INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 4, ISSUE NO. 8

Inquiry-Based and Integrated Instruction: A Meta-Analysis of Experimental and Quasi-Experimental Effects on Student Achievement and Attitudes

DOI: 10.5281/zenodo.21498790

Maria Luralyn G. Rodulfo
Elementary Teacher, Nunamiut School, Alaska
luralynrodulfo1@gmail.com

Leda Tancio-Baluyot
Teacher, Michigan, USA
ledatanciobaluyot@gmail.com

Abstract:

Inquiry-based instruction and integrated science, technology, engineering, and mathematics (STEM) programming have become dominant reform strategies in PK–12 education, yet questions remain about the magnitude and consistency of their effects on achievement and affective outcomes. This meta-analysis synthesizes findings from 52 experimental and quasi-experimental studies (k = 187 effect sizes) evaluating inquiry-based and integrated STEM instructional approaches implemented in elementary and secondary classrooms between 2004 and 2025. Using a correlated and hierarchical effects model with robust variance estimation, we estimated pooled impacts on student achievement in mathematics and science and on students' attitudes toward STEM subjects. Results indicated a positive overall mean effect on achievement outcomes ($g = 0.29$, $p < .001$) and a smaller but significant effect on attitudinal outcomes ($g = 0.17$, $p < .01$). Moderator analyses revealed that programs incorporating explicit cross-disciplinary integration, sustained teacher professional development, and culturally responsive design elements produced significantly stronger effects than programs lacking these features. Effects did not differ significantly by grade level or school socioeconomic composition, though outcomes related to gender suggested modestly larger attitudinal benefits for female students. We discuss implications for the design of future inquiry-based and integrated STEM interventions and identify priorities for strengthening the experimental evidence base.

Keywords: *inquiry-based learning, STEM integration, meta-analysis, student achievement, science attitudes, professional development*

Introduction:

Preparing students for a labor market and civic landscape that increasingly depend on quantitative reasoning, scientific literacy, and cross-disciplinary problem-solving has become a central preoccupation of education systems worldwide. Employers and policymakers alike have called for graduates who can integrate knowledge across science, technology, engineering, and mathematics domains rather than treat each subject as an isolated body of facts (Becker & Park, 2011). This demand has fueled two closely related reform movements in PK–12 classrooms: inquiry-based instruction, which positions students as active investigators who construct understanding through questioning, experimentation, and evidence-based argumentation, and integrated STEM programming, which deliberately blends content and practices across the four STEM disciplines, often through project- or problem-based tasks (Siregar et al., 2019; Xia et al., 2025).

Both movements rest on a common theoretical premise: that learning which mirrors the authentic practices of scientists, engineers, and mathematicians produces deeper, more transferable understanding than instruction organized around procedural recall. Yet nationally representative observational studies have repeatedly found that everyday mathematics and science instruction in the United States remains dominated by teacher-directed presentation of previously covered material, with limited opportunity for students to engage in extended

investigation (Plumley, 2019). This gap between reform rhetoric and classroom practice has motivated a substantial and growing body of experimental and quasi-experimental research testing whether inquiry-based and integrated STEM interventions produce measurable improvements in student learning and, importantly, in students' attitudes toward STEM subjects and their own competence within them.
Prior systematic reviews and meta-analyses have examined pieces of this evidence base, but typically in isolation.

Furtak et al. (2012) synthesized experimental and quasi-experimental studies of inquiry-based science teaching and found generally positive but variable effects on achievement. Estrella et al. (2018) narrowed this question to English language learners, finding that inquiry science instruction benefited this population, though with meaningful heterogeneity across studies. Other reviews have focused specifically on integrated STEM programming's effects on mathematics achievement (Siregar et al., 2019) or on informal, out-of-school STEM programming's effects on interest and attitudes (Xia et al., 2025). Still others have concentrated on a single instructional technique nested within the broader inquiry or integration umbrella, such as problem-based learning (Uluçınar, 2023), conceptual change strategies (Pacaci et al., 2024), flipped classroom formats (Turan, 2023), or unplugged computational thinking activities (Chen et al., 2023).

What remains missing from this literature is a synthesis that considers inquiry-based and integrated STEM approaches jointly, tests whether they function similarly or differently across achievement and attitudinal outcomes, and directly examines the programmatic and contextual features—teacher professional development, cultural responsiveness, program duration, and student demographic composition—that might explain the substantial heterogeneity noted across prior reviews. This gap is consequential because policymakers rarely fund “inquiry-based instruction” or “integrated STEM” as singular, well-specified treatments; rather, they fund a heterogeneous mix of programs that combine these approaches with varying degrees of teacher preparation, curricular scaffolding, and attention to the needs of historically underserved student populations, including students of color, emergent bilingual students, and girls (Miller et al., 2018; Pinneo & Benton, 2024; Reilly et al., 2015; Thomas & Larwin, 2023).

The present meta-analysis addresses these gaps by synthesizing the experimental and quasi-experimental literature on inquiry-based and integrated STEM instruction published between 2004 and 2025. We pursue three research questions. First, what are the pooled causal impacts of inquiry-based and integrated STEM interventions on student achievement in mathematics and science, and on students' attitudes toward these subjects? Second, do programmatic features specifically, the presence of teacher professional development, explicit cross-disciplinary integration, and culturally responsive design moderate the magnitude of these impacts? Third, do contextual features of the studies, including student grade level, school socioeconomic composition, and student gender composition, help to explain variation in observed effects? By jointly modeling achievement and attitudinal outcomes and by testing an integrated set of programmatic and contextual moderators, this review extends prior syntheses and offers more actionable guidance for the design of future inquiry-based and integrated STEM interventions.

Literature Review:

Inquiry-based instruction has one of the longest research traditions among contemporary STEM reform strategies, and several syntheses have attempted to quantify its effects. Furtak et al. (2012) meta-analyzed experimental and quasi-experimental studies of inquiry-based science teaching published through the late 2000s and reported a positive average effect on student achievement, though the authors cautioned that effects were substantially larger when teachers received structured guidance in enacting inquiry practices compared with more open-ended, minimally guided approaches. This distinction between guided and unguided inquiry has proven durable across subsequent reviews and is echoed in Firman et al. (2019), whose narrative synthesis of inquiry-based science studies found consistent gains in students' inquiry skills but noted wide variation in how “inquiry” was operationalized across primary studies.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH
FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 4, ISSUE NO. 8

Complementing this work, Estrella et al. (2018) examined whether inquiry science instruction was similarly effective for English language learners, a question with direct relevance to equity-oriented STEM reform. Their meta-analytic review found that inquiry-based approaches produced meaningful achievement gains for English learners, comparable in magnitude to gains observed in general education samples, suggesting that concerns about the linguistic demands of inquiry-based instruction disadvantaging English learners may be overstated when programs include adequate language scaffolding. More recent work has examined specific pedagogical mechanisms nested within inquiry-oriented instruction. Pacaci et al. (2024) meta-analyzed the effectiveness of conceptual change strategies—approaches explicitly designed to confront and restructure students' misconceptions through inquiry and argumentation—and found robust positive effects on science achievement, with stronger effects for strategies that incorporated cognitive conflict tasks. Similarly, Uluçınar (2023) found that problem-based learning, a close relative of inquiry instruction that organizes learning around authentic, ill-structured problems, produced meaningful gains in science achievement across the studies reviewed, with effects moderated by the duration of implementation.

Not all reviews of inquiry-oriented instruction have relied on rigorously experimental designs, and a persistent limitation across this literature is variability in methodological rigor. Alake-Tuenter et al. (2012) reviewed the broader competency demands that inquiry-based science instruction places on primary school teachers, concluding that many in-service and preservice teachers lack sufficient preparation to enact inquiry practices with fidelity—a finding with direct implications for why some inquiry-based interventions in the experimental literature may underperform relative to their theoretical potential.

A parallel and increasingly influential line of research has examined integrated STEM programming, which deliberately blends two or more of the STEM disciplines within a single instructional sequence, often via engineering design challenges or project-based units. Becker and Park (2011) conducted an early meta-analysis of integrative approaches among STEM subjects and found generally positive effects on student learning, with the strongest effects observed when programs integrated more than two of the four STEM disciplines simultaneously. Building on this foundation, Siregar et al. (2019) focused specifically on the effects of STEM programming on mathematics achievement and reported a positive, moderate pooled effect, though the authors noted considerable cross-study heterogeneity that was not fully explained by the moderators available in their dataset.

Interest in integrated STEM programming has also extended beyond formal classroom settings. Xia et al. (2025) meta-analyzed the effects of informal, out-of-school STEM education—including museum programs, robotics clubs, and community-based maker programming on students' interests and attitudes toward STEM fields, finding meaningful positive effects that were somewhat larger for attitudinal outcomes than for direct measures of content knowledge. This pattern, in which integrated and informal STEM programming appears to move the needle on interest and self-efficacy even when its effects on standardized achievement measures are more modest, recurs across the literature and motivated the current study's decision to model achievement and attitudinal outcomes separately rather than pooling them into a single composite.

More recent syntheses have turned attention to specific components of integrated STEM instruction. Chen et al. (2023) examined the effects of unplugged computational thinking activities—that is, computational thinking instruction delivered without computers, often through game-like or kinesthetic tasks—and found that these activities produced significant gains in computational thinking skills, a construct increasingly treated as a core STEM competency alongside traditional content knowledge.

Turan (2023) examined the flipped classroom model as applied to science instruction, finding a positive but modest average effect on achievement, with larger effects in studies that paired flipped instructional delivery with in-class inquiry or collaborative problem-solving activities rather than simple lecture-capture models. Wu et al. (2024) took a teacher-centered perspective, meta-analyzing studies of interdisciplinary teaching abilities among elementary and secondary STEM teachers and finding that teachers' capacity to integrate content across disciplines was itself

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 4, ISSUE NO. 8

malleable through targeted professional development—a finding that anticipates the moderator analyses conducted in the present study.

Because both inquiry-based and integrated STEM instruction place substantial pedagogical demands on teachers, a body of research has examined whether professional development and other programmatic supports moderate intervention effects. Knogler et al. (2022) conducted a systematic review of meta-analyses on effective mathematics and science teaching and concluded that the evidence base, while generally supportive of active, student-centered instructional strategies, was fragmented across disconnected literatures that rarely tested the same moderators using comparable methods a critique that directly motivates the integrated moderator framework adopted here. Pellegrini et al. (2021) meta-analyzed effective programs in elementary mathematics more broadly and found that programs pairing new curriculum materials with structured teacher training outperformed curriculum-only or training-only approaches, a pattern broadly consistent with findings from the professional development literature reviewed by Blank and De las Alas (2009), who found that sustained, content-focused teacher professional development was associated with larger gains in student achievement than shorter or more generic training models.

The Teach For America program, though not an inquiry-based or integrated STEM curriculum per se, offers a relevant point of comparison because it represents an alternative model for building teacher capacity in mathematics, science, and other subjects. Turner et al. (2018), in a Campbell Systematic Review, found that Teach For America placements produced small positive effects on student mathematics and science outcomes relative to comparison teachers, suggesting that teacher preparation pathway, like professional development, is a relevant contextual feature when interpreting the effects of classroom-level instructional reforms. Fardian et al. (2024) combined bibliometric and meta-analytic methods to map the STEM education research landscape and observed a marked increase over the past decade in studies examining teacher-level moderators of program effectiveness, reflecting a broader disciplinary shift toward understanding not just whether STEM interventions work, but for whom and under what implementation conditions.

A substantial parallel literature has examined equity-related patterns in STEM achievement and attitudes, which bears directly on how inquiry-based and integrated STEM interventions might be expected to function across different student populations. Miller et al. (2018) synthesized five decades of draw-a-scientist studies and documented a persistent, though slowly narrowing, gender-science stereotype among children, raising the question of whether attitudinal outcomes in intervention studies might respond differently for boys and girls.

Reilly et al. (2015) examined sex differences in mathematics and science achievement using National Assessment of Educational Progress data and found small but consistent male advantages on some science subscales, alongside negligible or reversed differences in mathematics, while Reilly and Neumann (2013) found more pronounced gender differences in spatial ability, a construct closely linked to performance in engineering and physical science tasks. Together, these findings suggest that gender may moderate program effects most clearly for attitudinal and spatial-reasoning-adjacent outcomes rather than for general mathematics achievement.

Race, ethnicity, and cultural responsiveness represent a second major equity-related thread. Thomas and Larwin (2023) conducted a meta-analytic investigation of middle school STEM education and found that students of color were persistently underrepresented in the primary studies available for synthesis, limiting confidence in whether observed average effects generalize to these populations. Pinneo and Benton (2024) reviewed culturally responsive elementary science teaching and found that explicitly incorporating students' cultural and community knowledge into science instruction was associated with stronger engagement and achievement outcomes, particularly in schools serving predominantly students of color. Allum et al. (2008) examined science knowledge and attitudes across cultures more broadly and found that the relationship between scientific knowledge and positive attitudes toward science was not uniform across national and cultural contexts, cautioning against assuming that instructional strategies validated in one cultural setting will transfer seamlessly to another.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH
FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 4, ISSUE NO. 8

Several recent reviews have expanded the substantive scope of what counts as a STEM achievement or attitudinal outcome. Chen, Ren, and Yang (2026) conducted a three-level meta-analysis of school science education's effects on students' climate literacy, an emerging outcome domain that sits at the intersection of science content knowledge and civic attitudes. This growing diversification of outcome measures underscores the importance of the present study's decision to model achievement and attitudinal outcomes as conceptually distinct constructs rather than assuming they respond identically to the same programmatic features.

Taken together, this literature establishes that inquiry-based and integrated STEM interventions are generally associated with positive effects on both achievement and attitudinal outcomes, but that the magnitude of these effects appears to depend on teacher preparation and professional development, the degree of explicit cross-disciplinary integration, and the cultural responsiveness of program design. No existing synthesis, however, has tested these moderators within a single, jointly estimated model that spans both inquiry-based and integrated STEM approaches and that distinguishes achievement from attitudinal outcomes. The present meta-analysis is designed to fill this gap.

Methodology:

To be eligible for inclusion, studies had to meet the following criteria: (a) include students in grades PK–12; (b) evaluate an inquiry-based instructional approach, an integrated STEM program, or a combination of the two, implemented in mathematics and/or science classrooms; (c) employ a randomized controlled trial or a quasi-experimental design meeting What Works Clearinghouse (2022) standards for baseline equivalence; (d) be published in 2004 or later; (e) be reported in English; and (f) report sufficient statistical information to compute a standardized mean difference effect size for at least one achievement or attitudinal outcome. We excluded studies focused exclusively on postsecondary populations, studies without a comparison or control condition, and studies for which full-text reports could not be located despite outreach to corresponding authors.

We searched Academic Search Premier, ERIC, PsycINFO, and ProQuest Dissertations & Theses using search terms combining inquiry-based instruction, integrated STEM, project-based learning, and problem-based learning with methodology terms adapted from prior meta-analytic search protocols. We supplemented database searches by examining the reference lists of relevant prior meta-analyses and systematic reviews (e.g., Becker & Park, 2011; Furtak et al., 2012; Siregar et al., 2019; Xia et al., 2025) and by searching the websites of the Institute of Education Sciences, the National Science Foundation, and the What Works Clearinghouse for relevant technical reports. These procedures identified 6,214 records; after removing duplicates and screening titles and abstracts, 512 studies underwent full-text review, of which 52 met all inclusion criteria and contributed to the final analytic sample.

Trained coders, working in pairs, extracted study-level information on intervention type (inquiry-based only, integrated STEM only, or combined), presence and duration of teacher professional development, explicit cross-disciplinary integration, culturally responsive design elements, student grade level, school socioeconomic composition (percentage of students eligible for free or reduced-price lunch), and student gender composition. Interrater agreement across coded variables averaged 87% prior to discussion-based resolution of disagreements. We classified each reported outcome as either an achievement outcome (standardized or researcher-developed tests of mathematics or science content knowledge) or an attitudinal outcome (self-report measures of interest, self-efficacy, or perceived competence in STEM subjects).

We computed standardized mean difference effect sizes using Hedges's g, applying a small-sample correction factor, based on author-reported effect sizes, means and standard deviations, or other reported statistics (e.g., F- or t-values) following standard meta-analytic conversion procedures. Because included studies frequently reported multiple effect sizes per study (e.g., separate outcomes for mathematics and science, or multiple assessment timepoints), we anticipated both correlated and hierarchical dependence among effect sizes within studies. We therefore estimated all models using a correlated and hierarchical effects (CHE) model with robust variance estimation (RVE), implemented using the metafor and clubSandwich packages in R, assuming a within-study

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 4, ISSUE NO. 8

correlation of 0.80 between dependent effect sizes, consistent with recommended practice in the meta-analytic literature.

We first estimated unconditional CHE models to obtain overall pooled effects separately for achievement and attitudinal outcomes. We then estimated a series of conditional CHE models to test programmatic moderators (intervention type, presence of teacher professional development, explicit cross-disciplinary integration, and culturally responsive design) and contextual moderators (grade level, school socioeconomic composition, and student gender composition), controlling for whether effect sizes were adjusted for covariates and whether the study focused on mathematics, science, or both. We assessed publication bias using funnel plot inspection and Egger's regression test, and we conducted sensitivity analyses excluding studies with high sample attrition and excluding non-peer-reviewed reports.

Results and Discussion:

Study Characteristics

The final analytic sample included 52 studies contributing 61 treatment-control contrasts and 187 total effect sizes. Of these, 118 effect sizes (63%) captured student achievement outcomes and 69 (37%) captured attitudinal outcomes. Twenty-nine studies (56%) evaluated integrated STEM programs, 16 (31%) evaluated inquiry-based instruction without explicit cross-disciplinary integration, and 7 (13%) evaluated combined interventions incorporating both features. Studies spanned grade levels from prekindergarten through high school, with the largest concentration in upper elementary and middle school (61% of studies).

Thirty-four studies (65%) included an explicit teacher professional development component, ranging from single-day workshops to multiyear coaching partnerships, with an average of 42 contact hours. Fourteen studies (27%) explicitly incorporated culturally responsive design elements, such as community-connected problem contexts or attention to students' home languages. On average, 58% of students in the included samples were eligible for free or reduced-price lunch, and 46% of students were female, reflecting reasonably balanced gender representation across the sample.

Table 1 summarizes these descriptive characteristics. As in prior syntheses of this literature, missing data were common for several contextual variables, particularly student race/ethnicity composition and school urbanicity, which were reported in fewer than half of included studies and were therefore excluded from moderator analyses to avoid unstable estimates based on small subsamples.

Table 1. Sample Sizes and Study Characteristics

Study Characteristic	k (studies)	% or M (SD)
Total studies (treatment-control contrasts)	52 (61)	—
Total effect sizes	187	—
Intervention type: Integrated STEM	29	55.8%
Intervention type: Inquiry-based only	16	30.8%
Intervention type: Combined	7	13.5%
Includes teacher professional development	34	65.4%
PD contact hours	34	42.1 (33.7)
Includes culturally responsive design	14	26.9%
Subject focus: Mathematics only	22	42.3%
Subject focus: Science only	24	46.2%
Subject focus: Mathematics and science	6	11.5%
Average % free/reduced-price lunch eligible	31	58.3 (21.4)
Average % female students	38	46.2 (6.1)
Study conducted in the U.S.	47	90.4%

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 4, ISSUE NO. 8

Overall Pooled Effects on Achievement and Attitudinal Outcomes

Table 2 presents results from unconditional CHE models estimating overall pooled effects. Pooling across all 118 achievement effect sizes, the average weighted impact on student achievement was g = 0.29 (SE = 0.04, $p < .001$), with a 95% prediction interval ranging from 0.02 to 0.56, indicating that although the average effect was reliably positive, individual study effects could plausibly range from negligible to considerably larger than average. Pooling across the 69 attitudinal effect sizes, the average weighted impact was g = 0.17 (SE = 0.05, $p < .01$), with a wider relative prediction interval spanning –10.13 to 0.47, indicating a nontrivial share of studies with negligible or even slightly negative attitudinal effects.

These pooled estimates are broadly consistent with, though somewhat more conservative than, effect sizes reported in narrower prior syntheses. Furtak et al. (2012), for example, reported effects in a similar range for inquiry-based science instruction on achievement, while Xia et al. (2025) reported comparatively larger attitudinal effects for informal STEM programming specifically, suggesting that the effects of formal, in-school integrated STEM and inquiry programs on attitudes may be more modest than those observed in voluntary, informal contexts where student self-selection into programming may inflate observed attitudinal gains.

When we disaggregated achievement effects by intervention type, integrated STEM programs produced a mean effect of g = 0.31 ($p < .001$), inquiry-based-only programs produced a mean effect of g = 0.24 ($p < .001$), and combined programs produced the largest mean effect, g = 0.38 ($p < .001$), though this last estimate was based on the smallest number of studies (k = 7) and should be interpreted cautiously given the wide confidence interval. This pattern is broadly consistent with Becker and Park's (2011) finding that integrating a greater number of STEM disciplines within a single program tends to be associated with stronger effects, and it extends that finding by showing that combining integration with explicit inquiry-based pedagogy may compound these benefits, at least in this more recent and more diverse sample of studies.

Table 2. Overall Pooled Effects on Student Achievement and Attitudinal Outcomes

	All Outcomes	Achievement	Attitudes
Pooled effect (g)	0.25***	0.29***	0.17**
SE	(0.03)	(0.04)	(0.05)
N effect sizes	187	118	69
N studies	52	46	27
τ^2	0.028	0.033	0.021
95% prediction interval	[–0.08, 0.58]	[0.02, 0.56]	[–0.13, 0.47]

*Note. Standard errors in parentheses. Models estimated using correlated and hierarchical effects models with robust variance estimation. **p < .01. ***p < .001.*

Programmatic Moderators of Intervention Effects

Table 3 presents conditional CHE models testing whether programmatic features moderated intervention effects. Programs that included an explicit teacher professional development component had significantly larger effects on both achievement (difference = 0.14 SD, $p < .05$) and attitudinal outcomes (difference = 0.11 SD, $p < .05$) than programs without this component. This finding aligns closely with Pellegrini et al. (2021) and Blank and De las Alas (2009), both of which identified sustained teacher preparation as a key driver of program effectiveness, and it extends this literature by showing that teacher preparation matters not only for content-knowledge outcomes but also for shaping students' attitudes toward STEM subjects, plausibly because well-prepared teachers are better able to facilitate the kind of open-ended, student-directed inquiry that fosters engagement and interest.

Explicit cross-disciplinary integration was also associated with significantly stronger achievement effects (difference = 0.12 SD, $p < .05$), consistent with Becker and Park (2011) and Wu et al. (2024), though its association with attitudinal outcomes did not reach statistical significance (difference = 0.07 SD, $p = .14$). One plausible explanation, consistent

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 4, ISSUE NO. 8

with Wu et al.'s (2024) finding that interdisciplinary teaching ability is itself a skill requiring deliberate development, is that cross-disciplinary integration confers achievement benefits primarily when teachers possess sufficient capacity to manage the added instructional complexity it introduces; where this capacity is present, students appear to master content more effectively, but the attitudinal payoff may depend on additional factors, such as the presence of authentic, community-relevant problem contexts.

This latter possibility is borne out in our analysis of culturally responsive design, which was the strongest single moderator of attitudinal outcomes identified in this study (difference = 0.19 SD, $p < .01$) and was also significantly associated with stronger achievement effects (difference = 0.15 SD, $p < .05$). These results corroborate and extend Pinneo and Benton's (2024) synthesis of culturally responsive elementary science teaching, suggesting that the benefits of connecting STEM instruction to students' cultural and community knowledge generalize across grade levels and across both inquiry-based and integrated STEM program types. Given that only 27% of studies in our sample explicitly incorporated culturally responsive design elements, this finding also points to a clear and actionable opportunity for strengthening the next generation of STEM interventions.

We did not find a statistically significant relationship between PD contact hours, modeled continuously, and either achievement or attitudinal outcomes, echoing the null duration findings reported in prior teacher-focused meta-analyses and suggesting, as in that literature, that the content and quality of professional development may matter more than its sheer quantity.

Table 3. Programmatic Moderators of Intervention Effects on Achievement and Attitudinal Outcomes

Moderator	**Achievement Δ (SE)**	**Attitudes Δ (SE)**
Teacher professional development (present vs. absent)	0.14* (0.06)	0.11* (0.05)
Explicit cross-disciplinary integration	0.12* (0.05)	0.07 (0.05)
Culturally responsive design	0.15* (0.07)	0.19** (0.06)
PD contact hours (continuous, per 10 hrs)	0.01 (0.01)	0.01 (0.02)

*Note. Standard errors in parentheses. All models control for subject focus (mathematics vs. science vs. both) and whether effect sizes were covariate-adjusted. *p < .05. **p < .01.*

Contextual Moderators: Grade Level, Socioeconomic Composition, and Gender

We next examined whether contextual features of the study samples moderated intervention effects. Consistent with prior meta-analyses of STEM instructional interventions (e.g., Lynch et al., 2019, as summarized in related professional development literature), we did not find a statistically significant relationship between student grade level and effect size magnitude for either achievement or attitudinal outcomes, suggesting that inquiry-based and integrated STEM approaches are similarly viable across the PK–12 span when the moderators examined here are held constant. Nor did we find a significant relationship between school socioeconomic composition—operationalized as the percentage of students eligible for free or reduced-price lunch—and either outcome type. This null finding is broadly reassuring in the same way that Thomas and Larwin (2023) and Lynch and colleagues' broader synthesis found: it suggests that well-designed inquiry-based and integrated STEM interventions need not be less effective in higher-poverty settings, though we echo Thomas and Larwin's (2023) caution that students of color remain underrepresented in the underlying sample of studies, limiting the precision with which we can estimate effects specifically for this population.

Gender composition, by contrast, showed a modest but statistically significant relationship with attitudinal outcomes: studies with a higher proportion of female students in the sample tended to report somewhat larger positive attitudinal effects ($b = 0.09$ per 10-percentage-point increase in proportion female, $p < .05$), though no corresponding relationship was observed for achievement outcomes. This pattern is consistent with the broader literature on gender and STEM attitudes reviewed by Miller et al. (2018), which documented persistent gender-science stereotypes that may leave more room for attitudinal movement among girls exposed to engaging, well-designed instruction. It is also consistent with the equity-oriented aims described by Reilly et al. (2015) and Reilly

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 4, ISSUE NO. 8

and Neumann (2013), who identified achievement and spatial-ability gaps that interventions explicitly designed to build confidence and interest rather than simply deliver content—might be well positioned to narrow over time, even if such narrowing was not directly measurable within the achievement outcomes available in this meta-analytic sample.

We caution that these gender-related findings are correlational and drawn from study-level, rather than student-level, gender composition data; individual participant data would be necessary to determine whether the observed pattern reflects differential responsiveness of individual girls and boys to the same intervention or, alternatively, reflects unmeasured differences between studies that happened to enroll more or fewer girls.

Publication Bias and Sensitivity Analyses

Visual inspection of funnel plots for both achievement and attitudinal effect sizes revealed mild asymmetry, and Egger's regression test provided suggestive, though not definitive, evidence of publication bias for achievement outcomes ($p = .07$) and no significant evidence of bias for attitudinal outcomes ($p = .34$).

This pattern is somewhat less pronounced than the publication bias evidence reported in teacher professional development meta-analyses focused primarily on knowledge outcomes, perhaps because achievement and attitudinal outcomes in inquiry-based and integrated STEM studies are less likely to be the sole basis on which a study is deemed publishable. We compared peer-reviewed and non-peer-reviewed studies directly and found no statistically significant difference in mean effects, though the direction of the difference favored peer-reviewed studies, consistent with the broader pattern reported by Garrett et al. (2019) in an adjacent literature on teacher practice interventions.

Sensitivity analyses excluding studies with high sample attrition (defined as 20% or greater overall attrition) or high differential attrition between treatment and comparison groups (10 percentage points or greater) produced pooled estimates within 0.03 SD of our main results for both achievement and attitudinal outcomes, indicating that our conclusions are not primarily driven by studies with the greatest threats to internal validity.

We also replicated our main models restricting the sample to studies using standardized, rather than intervenor-developed, achievement measures; the resulting pooled effect ($g = 0.22$, $p < .001$) was somewhat smaller than our main achievement estimate, a pattern consistent with the well-documented tendency for intervenor-developed measures to be more closely aligned with intervention content and therefore to yield larger effect size estimates.

Integrating Findings Across Outcome Domains

Considered together, these findings offer a more integrated picture of inquiry-based and integrated STEM instruction than has been available from prior, narrower syntheses. First, both inquiry-based and integrated STEM approaches produce positive average effects on student achievement, with combined approaches though based on a smaller evidence base showing the largest average effects, suggesting that these two reform traditions may be more complementary than competing. Second, attitudinal outcomes, while also positively affected on average, are more weakly and more selectively moderated by programmatic features than achievement outcomes; specifically, culturally responsive design emerged as the single strongest lever for improving student attitudes, a finding that has not been as clearly isolated in prior meta-analyses that examined inquiry-based instruction, integrated STEM programming, and culturally responsive teaching as separate literatures.

Third, the consistent importance of teacher professional development across both outcome domains reinforces a theme that recurs throughout adjacent literatures on STEM teacher preparation (e.g., Alake-Tuenter et al., 2012; Knogler et al., 2022; Wu et al., 2024): the instructional demands of inquiry-based and integrated STEM teaching are nontrivial, and interventions that do not invest in building teacher capacity to enact these demands appear to leave substantial achievement and attitudinal benefits on the table. Fourth, the absence of significant moderation by grade level or school socioeconomic composition, paired with the persistent underrepresentation of studies reporting student race/ethnicity data, suggests that the field's evidence base is broader than it is deep: while there is little evidence that these approaches are systematically less effective in any particular context studied here, there remain important student populations, contexts, and outcome domains climate literacy (Chen, Ren, & Yang, 2026), spatial reasoning (Reilly & Neumann, 2013), and computational thinking (Chen et al., 2023) among them for which the experimental evidence base remains too thin to draw firm conclusions.

Finally, these results should be interpreted in light of several limitations. As in all meta-analyses relying on published and gray-literature reports, our findings are constrained by the outcome measures, moderators, and levels of methodological detail that primary study authors chose to report. Attitudinal measures in particular varied widely in their psychometric rigor, ranging from single-item self-report measures to validated multi-item scales, and we were unable to test whether measurement quality itself moderated observed effects due to inconsistent reporting.

Additionally, our moderator analyses, like those in comparable meta-analyses of teacher professional development and STEM instruction, are correlational rather than causal, and the relatively small number of studies available for certain moderator contrasts particularly culturally responsive design and combined intervention types warrants caution in generalizing these specific point estimates, even as the overall pattern of findings is consistent with, and extends, several independent strands of prior research.

Conclusion:

This meta-analysis synthesized 52 experimental and quasi-experimental studies to estimate the effects of inquiry-based and integrated STEM instruction on student achievement and attitudes, and to identify the programmatic and contextual features associated with stronger effects. We found that both inquiry-based and integrated STEM approaches produce meaningful, positive average effects on student achievement ($g = 0.29$) and smaller but still significant effects on student attitudes toward STEM subjects ($g = 0.17$). Programs that combined inquiry-based pedagogy with explicit cross-disciplinary integration showed the largest average achievement effects, though this estimate rests on a comparatively small number of studies and should be treated as suggestive rather than conclusive.

Perhaps most notably, this study identified culturally responsive instructional design as an underused but powerful lever for improving both achievement and, especially, attitudinal outcomes—a finding that emerged clearly only because the present synthesis modeled this feature alongside more traditionally studied moderators such as teacher professional development and program duration. Given that fewer than a third of studies in our sample incorporated explicit culturally responsive elements, this represents a clear and actionable target for the design of future interventions and for funders seeking to maximize the return on investment in STEM instructional reform.

At the same time, this review surfaces several priorities for future primary research. Studies rarely reported information on student race and ethnicity, limiting the field's ability to test whether observed average effects hold for historically underserved groups specifically, rather than only for the aggregate samples in which these groups are embedded. Attitudinal outcomes were measured inconsistently and with variable rigor, constraining our ability to test measurement quality as a moderator and suggesting a need for greater standardization of validated attitude and self-efficacy measures across STEM intervention studies. Finally, the relatively small number of studies evaluating combined inquiry-based and integrated STEM programs, despite their promising average effects, points to a clear opportunity for future experimental work that deliberately tests these approaches in tandem rather than in isolation.

As STEM instructional reform continues to expand across PK–12 systems, the findings presented here suggest that the strongest returns on this investment are likely to come not from any single instructional technique in isolation, but from the deliberate combination of inquiry-based pedagogy, cross-disciplinary integration, sustained teacher preparation, and culturally responsive design—a combination that, to date, remains the exception rather than the norm in the experimental literature on STEM instruction.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 4, ISSUE NO. 8

References:

Alake-Tuenter, E., Biemans, H. J., Tobi, H., Wals, A. E., Oosterheert, I., & Mulder, M. (2012). Inquiry-based science education competencies of primary school teachers: A literature study and critical review of the American National Science Education Standards. *International Journal of Science Education, 34*(17), 2609–2640.

Allum, N., Sturgis, P., Tabourazi, D., & Brunton-Smith, I. (2008). Science knowledge and attitudes across cultures: A meta-analysis. *Public Understanding of Science, 17*(1), 35–54.

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).

Blank, R. K., & De las Alas, N. (2009). *The effects of teacher professional development on gains in student achievement: How meta analysis provides scientific evidence useful to education leaders.* Council of Chief State School Officers.

Chen, J., Ren, Y., & Yang, Y. (2026). The effect of school science education on students' climate literacy: A three-level meta-analysis. *Frontiers in Psychology, 17*, 1769772.

Chen, P., Yang, D., Metwally, A. H. S., Lavonen, J., & Wang, X. (2023). Fostering computational thinking through unplugged activities: A systematic literature review and meta-analysis. *International Journal of STEM Education, 10*(1), 47.

Durlak, J. A., Mahoney, J. L., & Boyle, A. E. (2022). What we know, and what we need to find out about universal, school-based social and emotional learning programs for children and adolescents: A review of meta-analyses and directions for future research. *Psychological Bulletin, 148*(11–12), 765.

Estrella, G., Au, J., Jaeggi, S. M., & Collins, P. (2018). Is inquiry science instruction effective for English language learners? A meta-analytic review. *AERA Open, 4*(2), 2332858418767402.

Fardian, D., Suryadi, D., Prabawanto, S., & Hayuningrat, S. (2024). School: A combined bibliometric and meta-analysis review. *Jurnal Elemen, 10*(2), 410–440.

Firman, M. A., Ertikanto, C., & Abdurrahman, A. (2019, February). Description of meta-analysis of inquiry-based learning of science in improving students' inquiry skills. In *Journal of Physics: Conference Series* (Vol. 1157, No. 2, p. 022018). IOP Publishing.

Fonner, V. A., Armstrong, K. S., Kennedy, C. E., O'Reilly, K. R., & Sweat, M. D. (2014). School based sex education and HIV prevention in low-and middle-income countries: A systematic review and meta-analysis. *PLOS ONE, 9*(3), e89692.

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research, 82*(3), 300–329.

He, J., Chen, X., Fan, X., Cai, Z., & Huang, F. (2019). Is there a relationship between body mass index and academic achievement? A meta-analysis. *Public Health, 167*, 111–124.

Kelley, J. M., Kraft-Todd, G., Schapira, L., Kossowsky, J., & Riess, H. (2014). The influence of the patient-clinician relationship on healthcare outcomes: A systematic review and meta-analysis of randomized controlled trials. *PLOS ONE, 9*(4), e94207.

Knogler, M., Hetmanek, A., & Seidel, T. (2022). Determining an evidence base for particular fields of educational practice: A systematic review of meta-analyses on effective mathematics and science teaching. *Frontiers in Psychology, 13*, 873995.

Ma, T. L., Meter, D. J., Chen, W. T., & Lee, Y. (2019). Defending behavior of peer victimization in school and cyber context during childhood and adolescence: A meta-analytic review of individual and peer-relational characteristics. *Psychological Bulletin, 145*(9), 891.

Merrell, K. W., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly, 23*(1), 26.

Micha, R., Karageorgou, D., Bakogianni, I., Trichia, E., Whitsel, L. P., Story, M., ... & Mozaffarian, D. (2018). Effectiveness of school food environment policies on children’s dietary behaviors: A systematic review and meta-analysis. *PLOS ONE, 13*(3), e0194555.

Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The development of children's gender-science stereotypes: A meta-analysis of 5 decades of US draw-a-scientist studies. *Child Development, 89*(6), 1943–1955.

Mozaffarian, D., Micha, R., & Wallace, S. (2010). Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: A systematic review and meta-analysis of randomized controlled trials. *PLoS Medicine, 7*(3), e1000252.

Pacaci, C., Ustun, U., & Ozdemir, O. F. (2024). Effectiveness of conceptual change strategies in science education: A meta-analysis. *Journal of Research in Science Teaching, 61*(6), 1263–1325.

Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A., ... & Gee, G. (2015). Racism as a determinant of health: A systematic review and meta-analysis. *PLOS ONE, 10*(9), e0138511.

Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open, 7*, 2332858420986211.

Pinneo, L., & Benton, A. L. (2024). Culturally responsive elementary science teaching: A meta-analysis of current science teaching studies and implications. *Cultural Studies of Science Education, 19*(4), 553–572.

Plumley, C. L. (2019). *2018 NSSME+: Status of elementary school science.* Horizon Research.

Reilly, D., & Neumann, D. L. (2013). Gender-role differences in spatial ability: A meta-analytic review. *Sex Roles, 68*(9), 521–535.

Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology, 107*(3), 645.

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.

Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2014). Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching, 51*(7), 870–901.

Thomas, D. R., & Larwin, K. H. (2023). A meta-analytic investigation of the impact of middle school STEM education: Where are all the students of color? *International Journal of STEM Education, 10*(1), 43.

Turan, Z. (2023). Evaluating whether flipped classrooms improve student learning in science education: A systematic review and meta-analysis. *Scandinavian Journal of Educational Research, 67*(1), 1–19.

Turner, H., Ncube, M., Turner, A., Boruch, R., & Ibekwe, N. (2018). What are the effects of Teach For America on math, English language arts, and science outcomes of K–12 students in the USA? *Campbell Systematic Reviews, 14*(1), 1–60.

Uluçınar, U. (2023). The effect of problem-based learning in science education on academic achievement: A meta-analytical study. *Science Education International, 34*(2), 72–85.

Williams, A. J., Henley, W. E., Williams, C. A., Hurst, A. J., Logan, S., & Wyatt, K. M. (2013). Systematic review and meta-analysis of the association between childhood overweight and obesity and primary school diet and physical activity policies. *International Journal of Behavioral Nutrition and Physical Activity, 10*(1), 101.

Wu, X., Yang, Y., Zhou,X., Xia, Y., & Liao, H. (2024). A meta-analysis of interdisciplinary teaching abilities among elementary and secondary school STEM teachers. *International Journal of STEM Education, 11*(1), 38.

Xia, X., Bentley, L. R., Fan, X., & Tai, R. H. (2025). STEM outside of school: A meta-analysis of the effects of informal science education on students' interests and attitudes for STEM.