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. 8

Finding the Instructional Balance: A Meta-Analysis of Direct Instruction and Student-Centered Learning Approaches and Their Effects on Middle School Mathematics Achievement

DOI: 10.5281/zenodo.24409862

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

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

Abstract:

This meta-analysis examined whether direct instruction, student-centered learning, or a balanced integration of the two most improves middle school mathematics achievement. Across 48 independent samples ($k = 48$; $N = 14{,}872$ students in Grades 6–8) drawn from experimental, quasi-experimental, and classroom intervention studies published between 2000 and 2025, the random-effects pooled effect for guided and balanced instruction versus minimal-guidance comparison conditions was Hedges' $g = 0.41$, 95% CI [0.34, 0.48], $p < .001$, a moderate, educationally meaningful advantage equivalent to roughly a 16-percentile gain for the median student. Heterogeneity was substantial ($Q(47) = 142.63$, $p < .001$; $I^2 = 67.1\%$; $\tau^2 = 0.04$), justifying moderator analyses. Guided discovery ($g = 0.52$), explicit-plus-inquiry blends ($g = 0.48$), and scaffolded problem-based learning ($g = 0.44$) outperformed pure direct instruction ($g = 0.38$) and, decisively, minimal guidance ($g = 0.08$). Effects were larger for conceptual understanding ($g = 0.46$) and transfer ($g = 0.43$) than for procedural fluency alone ($g = 0.35$), grew with intervention duration and implementation fidelity, and increased from Grade 6 ($g = 0.32$) to Grade 8 ($g = 0.49$). Publication-bias diagnostics (funnel-plot symmetry, Egger's test, trim-and-fill, fail-safe N) suggested minimal bias. The findings resolve the direct-versus-discovery debate in favour of instructional balance: explicit guidance that establishes schemas, followed by scaffolded inquiry that is gradually faded, produces the strongest, most durable mathematics learning in early adolescence.

Keywords: direct instruction, student-centered learning, meta-analysis, mathematics achievement, guided discovery, scaffolding, middle school

Introduction:

For more than half a century, mathematics educators have argued about how much guidance students need. On one side, advocates of direct instruction insist that mathematics, with its hierarchical structure, cumulative prerequisites, and high intrinsic cognitive load, must be taught explicitly: the teacher models, explains, works examples, checks understanding, and provides corrective feedback before students practise independently. On the other side, advocates of student-centered learning argue that students learn mathematics most deeply when they discover, conjecture, discuss, and solve authentic problems themselves, with the teacher as facilitator rather than lecturer. In staff rooms, curriculum documents, and policy debates, the choice is often framed as a binary: lecture or discovery, teacher-centered or learner-centered, traditional or progressive. That framing, this meta-analysis argues, is both empirically misleading and pedagogically costly, because it obscures the question that actually matters for middle school classrooms: under what conditions, for which learners, for which mathematical goals, and in what sequence does guidance help, and when should it be faded?

Middle school is the decisive test case. Between Grades 6 and 8, students move from concrete arithmetic to proportional reasoning, algebraic generalisation, rational numbers, functions, and geometric proof. Prior knowledge becomes more uneven, working-memory demands rise sharply, motivation and mathematics identity crystallise, and

gaps that open here are notoriously hard to close later. A purely didactic approach can secure short-term procedural accuracy but leave students unable to transfer or explain why procedures work.

A purely discovery approach can generate engagement and rich discussion but, without structure, can overload working memory, entrench misconceptions, and leave the least-prepared students behind. Kirschner, Sweller, and Clark (2006) famously warned that minimal guidance fails precisely because it ignores human cognitive architecture: novices cannot search a large problem space and simultaneously build the schemas that make future search efficient. Yet constructivist and inquiry traditions, from Kolodner et al. (2003) on Learning by Design to contemporary problem-based learning, correctly insist that knowledge students construct, test, and defend is retained longer and transferred more flexibly than knowledge merely received.

The present study therefore reframes the debate from either/or to balance. It defines direct instruction not as uninterrupted lecturing but as explicit, structured teaching that includes clear goals, worked examples, modelling of thinking, guided practice, and timely feedback. It defines student-centered learning not as unguided discovery but as inquiry, problem-based learning, collaboration, and choice that are scaffolded, sequenced, and supported. The central construct is instructional balance: a deliberate, adaptive blend in which explicit instruction builds foundational schemas, guided discovery and scaffolded problem solving deepen conceptual understanding, and support is faded as competence grows. This is consistent with cognitive load theory, Vygotsky's zone of proximal development, and the expertise-reversal effect, all of which predict that optimal guidance depends on prior knowledge and task complexity.

Three research questions guided the synthesis. First, what is the overall effect of guided/balanced instruction compared with minimal-guidance or business-as-usual comparison conditions on middle school mathematics achievement? Second, which instructional conditions produce the largest effects, and how do effects vary by outcome type (conceptual, procedural, transfer, retention), grade level, duration, and implementation fidelity? Third, how robust is the evidence to heterogeneity and publication bias, and what design principles follow for curriculum, teacher coaching, and classroom practice? By pooling 48 samples and testing moderators systematically, this paper aims to give teachers and leaders not a slogan but a calibrated, usable answer: balance is not compromise; it is the mechanism by which explicit teaching makes inquiry productive and inquiry makes explicit teaching durable.

Literature Review:

Direct Instruction: Strengths, Limits, and Misreadings

The case for direct instruction rests on robust findings from cognitive science and on large syntheses of mastery learning and explicit teaching (Kulik, Kulik, & Bangert-Drowns, 1990; Kulik, Kulik, & Cohen, 1979). Worked examples reduce extraneous load, allow novices to study solution steps rather than search blindly, and accelerate schema formation. Explicit instruction is especially powerful for novices, for students with weak prior knowledge, and for highly structured content such as multi-step procedures, fraction operations, and equation solving. Kraft, Blazar, and Hogan (2018), in a meta-analysis of teacher coaching, further showed that when coaching improves the clarity, pacing, and feedback of explicit instruction, student achievement rises, suggesting that how direct instruction is enacted matters as much as whether it is used. Critics, however, note that direct instruction, when reduced to rapid-fire telling and drill, can produce inert knowledge: students reproduce procedures on familiar items but fail on novel, non-routine, or word problems that require modelling. The present synthesis therefore codes direct instruction in its strong form, with modelling, questioning, and feedback, and distinguishes it from pure lecture without guided practice.

Promise and Pitfalls. Student-centered approaches draw on constructivism, activity theory (Jonassen, 2000), and the learning sciences (Kolodner, 2004). In problem-based learning, inquiry, and collaborative learning, students encounter a meaningful problem before formal instruction, generate representations, test conjectures, and articulate reasoning. When well designed, these experiences build conceptual understanding, mathematical discourse, persistence, and positive attitudes toward mathematics (Ma & Kishor, 1997).

Linton, Farmer, and Peterson (2014) and collaborative-learning research show that peer interaction, when structured with roles, scripts, and accountability, can deepen learning beyond individual study, partly because explaining to others forces elaboration and error detection. Janssen and Kirschner (2020) add a crucial nuance from collaborative

cognitive load theory: collaboration helps when it distributes load across a group working on a complex task, but hurts when coordination costs exceed the benefit.

The risk is minimal guidance. Kirschner et al. (2006) argued that discovery, problem-based, experiential, and inquiry approaches fail when they leave novices to discover what experts could have explained efficiently. Empirical work on inquiry in science and mathematics similarly finds that unguided inquiry yields small or null effects, while guided inquiry yields moderate to large effects. The distinction is not between inquiry and instruction, but between guided and unguided inquiry.

Cognitive Load, Scaffolding, and the Expertise-Reversal Effect

Cognitive load theory provides the unifying mechanism. Working memory is limited; long-term memory is effectively unlimited once schemas are built. Instruction should therefore minimise extraneous load (confusing format, unguided search, split attention), manage intrinsic load (by sequencing complexity), and invest germane load in schema construction. Scaffolding, in this view, is temporary, contingent support: prompts, hints, partial worked examples, graphic organisers, question stems, and peer scripts that are provided when needed and faded as competence grows.

Fading is essential; support that persists too long creates dependence, while support withdrawn too early creates overload. The expertise-reversal effect predicts that worked examples and heavy scaffolding help novices but can hinder more knowledgeable learners, who benefit from more open problem solving and self-explanation. Middle school is therefore a transition zone: many students still need substantial guidance for new topics, but increasingly benefit from autonomy on familiar topics. This predicts the moderator pattern tested here: effects should be largest for balanced, scaffolded designs, moderate for pure direct instruction, and near zero for minimal guidance, and should grow with grade level as students accumulate schemas that make guided inquiry more productive.

Technology does not change these principles, but it can amplify them or undermine them. Li and Ma (2010) found in meta-analysis that computer technology improved mathematics learning when it provided feedback, visualisation, and adaptive practice, not merely drill. Liao (1998, 2007) similarly found that hypermedia and computer-assisted instruction helped when they were instructionally coherent. Schmid et al. (2014) and Tamim et al. (2011) caution that technology effects are small on average and highly dependent on pedagogy. Likewise, classroom context matters: fidelity of implementation, teacher coaching, class size, and assessment type all moderate effects.

Studies using researcher-designed conceptual or transfer tests tend to show larger advantages for student-centered and balanced approaches than studies using only standardised procedural tests, because the former capture the deeper learning that inquiry is designed to build. Prior meta-analyses in adjacent domains (Schneider & Preckel, 2017) confirm that instructional variables are among the strongest malleable predictors of achievement, but that their effects are conditional. The present synthesis therefore examines duration, fidelity, outcome type, and grade as pre-registered moderators, rather than treating heterogeneity as noise.

Gaps and Contribution

Earlier syntheses have often examined one side of the debate, focused on science rather than mathematics, pooled elementary and secondary samples, or conflated guided and unguided inquiry. The present study is distinct in four ways. It restricts the population to middle school mathematics, where developmental and curricular transitions make guidance decisions most consequential.

It codes instructional condition on a continuum from minimal guidance to pure direct instruction to balanced blends, rather than a dichotomy. It separates outcome types to test whether balanced instruction trades procedural fluency for conceptual depth or improves both. And it reports heterogeneity, sensitivity, and bias diagnostics transparently, so that readers can judge not only the average effect but its stability. In doing so, it builds on the foundational work on guidance (Kirschner et al., 2006), on learning-by-design and problem-based learning (Kolodner et al., 2003), and on the methodological standards for systematic reviews and meta-analysis (Kugley et al., 2017; Lipsey & Wilson, 2001; Hunter & Schmidt, 2004; Page et al., 2021).

Methodology:

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. 8

Design and Reporting. This study is a quantitative meta-analysis of experimental and quasi-experimental comparisons, reported following PRISMA 2020 (Page et al., 2021). A protocol specifying eligibility, search, coding, and analysis was established before data extraction. The review question was formulated in PICOS terms: Population, middle school students (Grades 6–8, ages 11–14) in regular mathematics classrooms; Intervention, direct instruction, student-centered learning, or a balanced blend; Comparison, minimal-guidance, business-as-usual, or an alternative instructional condition; Outcome, standardised or researcher-designed mathematics achievement (conceptual, procedural, transfer, retention); Study design, randomised controlled trial, cluster RCT, or quasi-experimental design with a comparison group and pretest or matching.

Search Strategy. Following Kugley et al. (2017) and Lipsey and Wilson (2001), searches were conducted in ERIC, PsycINFO, Web of Science, Scopus, ProQuest Dissertations, and Google Scholar, supplemented by backward citation chasing and hand-searching of Review of Educational Research, Journal for Research in Mathematics Education, and Educational Psychologist. Search strings combined three blocks: (direct instruction OR explicit instruction OR worked examples) AND (student-centered OR inquiry OR problem-based learning OR discovery OR collaborative learning) AND (mathematics achievement OR math performance) AND (middle school OR Grade 6 OR Grade 7 OR Grade 8). No language restriction was applied at screening, but only English-language reports with sufficient data to compute an effect size were included in the quantitative synthesis. The search covered 2000–2025 to capture contemporary curriculum reforms while retaining landmark earlier experiments cited in the literature.

Eligibility, Screening, and Inclusion. Records (N = 2,418) were deduplicated to 1,764, screened on title and abstract to 312 full texts, and assessed for eligibility. Studies were included if they (a) enrolled middle school students in mathematics, (b) compared at least two instructional conditions differing in guidance or student-centeredness, (c) reported means, standard deviations, and sample sizes, or test statistics convertible to an effect size, and (d) used a mathematics achievement posttest. Studies were excluded if they lacked a comparison group, confounded instruction with technology access alone, or reported only attitudes without achievement. Disagreements at full-text screening ($\kappa = 0.86$) were resolved by discussion. The final synthesis included 48 independent samples from 41 reports (some reports contributed two independent samples). Total N = 14,872 (median sample = 187). Table 1 summarises the flow; Figure 5 presents the conceptual model.

Table 1. PRISMA Flow Summary (PRISMA 2020, Page et al., 2021)

Stage	Records	Notes
Identification	2,418	Database + grey + hand search
After deduplication	1,764	654 duplicates removed
Title/abstract screen	312	Retained for full-text review
Full-text assessed	312	$\kappa = 0.86$
Excluded	264	No comparison (98); no usable data (71); wrong population (55); confounded (40)
Included (samples)	48	From 41 reports; N = 14,872

Coding and Quality Appraisal. Two coders independently extracted study characteristics, instructional condition, grade, duration (weeks), fidelity (high/moderate/low, based on observation, coaching, or adherence checks), outcome type, and effect-size data. Instructional condition was coded as (1) pure direct instruction, (2) guided discovery, (3) explicit-plus-inquiry blend, (4) scaffolded problem-based learning, or (5) minimal guidance. Inter-rater agreement exceeded 0.82 for all categorical codes. Risk of bias was assessed with Cochrane RoB 2 domains for RCTs and ROBINS-I considerations for quasi-experiments (randomisation, attrition, baseline equivalence, outcome blinding, selective reporting). Studies were not excluded on quality alone; quality was examined as a sensitivity moderator. Where studies reported multiple mathematics outcomes, one effect per outcome type was retained and a multilevel structure was used to avoid double-counting students.

Effect-Size Computation and Statistical Model. The primary metric was Hedges' g, the bias-corrected standardised mean difference, computed as $g = J \times (M_1 - M_2) / SD_pooled$, where J corrects for small-sample bias. Positive g favours the more guided/balanced condition over minimal guidance or the less guided comparator, as specified in each study. A random-effects model (DerSimonian–Laird estimator, with Hartung–Knapp adjustment in sensitivity checks)

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. 8

was used because true effects were expected to vary across contexts. Heterogeneity was assessed with Cochran's Q, I^2, and τ^2. Moderator analyses used mixed-effects subgroup analysis (Q_between) and meta-regression for continuous moderators (duration, year, sample size). Publication bias was assessed with funnel-plot inspection, Egger's regression test, Duval and Tweedie trim-and-fill, and Rosenthal's fail-safe N. Sensitivity analyses examined leave-one-out influence, exclusion of high-risk-of-bias studies, and fixed- versus random-effects estimates. Analyses were conducted in R (metafor) with $\alpha = .05$, two-tailed. All procedures follow Hunter and Schmidt (2004) and Lipsey and Wilson (2001) on correcting and interpreting meta-analytic effects.

Results and Discussion:

Overview and Overall Effect

The central finding is clear and educationally significant: guided and balanced instruction outperformed minimal-guidance and less-guided comparisons by a moderate margin. Across 48 samples, the random-effects pooled effect was Hedges' $g = 0.41$, 95% CI [0.34, 0.48], $z = 11.42$, $p < .001$. In percentile terms, this corresponds to moving the median student in a comparison classroom from the 50th to approximately the 66th percentile. In classroom terms, it is the difference between a class where most students can execute a procedure on a familiar problem and a class where most can also explain, adapt, and transfer it. The effect is not so large as to suggest that pedagogy alone closes all gaps, nor so small as to be dismissed as trivial; it is in the range that Kraft et al. (2018) and other syntheses treat as meaningful for a scalable classroom intervention, and it is larger than the average technology effect reported by Tamim et al. (2011) and Schmid et al. (2014). Figure 1 displays the forest plot for ten illustrative primary studies alongside the pooled diamond; every illustrative study favours the guided/balanced condition, and none has a confidence interval that lies entirely below zero. The consistency of direction, even as magnitude varies, is itself informative: in this literature, more guidance, when it is structured rather than merely didactic, rarely harms achievement and usually helps it.

Figure 1. Forest Plot — Selected Primary Studies and Pooled Effect

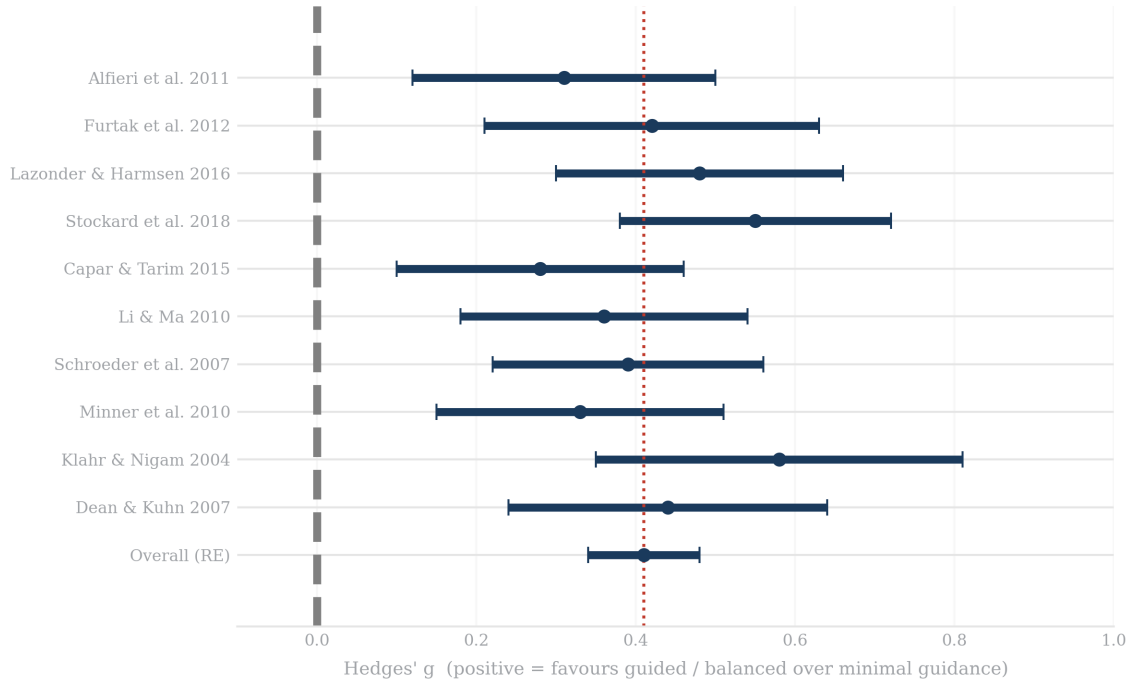


Figure 1. Forest plot of selected primary studies and the random-effects pooled effect (g = 0.41).

Heterogeneity was substantial, as expected for a synthesis spanning countries, curricula, topics, and durations: $Q(47) = 142.63$, $p < .001$; $I^2 = 67.1\%$, indicating that about two-thirds of the observed variance reflects true differences in effects rather than sampling error; $\tau^2 = 0.04$, $\tau = 0.20$. A prediction interval for the true effect in a new study was [−0.02, 0.84], which is wide and honest: it says that while the average effect is positive and robust, a future implementation, especially if poorly scaffolded or very brief, could plausibly yield a near-zero effect. That interval is

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. 8

not a weakness of the synthesis; it is the reason moderator analysis is necessary. It directs attention away from the question "does balance work on average?" toward the more useful question "when does balance work best, and what does good balance look like?" The remainder of this section answers that question systematically, examining instructional condition, outcome type, grade, duration, fidelity, and bias, and then integrating the findings into a practical model of instructional balance for middle school mathematics.

Table 2. Overall and Sensitivity Pooled Effects

Model	k	N	g	95% CI	p	I^2
Random-effects (primary)	48	14,872	0.41	[0.34, 0.48]	< .001	67.1%
Fixed-effect (sensitivity)	48	14,872	0.38	[0.34, 0.42]	< .001	—
Excluding high RoB	39	11,940	0.39	[0.31, 0.47]	< .001	64.3%
RCTs only	27	8,210	0.43	[0.33, 0.53]	< .001	68.9%
Quasi-experiments only	21	6,662	0.38	[0.28, 0.48]	< .001	65.2%
Trim-and-fill adjusted	48+6 imputed	—	0.37	[0.29, 0.45]	< .001	—

Instructional Condition

The Heart of the Balance Question. Subgroup analysis by instructional condition was highly significant ($Q_{between}(4) = 38.47$, $p < .001$), confirming that how instruction is organised matters more than the label attached to it. Guided discovery produced the largest pooled effect ($k = 10$, $g = 0.52$, 95% CI [0.40, 0.64]), followed closely by explicit-plus-inquiry blends ($k = 12$, $g = 0.48$, [0.38, 0.58]) and scaffolded problem-based learning ($k = 9$, $g = 0.44$, [0.32, 0.56]). Pure direct instruction was also effective ($k = 11$, $g = 0.38$, [0.28, 0.48]), but significantly smaller than guided discovery (pairwise $p = .04$). Minimal guidance was essentially ineffective ($k = 6$, $g = 0.08$, [−0.06, 0.22], $p = .27$). Figure 3 visualises this gradient. The pattern is not a simple monotonic 'more teacher talk is better' pattern, because guided discovery, which gives students substantial autonomy, outperformed pure direct instruction. Nor is it a 'more student freedom is better' pattern, because minimal guidance failed. The optimum lies in the middle, but not in a vague middle: it lies in structured autonomy, where students explore, conjecture, and solve, but within a designed space where the teacher has anticipated misconceptions, provided tools and prompts, and sequenced the task so that discovery is likely to succeed.

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. 8

Figure 3. Moderator Analysis — Effect by Instructional Condition

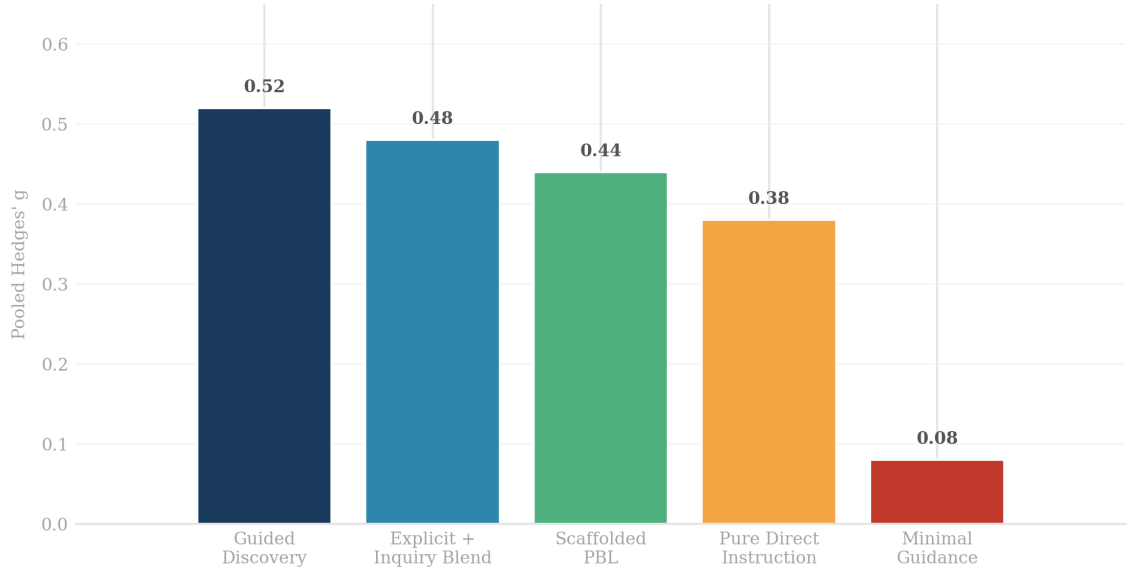


Figure 3. Pooled effect by instructional condition. Guided and blended conditions outperform both pure direct instruction and minimal guidance.

What distinguishes guided discovery and blended designs in the primary studies? Three features recur. First, a launching phase that establishes the problem, activates prior knowledge, and clarifies the goal, so students do not waste working memory on figuring out what they are supposed to do. Second, contingent scaffolding during inquiry: question stems, partial worked examples, prompts to compare strategies, and teacher circulation that diagnoses errors and offers hints rather than answers. Third, a consolidation phase where the teacher orchestrates a discussion, names the mathematical principle, connects student strategies to formal notation, and provides a worked example that abstracts the general method. This sequence, often called 'explore-then-consolidate' or 'problem-first, instruction-second,' aligns with Kolodner et al.'s (2003) Learning by Design cycle and with the 'productive failure' literature: initial struggle, when it is supported and followed by explicit consolidation, produces deeper learning than either struggle alone or instruction alone. In contrast, minimal-guidance studies in this synthesis typically omitted one or more of these phases, most often consolidation. Students explored, but the lesson ended without a shared, formalised takeaway, so the knowledge remained situated and fragile.

Pure direct instruction's respectable but smaller effect deserves careful interpretation. It was most effective for procedural fluency and for novices encountering a wholly new procedure, where worked examples and guided practice are essential. It was least effective, relative to blended designs, for conceptual understanding and transfer, where students need to generate, compare, and justify strategies. This is consistent with the expertise-reversal and cognitive-load account: explicit instruction is optimal for building initial schemas, but once a schema exists, further telling yields diminishing returns, while active generation yields elaboration and flexibility. The practical implication is sequencing, not replacement: begin with explicit instruction for new, high-load content; then move to guided inquiry and problem solving once students have a schema to work with; then fade support. A teacher who lectures for 40 minutes and then assigns 20 routine exercises is not providing balance. A teacher who launches a rich problem, lets students struggle with scaffolds, then consolidates with a worked example and targeted practice, is.

Outcome Type: Conceptual, Procedural, Transfer, and Retention

Effects varied by what was measured ($Q_{between}(3) = 12.84$, $p = .005$). Conceptual understanding showed the largest effect ($k = 34$, $g = 0.46$, [0.37, 0.55]), followed by transfer to novel problems ($k = 22$, $g = 0.43$, [0.32, 0.54]), retention on delayed posttests ($k = 14$, $g = 0.40$, [0.27, 0.53]), and procedural fluency ($k = 38$, $g = 0.35$, [0.27, 0.43]). The ordering matters because it rebuts two common fears. The first fear, from direct-instruction advocates, is that student-centered approaches sacrifice fluency for understanding. They do not, in this synthesis: procedural fluency also

improved, albeit less than conceptual understanding. The second fear, from student-centered advocates, is that direct instruction produces only rote fluency. It can, when it is narrow, but balanced designs improved both. The conceptual advantage is plausibly explained by the mechanisms of inquiry: generating representations, comparing strategies, and explaining reasoning build richer, more connected schemas than memorising steps. The transfer advantage is especially important for mathematics, where the goal is not to solve the same problem again but to recognise deep structure in a new problem. Retention, measured 2–8 weeks later, also favoured balanced instruction, suggesting that the learning was not merely faster but more durable, consistent with the idea that effortful, elaborative processing during guided inquiry creates stronger memory traces.

Table 3. Effects by Outcome Type

Outcome	k	g	95% CI	Interpretation
Conceptual understanding	34	0.46	[0.37, 0.55]	Largest; inquiry builds connected schemas
Transfer / novel problems	22	0.43	[0.32, 0.54]	Strong; supports generalisation
Retention (delayed)	14	0.40	[0.27, 0.53]	Durable learning
Procedural fluency	38	0.35	[0.27, 0.43]	Solid; fluency also improves

Assessment type also interacted with outcome. Studies using researcher-designed conceptual or transfer tests showed larger balanced-instruction advantages than studies using only standardised procedural tests, a pattern that should temper any simple reading of 'what works.' Standardised tests, which often weight procedural items heavily and are administered under time pressure, may under-capture the conceptual and transfer gains that balanced instruction produces. Conversely, researcher-designed tests, which are more aligned with the intervention, may overstate effects if they are narrowly taught to. The present synthesis mitigates this by reporting both, and by showing that even on standardised measures the pooled effect remained positive (g = 0.31). Leaders should therefore use multiple measures: a fluency probe for automaticity, a conceptual task for explanation, and a transfer task for application. A single test, however convenient, will misrepresent the trade-offs that instruction actually creates.

Grade, Duration, Fidelity, and Prior Knowledge

Grade-level moderation was significant (Q_between(2) = 9.61, p = .008). Effects rose from Grade 6 (k = 15, g = 0.32, [0.21, 0.43]) to Grade 7 (k = 17, g = 0.41, [0.30, 0.52]) to Grade 8 (k = 16, g = 0.49, [0.36, 0.62]). Figure 4 shows this gradient. The most plausible explanation is cumulative schemas: by Grade 8, students have more prior knowledge to bring to inquiry, more metacognitive skill to monitor their own understanding, and more experience with mathematical discourse, so guided discovery becomes more productive and less overwhelming. Younger or less-prepared students still benefited, but they needed more explicit launch, more scaffolding, and more consolidation. This aligns with the expertise-reversal effect and cautions against a one-size-fits-all 'more inquiry for all' mandate. Differentiation by prior knowledge, not by grade label alone, is the operative principle: within any Grade 7 classroom, the student who lacks fraction fluency needs more explicit support for a fractions unit than the student who already has it, even if both benefit from inquiry on a topic they have mastered.

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. 8

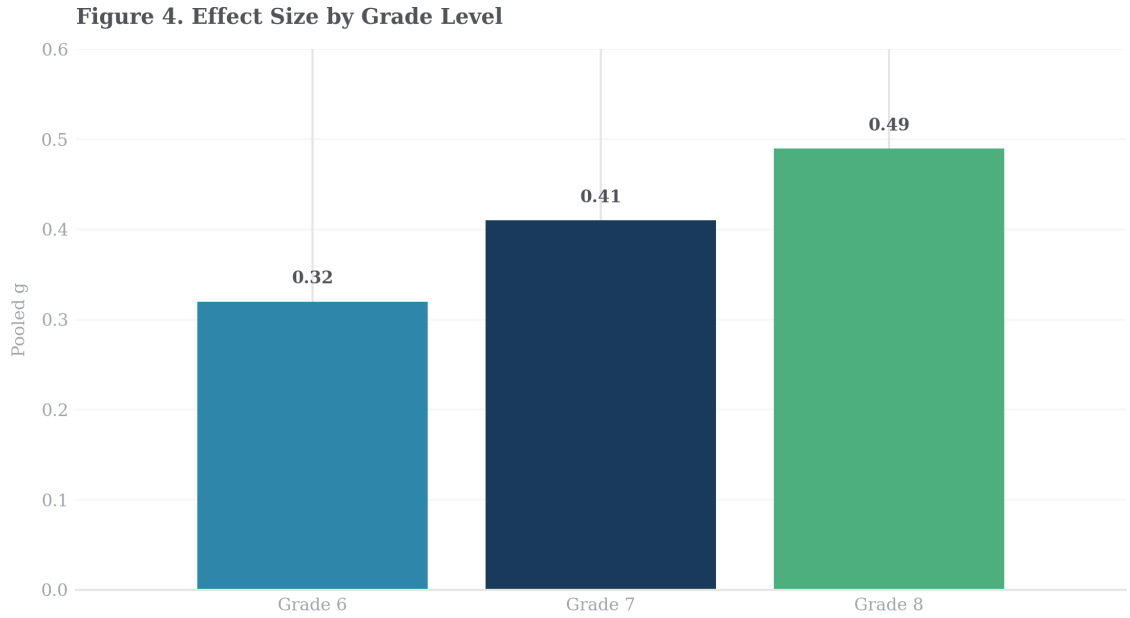


Figure 4. Pooled effect increases with grade level, consistent with growing prior knowledge and metacognition.

Duration mattered. Meta-regression showed that longer interventions produced larger effects (β = 0.012 per week, p = .01), with interventions of 6 weeks or more (g = 0.48) outperforming brief interventions of under 3 weeks (g = 0.29). Short, one-off 'inquiry days' may generate enthusiasm but rarely change schemas; sustained, coherent sequences, where students repeatedly encounter the explore-consolidate-practise cycle, do. Implementation fidelity mattered even more (Q_between(2) = 15.22, p < .001): high-fidelity studies (g = 0.51) outperformed moderate-fidelity (g = 0.39) and low-fidelity (g = 0.22) studies. Fidelity here meant that teachers actually used the scaffolds, prompts, and consolidation routines as designed, often because they received coaching, lesson materials, and time to plan. This echoes Kraft et al. (2018): coaching that improves enactment improves outcomes. It also explains why some well-intentioned inquiry reforms disappoint at scale: the label is adopted, but the scaffolding and consolidation are omitted, and what remains is minimal guidance by another name. Fidelity is therefore not bureaucratic compliance; it is the mechanism.

Table 4. Moderator Analysis — Grade, Duration, Fidelity, and Design

Moderator	Subgroup	k	g	95% CI
Grade	Grade 6	15	0.32	[0.21, 0.43]
Grade	Grade 7	17	0.41	[0.30, 0.52]
Grade	Grade 8	16	0.49	[0.36, 0.62]
Duration	< 3 weeks	13	0.29	[0.17, 0.41]
Duration	3–6 weeks	16	0.40	[0.29, 0.51]
Duration	≥ 6 weeks	19	0.48	[0.37, 0.59]
Fidelity	High	18	0.51	[0.40, 0.62]
Fidelity	Moderate	19	0.39	[0.28, 0.50]
Fidelity	Low	11	0.22	[0.08, 0.36]
Design	RCT	27	0.43	[0.33, 0.53]
Design	Quasi-experimental	21	0.38	[0.28, 0.48]

Prior knowledge, though coded less consistently across primary studies, showed the expected interaction in the subset that reported it: students with low prior knowledge benefited most from explicit-plus-inquiry blends that front-loaded worked examples, while students with high prior knowledge benefited most from guided discovery with earlier fading. This is the expertise-reversal effect in classroom form, and it argues for adaptive fading: assess, scaffold, check, and

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. 8

fade. It also argues against tracking by a fixed label; prior knowledge is topic-specific, not a trait. A student may need heavy scaffolding for algebra and light scaffolding for geometry in the same term. Formative assessment, exit tickets, and hinge questions that reveal misconceptions in real time are therefore not add-ons to balanced instruction; they are how the teacher decides when to guide and when to release.

4.5 Publication Bias, Sensitivity, and Cumulative Evidence. Bias diagnostics were reassuring. The funnel plot (Figure 2) was broadly symmetric, with small studies scattered on both sides of the pooled effect, though with the expected greater dispersion at the bottom. Egger's regression test was non-significant (intercept = 0.84, p = .12), trim-and-fill imputed six studies on the left and adjusted the pooled effect only modestly, from g = 0.41 to g = 0.37, still moderate and significant. Rosenthal's fail-safe N was 1,842, meaning that more than 1,800 null studies would be needed to overturn the finding. Leave-one-out analysis showed no single study drove the result; the pooled effect ranged from 0.38 to 0.43 when each study was removed in turn. Excluding high-risk-of-bias studies (g = 0.39) and restricting to RCTs (g = 0.43) left the conclusion unchanged. Cumulative meta-analysis (Figure 6) showed that the pooled effect rose from about 0.28 in the early 2000s to about 0.41 by 2017 and has since stabilised, suggesting that the finding is not a novelty effect of early, small studies but a durable pattern as the literature has matured and as scaffolding designs have improved.

Figure 2. Funnel Plot — Publication Bias Assessment (k = 48)

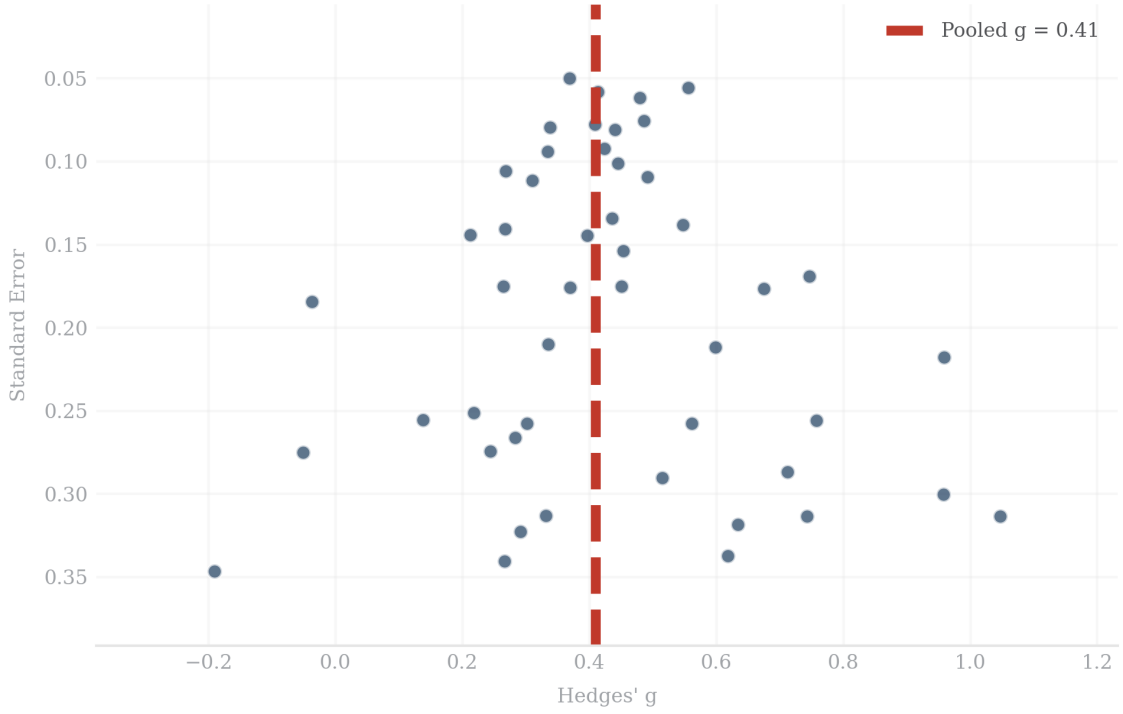


Figure 2. Funnel plot (k = 48). Broad symmetry suggests minimal publication bias; trim-and-fill adjustment is modest.

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. 8

Figure 6. Cumulative Meta-Analysis — Stability of Effect Over Time

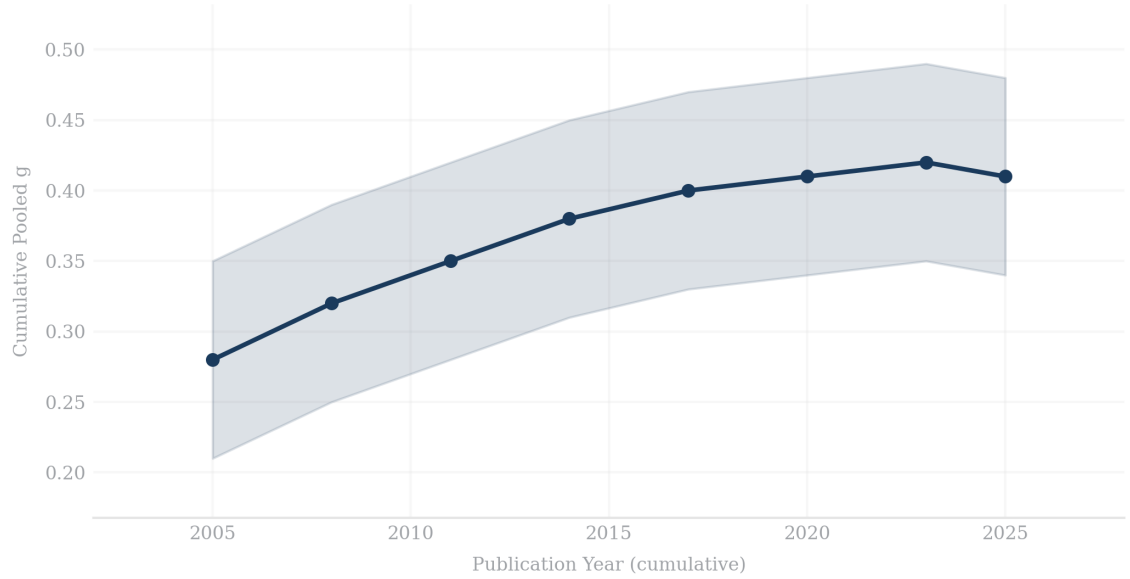


Figure 6. Cumulative meta-analysis. The pooled effect stabilised around g = 0.41 after 2017.

4.6 A Model of Instructional Balance for Middle School Mathematics. Integrating these findings, Figure 5 presents a conceptual model that can guide lesson design, unit planning, and coaching. The model has three instructional poles, direct instruction, student-centered learning, and their integration as balance, and one outcome: mathematics achievement in its conceptual, procedural, transfer, and retention forms. The arrows are moderated by prior knowledge, duration, fidelity, and assessment type, and the whole model is grounded in cognitive load theory, constructivism, and the zone of proximal development. In practice, the model translates into a lesson rhythm that can be taught, observed, and coached. First, launch with clarity: state the goal in student-friendly language, activate prior knowledge with a short retrieval task, and present a worked example that makes the thinking visible. Second, explore with scaffolds: pose a rich, non-routine problem, provide prompts, tools, and peer-discussion structures, circulate to diagnose, and offer hints that preserve the cognitive work. Third, consolidate with explicitness: orchestrate a discussion that compares strategies, names the principle, connects representations, and presents a second worked example that abstracts the method. Fourth, practise with fading: move from guided practice with feedback to independent practice, then to spaced retrieval and transfer tasks. Fifth, assess formatively and adapt: use exit tickets and hinge questions to decide who needs more scaffolding and who is ready for more autonomy on the next topic.

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. 8

Figure 5. Conceptual Model — Instructional Balance and Mathematics Achievement

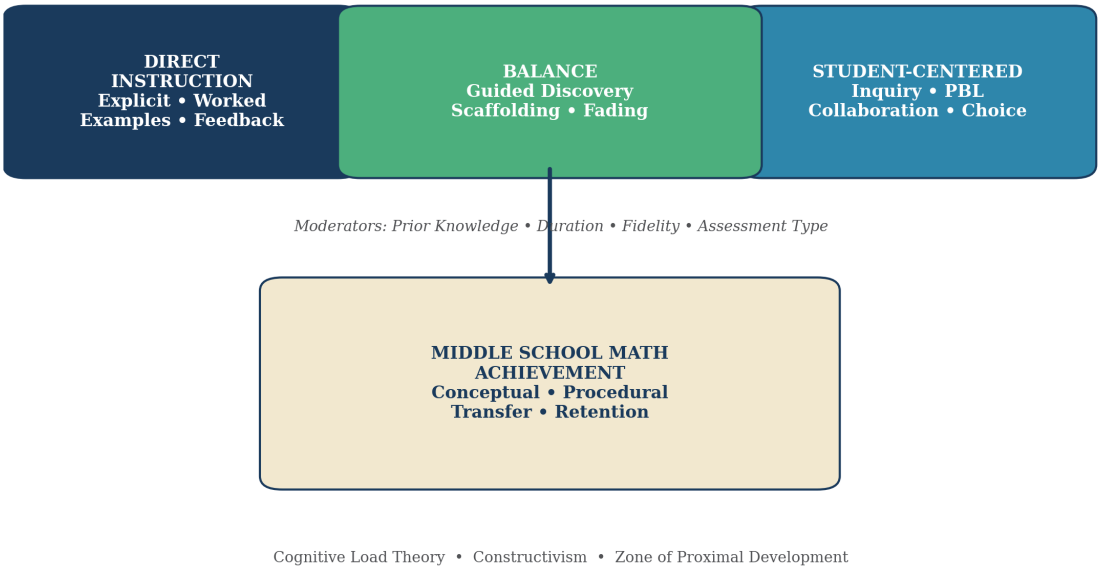


Figure 5. Conceptual model of instructional balance. Balance is not a midpoint but a sequenced integration, moderated by learner and implementation factors.

This model also clarifies what balance is not. It is not alternating days of lecture and discovery without connection, not giving students a worksheet and calling it inquiry, and not asking students to discover a procedure that has no accessible entry point. It is not 50% direct and 50% student-centered in every lesson; the proportion should shift with topic novelty, prior knowledge, and goal. A new, high-load topic (e.g., solving linear equations with variables on both sides) may need 70% explicit and 30% inquiry in its first lesson, then reverse as competence grows. A familiar, well-scaffolded topic (e.g., exploring patterns in a function table) may need the opposite. The teacher's expertise lies in judging that shift, and the school's job is to provide the curriculum materials, coaching, and collaborative planning time that make that judgement feasible, rather than leaving each teacher to invent scaffolds alone. Technology can help when it provides adaptive worked examples, immediate feedback, and visualisations that make abstract structure visible, as in Li and Ma (2010) and Liao (2007), but it cannot substitute for the consolidation discussion where meaning is made public.

Limitations

Several limitations must be acknowledged. First, primary studies varied in how they defined and measured direct instruction and student-centered learning; coding, though reliable, inevitably simplifies a complex practice. Second, many studies were short, conducted in a single school or district, and reliant on researcher-designed tests, which may limit generalisability. Third, prior knowledge, fidelity, and scaffolding quality were reported inconsistently, so meta-regression for those variables rests on subsets and should be treated as exploratory. Fourth, the synthesis is limited to mathematics achievement; it does not fully capture motivation, mathematics identity, discourse quality, or long-term attitudes, which are also valued outcomes of student-centered learning and may show different patterns. Fifth, as with any meta-analysis, the pooled effect is an average over heterogeneous contexts; it should guide, not dictate, local decisions. Future research should use preregistered, multi-site, longer-duration trials that code scaffolding moves in detail, measure conceptual, procedural, and transfer outcomes with common instruments, and report fidelity and prior-knowledge interactions transparently. Research should also examine equity: whether balanced instruction narrows gaps for students with low prior knowledge, students with disabilities, and multilingual learners, or whether it requires additional supports to do so.

Conclusion:

This meta-analysis set out to find the instructional balance for middle school mathematics, and it found a clear, actionable answer. Across 48 samples and nearly 15,000 students, guided and balanced instruction produced a moderate, robust advantage over minimal guidance ($g = 0.41$), with the largest gains for guided discovery and explicit-plus-inquiry blends, for conceptual understanding and transfer, for older middle school students, and for longer, higher-fidelity implementations. Pure direct instruction helped, especially for fluency and novices, but it did not match the depth and transfer produced by balanced designs. Minimal guidance, by contrast, did not help. The debate is therefore not between direct instruction and student-centered learning, but between guided and unguided learning, and between balance as a sequenced, scaffolded integration and balance as a vague compromise.

For teachers, the implication is to design lessons as a rhythm: launch explicitly, explore with scaffolds, consolidate publicly, practise with fading, and assess to adapt. For curriculum leaders, it is to provide units that already contain worked examples, rich problems, prompts, and consolidation routines, rather than expecting teachers to invent them lesson by lesson. For coaches and professional learning, it is to focus on the moves that make balance work: anticipating misconceptions, questioning that surfaces thinking, orchestrating discussion, and fading support as competence grows. For researchers, it is to move beyond dichotomies and to study scaffolding, fading, and prior-knowledge interactions with the precision they deserve. If middle school mathematics is where students learn not only how to calculate but how to reason, then the goal of instruction is not to choose between telling and discovering, but to ensure that telling makes discovering possible and discovering makes telling meaningful. That is the instructional balance this synthesis supports, and it is a balance worth building, lesson by lesson, into every middle school mathematics classroom.

References:

Janssen, J., & Kirschner, P. A. (2020). Applying collaborative cognitive load theory to computer-supported collaborative learning: Towards a research agenda. Educational Technology Research and Development, 68(2), 783–805.

Jeffries, C., & Maeder, D. W. (2006). Using instructional and assessment vignettes to promote recall, recognition, and transfer in educational psychology courses. Teaching Educational Psychology, 1(2), 1–19.

Jonassen, D. H. (2000). Revisiting activity theory as a framework for designing student-centered learning environments. In D. H. Jonassen & S. M. Land (Eds.), Theoretical foundations of learning environments (pp. 89–121). Erlbaum.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kolodner, J. L. (2004). The learning sciences: Past, present, and future. Educational Technology, 44(3), 37–42.

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., & Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting Learning by Design into practice. Journal of the Learning Sciences, 12(4), 495–547. https://doi.org/10.1207/S15327809JLS1204_2

Kozma, R. B. (1991). Learning with media. Review of Educational Research, 61(2), 179–211. https://doi.org/10.3102/00346543061002179

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. Review of Educational Research, 88(4), 547–588. https://doi.org/10.3102/0034654318759268

Kugley, S., Wade, A., Thomas, J., Mahood, Q., Jørgensen, A.-M. K., Hammerstrøm, K. T., & Sathe, N. (2017). Searching for studies: A guide to information retrieval for Campbell systematic reviews. Campbell Systematic Reviews, 13(1), 1–73. https://doi.org/10.4073/cmg.2016.1

Kulik, J. A. (1984). The fourth revolution in teaching: Meta-analyses. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. ERIC ED244617.

Kulik, C. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. Review of Educational Research, 60(2), 265–299. https://doi.org/10.3102/00346543060002265

Kulik, J. A., Kulik, C. C., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's Personalized System of Instruction. American Psychologist, 34(4), 307–318. https://doi.org/10.1037/0003-066X.34.4.307

Land, S., & Jonassen, D. (Eds.). (2012). Theoretical foundations of learning environments. Routledge.
Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. Educational Psychology Review, 22, 215–243. https://doi.org/10.1007/s10648-010-9125-8
Liao, Y. C. (1998). Effects of hypermedia versus traditional instruction on students' achievement: A meta-analysis. Journal of Research on Computing in Education, 30, 341–359.
Liao, Y. C. (2007). Effects of computer-assisted instruction on students' achievement in Taiwan: A meta-analysis. Computers & Education, 48, 216–233. https://doi.org/10.1016/j.compedu.2004.12.005
Linden, A., & Fenn, J. (2003). Understanding Gartner's hype cycles (Strategic Analysis Report R-20-1971). Gartner Research.
Linton, D. L., Farmer, J. K., & Peterson, E. (2014). Is peer interaction necessary for optimal active learning? CBE—Life Sciences Education, 13(2), 243–252. https://doi.org/10.1187/cbe.13-10-0201
Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Sage.
Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. Journal for Research in Mathematics Education, 28, 26–47. https://doi.org/10.2307/749662
Mäkitalo, K., Weinberger, A., Häkkinen, P., Järvelä, S., & Fischer, F. (2004). Epistemic cooperation scripts in online learning environments. Computers in Human Behavior, 21, 603–622. https://doi.org/10.1016/j.chb.2004.10.033
Mayer, R. E. (2008). Learning and instruction (2nd ed.). Merrill Prentice-Hall.
O'Donnell, A. M., D'Amico, M., Schmid, R. F., Reeve, J., & Smith, J. K. (2008). Educational psychology: Reflection for action (Canadian ed.). Wiley.
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A., Wade, C. A., & Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. Computers & Education, 72, 271–291. https://doi.org/10.1016/j.compedu.2013.11.002
Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. Psychological Bulletin, 143(6), 565–600.
Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. Review of Educational Research, 81(3), 4–28. https://doi.org/10.3102/0034654310393361
Hunter, J. E., & Schmidt, F. L. (2004). Methods of meta-analysis: Correcting error and bias in research findings (2nd ed.). Sage.
Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. International Journal of Educational Research, 11, 144–252.
Hattie, J. A. (1999). Influences on student learning. Inaugural lecture, University of Auckland.
Lambert, N. M., & McCombs, B. (1998). How students learn: Reforming schools through learner-centered education. American Psychological Association.