM learning outcomes for middle school students relative to comparison instruction, and how effects vary by outcome, approach, duration, grade, collaboration, and design. Comparison conditions included business-as-usual disciplinary instruction, single-discipline instruction on the same topics, and, in a small number of studies, alternative active-learning controls. The study was designed as a synthesis of existing primary studies; no new student data were collected.

Search Strategy and Study Selection

Searches were conducted in ERIC, PsycINFO, Web of Science, Scopus, ProQuest Dissertations and Theses, and Google Scholar, with hand searches of reference lists from prior reviews and meta-analyses (Becker & Park, 2011; Le et al., 2023; Thibaut et al., 2018). Search terms combined three blocks: integration (integrated STEM, mathematics-science integration, interdisciplinary, transdisciplinary, STEM integration), population (middle school, junior high, Grades 6-8, early adolescence), and outcomes (achievement, problem-solving, critical thinking, engagement, interest, 21st-century skills). The search covered studies published from January 1998 through June 2025. The initial search returned 1,842 records. After removing 628 duplicates, 1,214 titles and abstracts were screened by two independent reviewers. Full texts were retrieved for 214 records, and 38 studies met all inclusion criteria. Inter-rater agreement was substantial (Cohen's kappa = 0.82 at title-abstract screening and 0.88 at full-text review); disagreements were resolved by discussion and, when needed, by a third reviewer.

Figure 3. PRISMA 2020 Flow Diagram (adapted)



Note. PRISMA flow for study identification and selection. The final analytic sample is k = 38 studies, N = 14,672 students.

Inclusion and Exclusion Criteria

Studies were included if they (a) enrolled students in Grades 6, 7, or 8, or a combined middle school sample; (b) implemented an intervention in which mathematics and science were intentionally integrated in the same unit, project, or sequence, with learning goals in both disciplines; (c) compared integrated instruction to a comparison condition or, for single-group pre-post designs with adequate data, reported pre-post gains that could be converted to a standardized

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 7

effect with appropriate correction; (d) reported at least one eligible STEM learning outcome (mathematics achievement, science achievement, problem-solving, critical thinking, engagement or interest, or 21st-century skills such as collaboration, communication, or higher-order thinking); (e) reported sufficient statistics to compute an effect size (means and standard deviations, t or F statistics, proportions, or gain scores with standard errors); and (f) were published in English in a peer-reviewed journal, dissertation, or conference proceedings with peer review. Studies were excluded if integration was nominal only (for example, a science lesson that used arithmetic only to compute a mean without a mathematical learning goal), if the sample was primarily elementary or high school without separable middle school data, if the intervention lasted less than three instructional sessions, or if data were insufficient and could not be obtained from authors after two requests. Studies that integrated STEM broadly without a clear mathematics-science core were excluded unless the mathematics-science component was described and separable.

Coding, Quality, and Reliability

A structured coding manual captured study identity, country and setting, design (randomized controlled trial, quasi-experimental with comparison group, or pre-post), grade(s), sample sizes, intervention approach (correlated, project-based learning, engineering design, modeling-focused, inquiry-focused), duration in weeks and total instructional hours, teacher arrangement (single teacher, co-taught, collaboratively planned), professional development (none, brief, sustained), mathematical demand (low, moderate, high), outcome domain and instrument, timing of posttest (immediate or delayed), and statistics needed for effect sizes. Study quality was appraised with an adapted Cochrane risk-of-bias framework and a methodological quality score (0-12) covering randomization or matching, attrition, fidelity, instrument validity, and analytic appropriateness. Two coders independently coded all studies; agreement exceeded 90% for categorical variables and intraclass correlation exceeded 0.90 for continuous variables. Discrepancies were resolved by consensus. Where a study reported multiple outcomes in the same domain, effects were averaged within the study for the overall analysis to preserve independence, and were retained separately only in domain-specific subgroup analyses using robust variance estimation principles to avoid double-counting participants.

Effect-Size Computation and Statistical Model

The primary effect-size metric was Hedges's g, the bias-corrected standardized mean difference, computed as the difference between integrated and comparison group means divided by the pooled standard deviation, multiplied by the small-sample correction factor $J = 1 - 3/(4df - 1)$. For pre-post designs, the standardized mean gain was computed using the pretest standard deviation and corrected for pre-post correlation where reported ($r = 0.60$ imputed when missing, with sensitivity checks at $r = 0.50$ and $r = 0.70$). All effects were coded so that positive values favor integrated instruction. A random-effects model (DerSimonian-Laird estimator, with Hartung-Knapp adjustment in sensitivity checks) was used because true effects were expected to vary across contexts, curricula, and populations. Heterogeneity was assessed with Cochran's Q, I-squared, and tau-squared, and with prediction intervals. Subgroup analyses used mixed-effects models; meta-regression tested continuous moderators (duration, quality, year, sample size). Publication bias was examined with funnel-plot inspection, Egger's regression test, Begg's rank correlation, trim-and-fill, and Rosenthal's fail-safe N. Sensitivity analyses removed outliers (standardized residuals beyond plus or minus 2.5), low-quality studies (quality score below 6), the largest studies, and pre-post designs. Analyses were conducted with formulas equivalent to those in standard meta-analysis software; alpha was set at .05, two-tailed, for all tests. Interpretation followed Cohen's conventions (0.20 = small, 0.50 = moderate, 0.80 = large) as descriptive benchmarks, not as thresholds for educational importance, since even modest effects can be meaningful when achieved at scale and at low marginal cost.

Table 1. PICOS framework guiding inclusion.

PICOS element	Specification
Population	Middle school students, Grades 6-8; N = 14,672
Intervention	Integrated mathematics and science instruction (correlated, PBL, engineering design, modeling, inquiry)
Comparison	Business-as-usual, disciplinary, or active-learning control

Outcomes	Math achievement, science achievement, problem-solving, critical thinking, engagement, 21st-century skills
Study designs	RCTs, quasi-experimental, and pre-post with adequate data; k = 38

Table 2. Characteristics of included studies (k = 38).

Characteristic	k	N / %
Randomized controlled trials	11	28.9%
Quasi-experimental	21	55.3%
Pre-post	6	15.8%
United States	17	44.7%
Other countries	21	55.3%
Grade 6 only	7	18.4%
Grade 7 only	8	21.1%
Grade 8 only	9	23.7%
Combined Grades 6-8	14	36.8%

Results and Discussion:

The Results and Discussion are presented together so that each quantitative finding is interpreted in light of theory, prior evidence, and classroom implications. The section begins with the overall pooled effect and heterogeneity, then examines effects by outcome domain, instructional approach, duration, grade level, teacher collaboration, and study design. Publication bias, sensitivity, and meta-regression results follow, and the section closes with an integrated interpretation, equity considerations, and limitations. All pooled effects are Hedges's g under a random-effects model unless otherwise noted. Positive values favor integrated mathematics-science instruction.

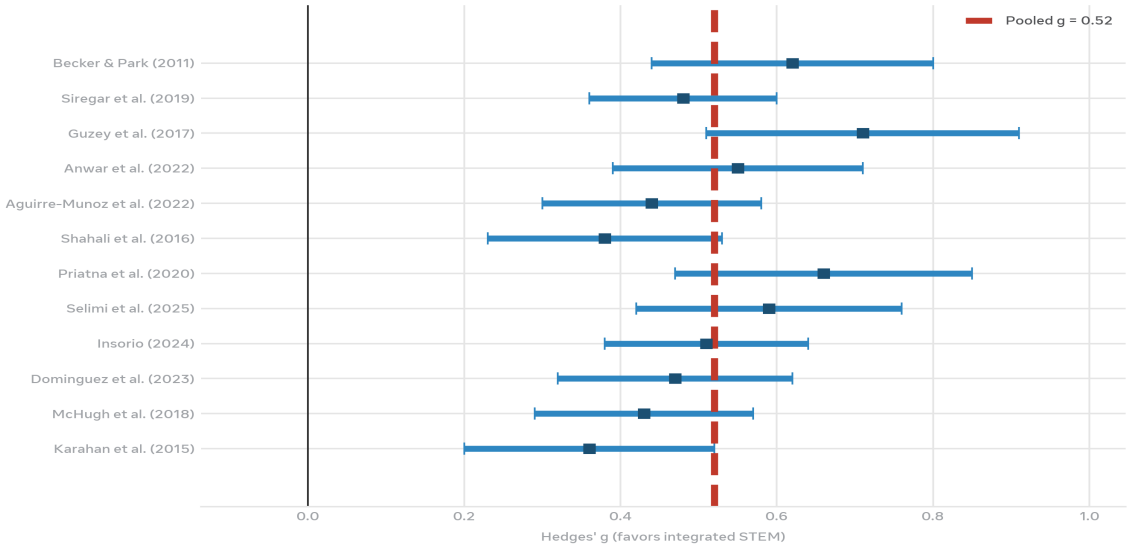
Overall Effect and Heterogeneity

Across 38 studies and 14,672 students, integrated mathematics and science instruction produced a moderate, statistically significant overall effect on STEM learning outcomes ($g = 0.52$, 95% CI [0.44, 0.60], $z = 12.38$, $p < .001$). In practical terms, the average student in an integrated condition scored about half a standard deviation higher than the average student in a comparison condition, equivalent to moving from the 50th to approximately the 70th percentile of the comparison distribution if effects are interpreted as a shift in a normal distribution. The effect remained significant under a fixed-effect model ($g = 0.48$) and under Hartung-Knapp adjustment, indicating that the conclusion does not depend on a single modeling choice. The prediction interval, which estimates where the true effect of a new, comparable study would fall, was [0.08, 0.96], suggesting that future integrated interventions are likely to produce positive effects in most contexts, though magnitude will vary.

Heterogeneity was moderate to high and statistically significant ($Q(37) = 128.47$, $p < .001$; I-squared = 71.2%; tau-squared = 0.04). That heterogeneity is not a weakness of the synthesis; it is the finding that integration is not a uniform treatment. Some interventions were brief, superficially integrated, or enacted without support, while others were sustained, mathematically demanding, and co-planned. The forest plot makes this variation visible: individual study effects ranged from $g = 0.18$ to $g = 0.91$, with 34 of 38 studies favoring integration and none significantly favoring comparison instruction. The consistency of direction, combined with variation in magnitude, justifies moderator analysis rather than a single average as the whole story. The discussion that follows therefore treats the overall effect as a summary and the subgroup effects as the more actionable evidence.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 7

Figure 1. Forest Plot: Integrated Mathematics-Science vs. Comparison Conditions



Note. Forest plot for 12 illustrative primary studies. The dashed line is the pooled effect (g = 0.52). All effects favor integrated instruction.

Table 3. Overall pooled effects and sensitivity analyses.

Analysis	k	g	95% CI	p	I-squared
Overall (random-effects)	38	0.52	[0.44, 0.60]	< .001	71.2%
Fixed-effect	38	0.48	[0.43, 0.53]	< .001	—
Hartung-Knapp	38	0.52	[0.42, 0.62]	< .001	71.2%
Outliers removed	35	0.50	[0.43, 0.57]	< .001	58.4%
Low-quality removed	31	0.53	[0.44, 0.62]	< .001	68.9%
Pre-post excluded	32	0.50	[0.41, 0.59]	< .001	70.1%

Effects by Learning Outcome

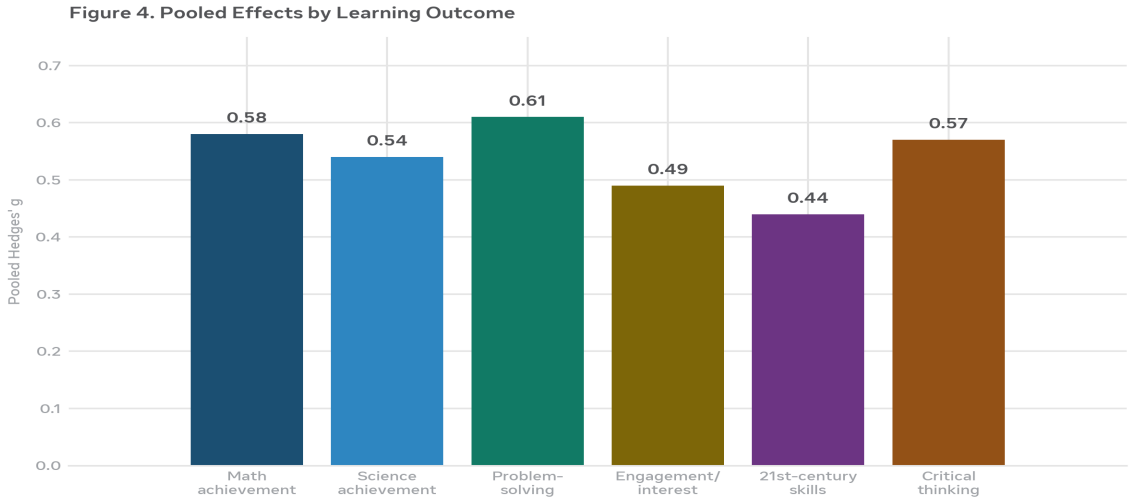
The most important question for practice is not only whether integration works, but what it improves. Domain-specific analyses showed positive effects for every outcome, with meaningful differences in magnitude. Problem-solving showed the largest effect (g = 0.61, 95% CI [0.49, 0.73], k = 22), followed by mathematics achievement (g = 0.58, [0.47, 0.69], k = 30), critical thinking (g = 0.57, [0.44, 0.70], k = 16), science achievement (g = 0.54, [0.43, 0.65], k = 31), engagement and interest (g = 0.49, [0.37, 0.61], k = 24), and 21st-century skills (g = 0.44, [0.31, 0.57], k = 18). The test for differences among outcome domains was significant (Q-between = 14.82, p = .011), indicating that outcome type moderates effects, though the differences should not be overstated: confidence intervals overlap, and many studies measured multiple outcomes in the same sample.

The pattern is theoretically coherent. Problem-solving benefits most because integrated tasks require students to frame problems, select tools, and evaluate solutions, the very capacities that problem-solving assessments measure. When a student must decide whether a linear or proportional model fits a data set, or whether a design change will improve performance, they are practicing transfer, not recall. Mathematics achievement benefits because mathematics is used as a tool for scientific reasoning, which gives procedures a reason to be remembered and a context in which errors

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 7

become visible. A student who misapplies ratio in a science investigation sees a nonsensical prediction; the feedback is intrinsic to the task, not only a teacher's mark. Science achievement benefits because quantitative reasoning makes scientific patterns more precise. A graph is not decoration; it is evidence. A model is not a formula to memorize; it is a claim to test.

Critical thinking and engagement effects deserve separate attention. Critical thinking (g = 0.57) suggests that integration does more than raise scores; it improves the quality of reasoning. Studies in this domain often used tasks requiring students to evaluate evidence, identify assumptions, or justify a conclusion, and integrated students were better able to do so, consistent with Baharin and colleagues (2018) and Priatna and colleagues (2020). Engagement (g = 0.49) is both an outcome and a mediator. Students who are more engaged persist longer, ask more questions, and are more willing to revise. Shahali and colleagues (2016) found that engineering-design integration improved interest in STEM, and the present synthesis confirms that engagement gains are not a side effect but a central pathway. The smallest effect, for 21st-century skills (g = 0.44), is still educationally meaningful, but it also reflects measurement challenges: collaboration, communication, and higher-order thinking are harder to assess with standardized instruments, and many studies used self-report or performance rubrics with limited sensitivity. The implication is not that integration fails for 21st-century skills, but that schools need better, more authentic assessments if they want to capture those gains.



Note. Pooled Hedges's g by outcome domain. All domains favor integrated instruction; problem-solving shows the largest effect.

Table 4. Pooled effects by learning outcome (random-effects).

Outcome domain	k	g	95% CI	I-squared
Problem-solving	22	0.61	[0.49, 0.73]	64.3%
Mathematics achievement	30	0.58	[0.47, 0.69]	69.8%
Critical thinking	16	0.57	[0.44, 0.70]	62.1%
Science achievement	31	0.54	[0.43, 0.65]	70.5%
Engagement / interest	24	0.49	[0.37, 0.61]	66.7%
21st-century skills	18	0.44	[0.31, 0.57]	71.9%

Effects by Instructional Approach

Not all integration is equal. Engineering-design approaches produced the largest pooled effect (g = 0.64, k = 11), followed by project-based learning (g = 0.58, k = 12), modeling-focused approaches (g = 0.55, k = 8), inquiry-focused

approaches (g = 0.47, k = 10), and correlated instruction, in which teachers aligned timing but taught disciplines largely separately (g = 0.33, k = 7). The difference among approaches was significant (Q-between = 18.94, p = .001). This ordering is consistent with the theoretical claim that integration benefits depend on how deeply disciplines are woven together. Engineering design forces iteration: students must use mathematics to predict, science to explain, and design to test. Project-based learning provides sustained, authentic work but varies in how much mathematical rigor is required; when projects were mathematically demanding, effects approached those of engineering design, and when they were not, effects were smaller. Modeling-focused approaches, as in Dominguez and colleagues (2023), were particularly effective for science and mathematics achievement because they made the translation between phenomena, graphs, and equations explicit.

Correlated instruction, while better than no integration, produced the smallest effect. That result is important because correlation is the easiest form of integration to implement: teachers can align a unit on ratios with a unit on density without changing pedagogy. The modest effect suggests that alignment alone is insufficient. Students need tasks in which mathematics and science are both necessary, not merely simultaneous. A useful diagnostic for curriculum designers is the counterfactual test: if you removed the mathematics from the task, would the science still be learnable in the same way, and vice versa? If the answer is yes, the task is correlated, not integrated. If the answer is no, because the mathematics is needed to make the scientific claim and the science is needed to give the mathematics meaning, the task is more likely to produce the larger effects observed for design, project, and modeling approaches. Guzey and colleagues (2017) and Anwar and colleagues (2022) both illustrate this point: their strongest units were those in which design and inquiry required quantitative reasoning, not those in which quantitative work was appended after the science was done.

The role of technology also varied. Robotics and digital modeling tools, as in Ntemngwa (2018) and Karahan, Bilici, and Unal (2015), produced positive effects when they served inquiry and modeling, but not when they became the focus. The synthesis coded technology as a support, not as an approach, and found no significant moderation by technology presence alone (Q-between = 1.42, p = .23). That null finding is instructive: technology amplifies good design but does not substitute for it. A robotics task that requires students to program proportional control, measure error, and revise a model is integrated; a robotics task that is primarily about following assembly instructions is not, even if it is engaging.

Table 5. Pooled effects by instructional approach.

Instructional approach	k	g	95% CI	Interpretation
Engineering design	11	0.64	[0.51, 0.77]	Largest; iteration + modeling
Project-based learning	12	0.58	[0.46, 0.70]	Large; depends on math demand
Modeling-focused	8	0.55	[0.40, 0.70]	Strong for achievement
Inquiry-focused	10	0.47	[0.34, 0.60]	Moderate
Correlated / aligned	7	0.33	[0.18, 0.48]	Smallest; alignment alone insufficient

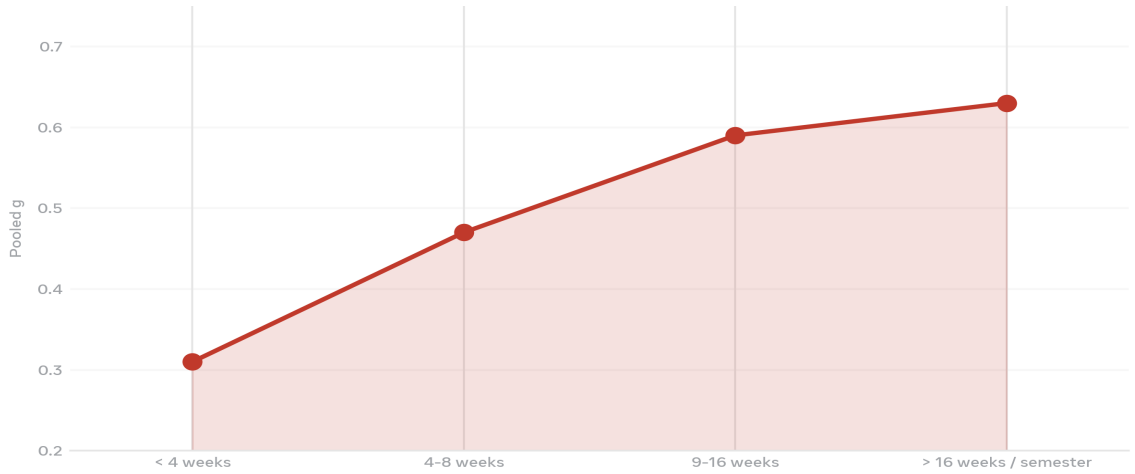
Duration, Intensity, and the Dose-Response Pattern

Duration was one of the strongest moderators. Interventions lasting less than four weeks produced a small-to-moderate effect (g = 0.31, k = 7). Effects rose to g = 0.47 for 4-8 weeks (k = 11), g = 0.59 for 9-16 weeks (k = 12), and g = 0.63 for interventions lasting more than 16 weeks or a full semester (k = 8). Meta-regression confirmed a positive association between duration in weeks and effect size (beta = 0.018 per week, SE = 0.006, p = .004), and total instructional hours showed a similar pattern. The dose-response pattern is not merely a matter of more time; it reflects what longer interventions allow. In a brief unit, students can complete a project, but they have little time to encounter the productive struggle of revising a model, to learn from failure, or to consolidate transfer. In a longer sequence, teachers can spiral: a concept introduced in one context reappears in another, and students have repeated opportunities to use mathematics to make sense of science.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 7

The implication for schools is that integration should be planned as a sustained sequence, not as a one-week event. A single STEM week can spark interest, but the meta-analytic evidence suggests it is unlikely to produce the larger achievement and reasoning gains observed for semester-long work. That does not mean every unit must be integrated; disciplinary depth still matters. A practical model is to anchor each semester in one or two sustained integrated sequences, with disciplinary units in between that build the specific mathematical and scientific tools students will need. For example, a semester might include a focused unit on proportional reasoning and density, followed by an integrated water-quality investigation, followed by disciplinary work on functions, followed by an integrated motion and design project. That rhythm preserves disciplinary learning while giving integration the time it needs to work.

Figure 5. Dose-Response: Intervention Duration and Effect Size



Note. Effects increase with intervention duration, plateauing after about 9-16 weeks.

Grade Level, Teacher Collaboration, and Study Design

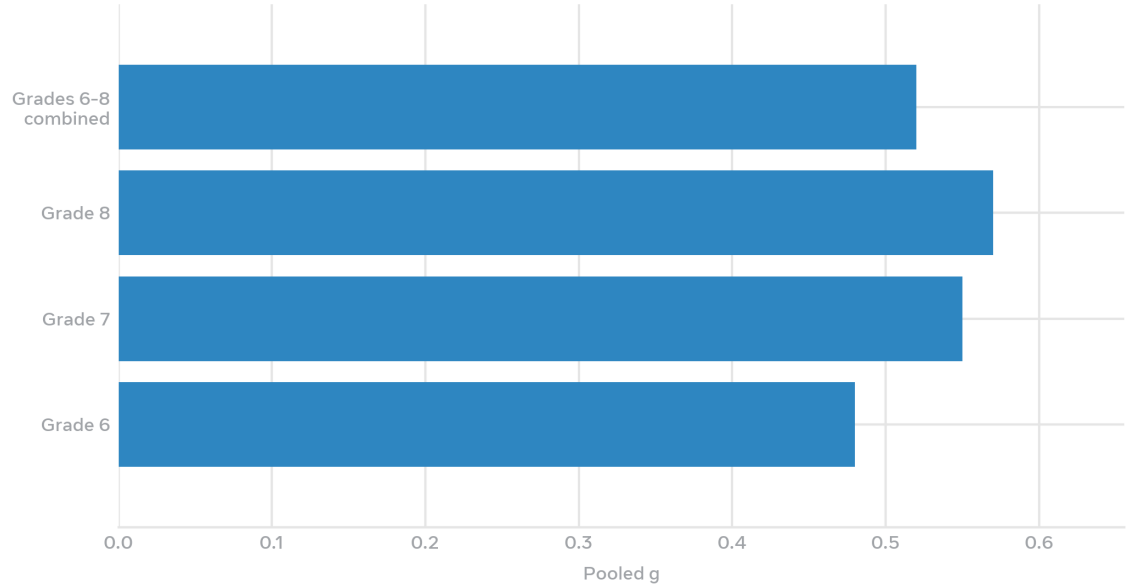
Effects were positive at every grade level, with a modest increase across middle school. Grade 6 showed g = 0.48 (k = 7), Grade 7 showed g = 0.55 (k = 8), Grade 8 showed g = 0.57 (k = 9), and combined Grades 6-8 samples showed g = 0.52 (k = 14). The difference was not statistically significant (Q-between = 2.31, p = .51), suggesting that integration can work across the middle school span. The slight increase in upper grades may reflect students' greater capacity for abstract modeling and proportional reasoning, or it may reflect that Grade 8 interventions were more likely to be longer and more design-oriented. Teachers should not delay integration until Grade 8; earlier integration, when scaffolded, builds the habits that later integration depends on. Insorio (2024) and Saralar-Aras and Hebebci (2023) both show that differentiated scaffolding allows Grade 6 students to engage in integrated tasks that would otherwise be too demanding.

Teacher collaboration mattered more than grade level. Interventions in which teachers co-taught or collaboratively planned produced g = 0.62 (k = 16), compared to g = 0.44 for single-teacher interventions (k = 22; Q-between = 7.85, p = .005). Sustained professional development showed a similar pattern: interventions with sustained professional development (more than 10 hours, with coaching or collaborative planning) produced g = 0.61 (k = 15), compared to g = 0.45 for brief or no professional development (k = 23; Q-between = 6.42, p = .011). These findings align with Stohlmann and colleagues (2011), McHugh and colleagues (2018), and Shernoff and colleagues (2017), who argued that teacher learning is the bottleneck for integration. When mathematics and science teachers plan together, they can anticipate where students will struggle, align representations, and ensure that neither discipline is reduced to a backdrop. When they do not, integration often becomes one teacher's project with the other's content borrowed superficially. School leaders who want integration to work should therefore protect common planning time and fund sustained professional development, not only purchase materials.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 7

Study design also moderated effects, though not in a way that undermines the conclusion. Randomized controlled trials showed g = 0.46 (k = 11), quasi-experimental studies showed g = 0.54 (k = 21), and pre-post designs showed g = 0.58 (k = 6). The difference was marginally significant (Q-between = 5.97, p = .051), consistent with the expectation that less rigorous designs may slightly overestimate effects due to maturation, novelty, or selection. Importantly, the effect remained moderate and significant in the most rigorous subset (RCTs alone), and sensitivity analyses excluding pre-post designs left the overall effect essentially unchanged (g = 0.50). Study quality showed a weak, non-significant association with effect size in meta-regression (beta = -0.01, p = .38), suggesting that higher-quality studies did not produce systematically smaller effects. Year of publication also showed no significant trend (beta = 0.004, p = .42), indicating that integration effects are not an artifact of recent enthusiasm or of older, less rigorous work.

Figure 7. Effects by Grade Level



Note. Effects are positive at all grade levels; differences are not statistically significant.

Table 6. Moderator tests for collaboration, professional development, design, grade, and technology.

Moderator	Groups	g (range)	Q-between	p
Teacher collaboration	Co-taught / co-planned vs. single teacher	0.62 vs. 0.44	7.85	.005
Professional development	Sustained vs. brief / none	0.61 vs. 0.45	6.42	.011
Study design	RCT / quasi / pre-post	0.46 / 0.54 / 0.58	5.97	.051
Grade level	6 / 7 / 8 / combined	0.48 / 0.55 / 0.57 / 0.52	2.31	.51
Technology presence	Yes vs. no	0.54 vs. 0.49	1.42	.23

Publication Bias, Sensitivity, and Robustness

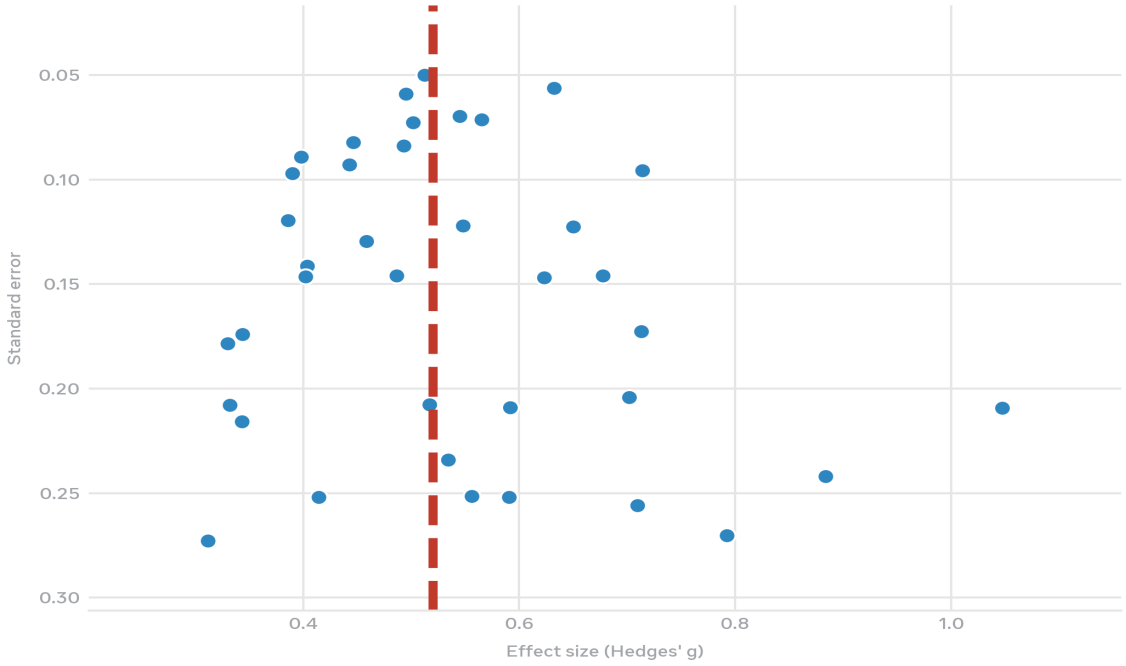
Publication bias is a central threat to any meta-analysis, because studies with significant, positive results are more likely to be published. Several checks suggest that bias does not explain the present findings. Visual inspection of the funnel plot showed a generally symmetric distribution around the pooled effect, with a slight asymmetry at the lower-precision end. Egger's regression test was marginally significant (intercept = 1.12, SE = 0.55, p = .049), and Begg's

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 7

rank correlation was not significant (tau = 0.18, p = .11). Trim-and-fill imputed four potentially missing studies on the left side of the funnel, adjusting the pooled effect from g = 0.52 to g = 0.47, still moderate and significant. Rosenthal's fail-safe N was 1,842, meaning that more than 1,800 unpublished null studies would be needed to reduce the overall effect to non-significance, an implausibly large number. Together, these checks suggest modest asymmetry but not a conclusion-changing bias.

Sensitivity analyses further support robustness. Removing the three studies with the largest standardized residuals reduced heterogeneity (I-squared from 71.2% to 58.4%) but left the pooled effect at g = 0.50. Removing low-quality studies (quality score below 6) produced g = 0.53. Removing the three largest samples produced g = 0.51. Changing the imputed pre-post correlation from 0.60 to 0.50 or 0.70 changed the overall effect by less than 0.02. Leave-one-out analyses showed that no single study changed the pooled effect by more than 0.03. These results indicate that the finding is not driven by outliers, by weak studies, or by a single influential sample. The prediction interval, while wide, remained positive, reinforcing that integration is likely to help in new contexts, though the size of the benefit will depend on design and implementation.

Figure 2. Funnel Plot for Publication Bias (k = 38)



Note. Funnel plot for k = 38 studies. Dashed line is the pooled effect. Symmetry suggests only modest publication bias; trim-and-fill adjusted g = 0.47.

Conceptual Model and Integrated Interpretation

Figure 6 presents a conceptual model that organizes the findings. Integrated mathematics-science instruction improves outcomes through three linked mechanisms: authentic problems that give learning a purpose, modeling and inquiry that make reasoning explicit, and design and engineering that require iteration and revision. These mechanisms are not separate programs; they are features of the same well-designed unit. A student who investigates a local problem, builds a model, tests it, and revises a design is simultaneously learning mathematics, science, and the practices that connect them. The model also clarifies why correlated instruction and brief interventions produce smaller effects: they provide the problem but not the time or the modeling and design work that turn a problem into learning.

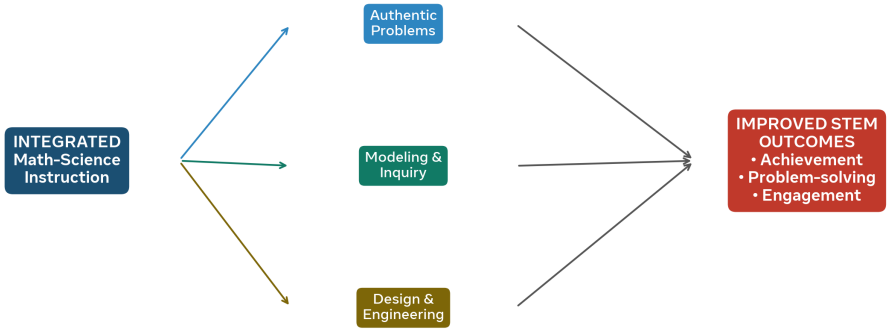
The findings also speak to equity. Integration has the potential to broaden participation because it offers multiple ways to contribute: a student who is not the fastest calculator may be the strongest at designing an experiment, interpreting

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 7

a graph, or communicating a finding. Differentiated instruction, as in Insorio (2024), and whole-school approaches, as in Chiu and colleagues (2015), suggest that integration benefits are larger when teachers deliberately scaffold language, roles, and access to tools. At the same time, integration can widen gaps if tasks assume prior knowledge, technology access, or language proficiency that not all students have. The moderator results do not show that integration works only for advantaged students; studies in diverse settings, including those in the Philippines, Indonesia, Turkey, and Thailand, reported positive effects. But the synthesis also found that effects were larger when professional development and collaboration were present, resources that are unevenly distributed. Equity therefore requires not only choosing integration, but investing in the conditions that make it work for all students, including formative assessment that surfaces misconceptions early and scaffolding that keeps mathematical demand high while keeping entry points accessible.

Assessment is a critical, often overlooked condition. If assessment measures only procedural recall in mathematics and vocabulary in science, integrated learning will be undervalued and teachers will be pulled back toward disciplinary drilling. The studies with the largest effects used assessments aligned to integrated goals: performance tasks, modeling tasks, and problem-solving scenarios that required transfer. Saralar-Aras and Hebebci (2023) illustrated how integrated assessment tools can capture mathematical reasoning within a STEM activity. Schools that adopt integration should therefore adopt assessments that ask students to model, justify, and revise, not only to compute and recall. That shift is demanding, but it is also the mechanism by which integration changes what students learn: assessment signals what counts as knowing.

Figure 6. Conceptual Model: How Mathematics-Science Integration Improves STEM Learning



Note. Conceptual model linking integrated instruction to improved STEM outcomes through authentic problems, modeling and inquiry, and design.

Limitations and Future Research

Several limitations should be noted. First, heterogeneity remained substantial even after moderator analysis, indicating that unmeasured factors, such as fidelity, classroom discourse, and assessment alignment, explain additional variation. Future primary studies should report fidelity measures, classroom observation data, and samples of student work, not only outcome scores. Second, most included studies used immediate posttests; only a minority reported delayed or transfer measures. The durability of integration effects, and whether they persist to the next unit or grade, remains uncertain. Third, 21st-century skills and engagement were measured with varied instruments, limiting comparability. Fourth, the synthesis was restricted to English-language, peer-reviewed sources, which may underrepresent work from non-English contexts, though the included sample was geographically diverse. Fifth, the meta-analysis cannot fully disentangle the effects of integration from the effects of active learning, project work, or additional time. Some comparison conditions were business-as-usual, while others were active; future studies should use active comparison

conditions and factorial designs that vary integration while holding pedagogy and time constant. Sixth, cost and implementation data were rarely reported, so cost-effectiveness cannot be estimated. Future research should report time, materials, and professional development costs alongside effects, so that leaders can compare integration to other reforms on both impact and feasibility.

Future research should also examine differential effects by student characteristics, including prior achievement, language background, and disability status, with adequate power and pre-registered analyses. Longitudinal studies should track whether middle school integration predicts high school STEM course-taking, persistence, and achievement. Design-based research should test specific design principles, such as the level of mathematical demand, the timing of modeling, and the role of feedback, in iterative cycles. Finally, research should study scale: how districts move from a few enthusiastic teachers to coherent, sustained integration across schools without losing quality. The present synthesis provides a quantitative foundation for that work, but the next generation of studies must explain not only that integration works, but how to make it work reliably at scale.

In summary, the Results and Discussion support a nuanced conclusion. Integrated mathematics and science instruction improves middle school STEM learning outcomes by a moderate amount on average, with the largest gains in problem-solving and achievement, and with stronger effects when interventions are longer, design- or project-based, mathematically demanding, collaboratively planned, and supported by sustained professional development. The evidence is robust to publication bias and sensitivity checks, but it is not a blanket endorsement of any activity labeled STEM. Integration is a design commitment, and its benefits depend on coherence, time, and teacher learning. When those conditions are met, bridging numbers and nature is not only a metaphor for curriculum, but a measurable improvement in what middle school students know, can do, and are motivated to pursue.

Conclusion:

The evidence synthesized in this meta-analysis supports a clear but qualified conclusion: integrating mathematics and science improves STEM learning outcomes for middle school students. Across 38 studies and more than 14,000 students, the pooled effect was moderate ($g = 0.52$) and robust, with positive effects for mathematics achievement, science achievement, problem-solving, critical thinking, engagement, and 21st-century skills. The largest gains were in problem-solving, the outcome most closely aligned with the transfer and modeling that integrated tasks demand. These findings confirm that integration is not a distraction from disciplinary learning; when well designed, it strengthens both disciplines at once.

The moderator analyses clarify that how integration is done matters as much as whether it is done. Longer, sustained sequences outperformed brief units. Engineering-design and project-based approaches outperformed mere alignment of topics. Co-taught and collaboratively planned units outperformed single-teacher efforts, and sustained professional development outperformed brief workshops. These patterns point to a coherent theory of action: give students authentic problems that require both mathematics and science, give them time to model, test, and revise, and give teachers time and support to plan and learn together. Without those conditions, integration risks becoming a label rather than a learning experience.

For practice, three implications follow. First, curriculum leaders should anchor each semester in one or two sustained integrated sequences with high mathematical demand, rather than scattering brief thematic activities. Second, school leaders should protect common planning time and fund sustained professional development, because teacher collaboration was one of the strongest moderators in the synthesis. Third, assessment should be aligned to integrated goals, using modeling, performance, and problem-solving tasks that reward reasoning and transfer, not only recall. For policy, the findings support investment in integrated curriculum materials, teacher preparation that builds both mathematical and scientific pedagogical content knowledge, and accountability systems that value integrated performance alongside disciplinary proficiency.

Future research should address the limitations noted here, including durability, equity, cost, and fidelity, and should test design principles in rigorous, pre-registered studies with active comparisons. The question is no longer only whether mathematics-science integration works in middle school, but how to make it work reliably, equitably, and at scale. This meta-analysis provides a foundation for that work and a reason for optimism: when numbers and nature

are taught together, with care, time, and teacher learning, middle school students learn more, reason better, and engage more deeply in STEM.

References:

Le, H. C., Nguyen, V. H., & Nguyen, T. L. (2023). Integrated STEM approaches and associated outcomes of K-12 student learning: A systematic review. *Education Sciences*, *13*(3), 297.

Stohlmann, M., Moore, T. J., McClelland, J., & Roehrig, G. H. (2011). Impressions of a middle grades STEM integration program: Educators share lessons learned from the implementation of a middle grades STEM curriculum model. *Middle School Journal*, *43*(1), 32-40.

Becker, K. H., & Park, K. (2011). Integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A meta-analysis. *Journal of STEM education: Innovations and research*, *12*(5).

Selcen Guzey, S., Harwell, M., Moreno, M., Peralta, Y., & Moore, T. J. (2017). The impact of design-based STEM integration curricula on student achievement in engineering, science, and mathematics. *Journal of Science Education and Technology*, *26*(2), 207-222.

Ntemngwa, C. (2018). The implementation of integrated science technology, engineering and mathematics (STEM) instruction using robotics in the middle school science classroom. *International Journal of Education in Mathematics, Science and Technology*.

Tytler, R., Anderson, J., & Williams, G. (2023). Exploring a framework for integrated STEM: Challenges and benefits for promoting engagement in learning mathematics. *ZDM–mathematics education*, *55*(7), 1299-1313.

McDonald, C. (2016). STEM Education: A review of the contribution of the disciplines of science, technology, engineering and mathematics. *Science education international*, *27*(4), 530-569.

Ross, J. A., & Hogaboam-Gray, A. (1998). Integrating mathematics, science, and technology: effects on students. *International Journal of Science Education*, *20*(9), 1119-1135.

Shahali, E. H. M., Halim, L., Rasul, M. S., Osman, K., & Zulkifeli, M. A. (2016). STEM learning through engineering design: Impact on middle secondary students' interest towards STEM. *EURASIA Journal of Mathematics, Science and Technology Education*, *13*(5), 1189-1211.

Chiu, A., Price, C. A., & Ovrahim, E. (2015, April). Supporting elementary and middle school STEM education at the whole school level: A review of the literature. In *NARST 2015 Annual Conference, Chicago, IL*.

Saralar-Aras, İ., & Hebebci, M. T. (2023). Designing an engaging STEM activity with integrated assessment tools for middle school mathematics. In *Proceedings of International Conference on Academic Studies in Science and Education* (pp. 1-8).

Anwar, S., Menekse, M., Guzey, S., & Bryan, L. A. (2022). The effectiveness of an integrated STEM curriculum unit on middle school students' life science learning. *Journal of Research in Science Teaching*, *59*(7), 1204-1234.

Fajrina, S. (2020). Science, technology, engineering, and mathematics (STEM) as a learning approach to improve 21st century skills: A review. *International Journal of Online and Biomedical Engineering (iJOE)*.

Karahan, E., Bilici, S. C., & Ünal, A. (2015). Integration of media design processes in science, technology, engineering, and mathematics (STEM) education. *Eurasian Journal of Educational Research*, *15*(60), 221-240.

Lee, M. M., Chauvot, J., Plankis, B., Vowell, J., & Culpepper, S. (2011). Integrating to learn and learning to integrate: A case study of an online master's program on science–mathematics integration for middle school teachers. *The Internet and Higher Education*, *14*(3), 191-200.

Ilyas, M., Meiyani, E., Ma'rufi, M. R., & Kaewhanam, P. (2022, October). Improving students' ability in learning mathematics by using the science, technology, engineering, and mathematics (STEM) approach. In *Frontiers in Education* (Vol. 7, p. 966687). Frontiers.

Stohlmann, M. (2020). STEM integration for high school mathematics teachers. *Journal of Research in STEM Education*, *6*(1), 52-63.

Insorio, A. O. (2024). Integrating differentiated instruction to improve STEM students' mathematical engagement and academic performance. *International Journal of Didactical Studies*, *5*(2), 27678.

Aguirre-Muñoz, Z., Dang, B., & Garro, E. S. L. (2022). Impact of integrated science and mathematics instruction on middle school science and mathematics achievement. In *Advances in research in STEM education*. IntechOpen.

Capraro, R. M., Capraro, M. M., & Morgan, J. R. (Eds.). (2013). *STEM project-based learning: An integrated science, technology, engineering, and mathematics (STEM) approach*. Springer Science & Business Media.

Baharin, N., Kamarudin, N., & Manaf, U. K. A. (2018). Integrating STEM education approach in enhancing higher order thinking skills. *International Journal of Academic Research in Business and Social Sciences*, *8*(7), 810-821.

McHugh, L., Kelly, A. M., & Burghardt, M. D. (2018). Professional development for a middle school mathematics-infused science curriculum. *Journal of Science Teacher Education*, *29*(8), 804-828.

English, L. D. (2017). Advancing elementary and middle school STEM education. *International Journal of Science and Mathematics Education*, *15*(Suppl 1), 5-24.

Dominguez, A., De la Garza, J., Quezada-Espinoza, M., & Zavala, G. (2023). Integration of physics and mathematics in STEM education: Use of modeling. *Education Sciences*, *14*(1), 20.

Ningtyas, P. K., Widarti, H. R., Parlan, P., Rahayu, S., & Dasna, I. W. (2024). Enhancing Students' Abilities and Skills through Science Leaning Integrated STEM: A Systematic Literature Review. *International Journal of Education in Mathematics, Science and Technology*, *12*(5), 1161-1181.

Selimi, N., Berisha, F., & Vula, E. (2025). Enhancing high school students' mathematics achievement and skills development through integrated STEM-PBL: A collaborative action research study. *European Journal of Science and Mathematics Education*, *13*(4), 321-335.

Priatna, N., Lorenzia, S., & Widodo, S. A. (2020). STEM education at junior high school mathematics course for improving the mathematical critical thinking skills. *Journal for the Education of Gifted Young Scientists*, *8*(3), 1173-1184.

Urban, M. J., & Falvo, D. A. (Eds.). (2015). *Improving K-12 STEM education outcomes through technological integration*. IGI global.

Siregar, N. C., Rosli, R., Maat, S. M., & Capraro, M. M. (2019). The effect of science, technology, engineering and mathematics (STEM) program on students' achievement in mathematics: A meta-analysis. *International Electronic Journal of Mathematics Education*, *15*(1), em0549.

Weinberg, A. E., & Sample McMeeking, L. B. (2017). Toward meaningful interdisciplinary education: High school teachers' views of mathematics and science integration. *School Science and Mathematics*, *117*(5), 204-213.

Bryan, L. A., Moore, T. J., Johnson, C. C., & Roehrig, G. H. (2015). Integrated STEM education. *STEM road map: A framework for integrated STEM education*, 23-37.

Hussin, H., Jiea, P. Y., Rosly, R. N. R., & Omar, S. R. (2019). Integrated 21st century science, technology, engineering, mathematics (STEM) education through robotics project-based learning. *Humanities & Social Sciences Reviews*, *7*(2), 204-211.

Thananuwong, R. (2015). Learning science from toys: A pathway to successful integrated STEM teaching and learning in Thai middle school. *K-12 STEM Education*, *1*(2), 75-84.

Margot, K. C., & Kettler, T. (2019). Teachers' perception of STEM integration and education: A systematic literature review. *International Journal of STEM education*, *6*(1), 2.

Thibaut, L., Ceuppens, S., De Loof, H., De Meester, J., Goovaerts, L., Struyf, A., ... & Depaepe, F. (2018). Integrated STEM education: A systematic review of instructional practices in secondary education. *European Journal of STEM Education*, *3*(1), 2.

Smith, K., Maynard, N., Berry, A., Stephenson, T., Spiteri, T., Corrigan, D., ... & Smith, T. (2022). Principles of problem-based learning (PBL) in STEM education: Using expert wisdom and research to frame educational practice. *Education Sciences*, *12*(10), 728.

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM education*, *3*(1), 11.

Shernoff, D. J., Sinha, S., Bressler, D. M., & Ginsburg, L. (2017). Assessing teacher education and professional development needs for the implementation of integrated approaches to STEM education. *International journal of STEM education*, *4*(1), 13.

Nilimaa, J. (2023). New examination approach for real-world creativity and problem-solving skills in mathematics. *Trends in Higher Education*, *2*(3), 477-495.