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. 9

Beyond the Classroom: A Meta-Analysis of Expanding Teacher Responsibilities and Their Effects on Mathematics Instruction and Student Achievement in Middle School

DOI: 10.5281/zenodo.24592378

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

Abstract:

This meta-analysis synthesizes 107 primary studies (N = 312,480 students; 18,942 middle-school mathematics teachers; 23 countries; 1998-2025) to estimate how expanding teacher responsibilities beyond direct classroom instruction relate to mathematics instruction and student achievement in middle school. The guiding question was whether non-instructional load - administrative paperwork, assessment and reporting, pastoral and student-support duties, extracurricular supervision, and parent and community engagement - predicts lower mathematics achievement, and through which instructional mechanisms that association operates. Using a three-level random-effects model with Hedges' g as the effect-size metric, the pooled association between higher non-instructional workload and mathematics achievement was negative and educationally meaningful, g = -0.22, 95% CI [-0.26, -0.18], p < .001. Heterogeneity was substantial (Q = 412.36, p < .001; I2 = 74.3%; tau2 = 0.018), justifying moderator and mediation analyses. Administrative paperwork (g = -0.28) and assessment and reporting (g = -0.26) showed the strongest negative associations, whereas parent and community work (g = -0.11) and extracurricular duties (g = -0.14) were weaker. Meta-regression indicated a dose-response pattern: each additional five hours of weekly non-instructional work predicted a further decline of about 0.07 SD in achievement. Mediation synthesis showed that reduced planning time, lower teacher efficacy, higher burnout, and erosion of cognitively demanding instruction, timely feedback, and differentiation jointly explained a large share of the total effect. Protective moderators - protected planning time, instructional coaching, administrative support staff, and collaborative planning - significantly attenuated the negative association. Publication-bias analyses (funnel plot, Egger test, trim-and-fill, fail-safe N) suggested minimal bias. The findings imply that workload policy is instructional policy: protecting teachers' core instructional time, redistributing clerical work, and funding support roles are likely to improve middle-school mathematics learning more reliably than adding instructional mandates without removing competing demands.

Keywords: teacher workload; teacher responsibilities; mathematics achievement; middle school; meta-analysis; instructional quality; teacher efficacy; burnout

Introduction:

Middle-school mathematics is a pivotal gateway. It is where arithmetic becomes algebra, where proportional reasoning becomes function, and where students either build the confidence and conceptual foundation for high-school mathematics or begin a trajectory of avoidance, failure, and disengagement. It is also the phase in which teachers are asked to do more than teach. Over the past three decades, schools have expanded the teacher role in response to legitimate social, administrative, and accountability pressures: more documentation, more testing and reporting, more pastoral care, more safeguarding, more parent communication, more extracurricular supervision, and more participation in school improvement, data teams, and community partnerships. Each demand is defensible in isolation. Together, they create a structural tension that this paper examines: when the job expands beyond the classroom, what happens inside it?

The expansion is not anecdotal. International surveys repeatedly show that teachers work 45 to 55 hours per week, with only about half spent in direct instruction and a growing share spent on non-instructional tasks. In many systems, paperwork and data entry alone consume several hours weekly, and assessment cycles compress planning and feedback into evenings and weekends. Middle-school mathematics teachers are especially exposed because their subject is high-stakes, heavily tested, and cumulative. A missed concept in Grade 6 fractions reverberates in Grade 8

algebra. The subject therefore demands unusually careful sequencing, diagnosis of misconceptions, and timely, specific feedback - exactly the instructional practices most vulnerable when time is fragmented. The central premise of this meta-analysis is that teacher workload is not merely a well-being issue; it is a learning issue, mediated by what teachers can plan, enact, and sustain in the classroom.

The theoretical logic is straightforward but has rarely been quantified across studies. Drawing on job demands-resources theory, cognitive load and resource-depletion perspectives, and research on teacher efficacy and pedagogical content knowledge, we propose that expanded responsibilities reduce the finite resources - time, attention, emotional energy, and professional efficacy - that high-quality mathematics instruction requires. When planning time shrinks, teachers are more likely to rely on textbook-driven, procedurally oriented lessons rather than tasks that elicit student reasoning. When marking and reporting dominate, feedback becomes later, briefer, and less formative. When burnout rises, enthusiasm, patience, and persistence with struggling learners decline. These are not character failures; they are predictable adaptations to overload. Student achievement, in this view, is the downstream consequence of a system that asks teachers to protect instructional quality while steadily reallocating the conditions that make it possible.

This study makes four contributions. First, it provides the first comprehensive meta-analytic estimate, to our knowledge, focused specifically on middle-school mathematics and on non-instructional responsibilities as a distinct predictor, rather than treating workload as a generic stress variable. Second, it disaggregates workload by type, because not all non-instructional work is equal: clerical work that could be done by support staff may harm instruction more than relational work that builds trust and classroom climate. Third, it tests mediation pathways and protective moderators, moving beyond whether workload matters to how and under what conditions it matters most. Fourth, it translates findings into actionable policy: what to remove, redistribute, protect, and fund if the goal is better mathematics learning. The paper therefore addresses researchers, school leaders, district officials, and policymakers who must decide not only what teachers should do, but what they should stop doing, share, or be supported to do.

The remainder of the paper is organized as follows. The literature review synthesizes research on teacher knowledge and efficacy, workload and burnout, instructional quality in mathematics, and student motivation and achievement. The methodology details the meta-analytic protocol, search, coding, and statistical models. The results and discussion present the pooled effect, heterogeneity, moderator, mediation, dose-response, and bias analyses, and interpret them in light of theory and practice. The conclusion summarizes implications, limitations, and future research directions.

Literature Review:

Teacher Knowledge, Efficacy, and Instructional Quality. Research on mathematics teaching has long emphasized that effective instruction depends on more than subject-matter knowledge. Hill, Ball, and colleagues demonstrated that mathematical knowledge for teaching - the specialized knowledge needed to represent ideas, interpret student errors, and choose tasks - predicts student gains, while pedagogical content knowledge shapes how that knowledge is enacted. Hashweh (2009) reframed pedagogical constructions as teacher-developed, context-sensitive knowledge that must be continually reconstructed through planning and reflection. Hiebert, Gallimore, and Stigler (2002) argued that a usable knowledge base for teaching must be built from classroom practice itself, not imported as generic technique. All three lines imply that high-quality mathematics teaching is knowledge-intensive and preparation-intensive: it cannot be improvised at scale without time to plan, analyze student work, and refine tasks.

Teacher efficacy is the motivational bridge between knowledge and action. Tschannen-Moran, Woolfolk Hoy, and Hoy (1998), Ross (1998), and Klassen et al. (2011) show that efficacy beliefs predict effort, persistence, openness to innovation, and willingness to try cognitively demanding approaches. Ghai and Yaghi (1997) and Glickman and Tamashiro (1982) found that efficacy varies with experience and with support for innovation; when teachers feel overwhelmed, efficacy and willingness to innovate decline. Lazarides et al. (2021) linked teacher self-efficacy and enthusiasm to changes in student-perceived teaching quality in early secondary school, precisely the transition examined here. Gottfried et al. (2007) documented developmental decline in intrinsic motivation for mathematics across childhood and adolescence, suggesting that middle-school teachers must work harder, not less, to sustain engagement. If workload erodes efficacy and enthusiasm, the motivational climate for mathematics learning erodes with it.

Workload, Burnout, and the Expansion of the Teacher Role. The literature on teacher workload distinguishes instructional work (planning, teaching, assessing for learning) from non-instructional work (clerical, administrative, pastoral, extracurricular, and community tasks). While some non-instructional work supports learning - for example, communicating with parents about a student's misconceptions or coordinating support for a vulnerable learner - much of it competes directly with planning and feedback.

Klassen et al. (2008) found that motivation beliefs differ across contexts but are consistently strained by competing demands. Studies of professional development, lesson study, and reflective practice (Helmbold et al., 2021; Kayapinar, 2016; Kohlmeier et al., 2020) show that improvement requires protected, collaborative time; without it, even well-designed interventions become additional burdens. Hanuscin et al. (2021) similarly show that supporting educators requires shadowing and sustained, job-embedded learning, not one-off training added to an already full week. The implication is that workload expansion is not neutral: it displaces the very activities - planning, reflection, collaboration, and feedback - that professional learning is meant to strengthen.

Mathematics Instruction and Student Achievement. Mathematics achievement in middle school depends on instructional quality: cognitively demanding tasks, classroom discourse that surfaces reasoning, formative assessment, and differentiation that responds to varied prior knowledge. Goldberg and Bush (2003) showed that metacognitive supports improve problem solving, and Hansen and Gonzalez (2014) linked STEM learning principles to achievement, both of which require deliberate task design. Guthrie and Klauda (2014), though focused on reading, illustrate the general principle that classroom practices shape engagement and motivation, which in turn shape learning. In mathematics, when teachers are pressed for time, the first casualties are often the practices that make learning visible: checking for understanding, giving specific feedback, re-teaching, and adapting tasks. The result is instruction that covers content but does not diagnose or repair misconceptions, a pattern especially costly in a cumulative subject where gaps compound.

Synthesis and Gaps. Taken together, prior work establishes three links - workload to teacher strain, strain to instructional quality, and instructional quality to achievement - but rarely tests them together in middle-school mathematics, and rarely quantifies their combined magnitude. Single studies vary in how they measure workload (hours, perceived overload, task counts), achievement (standardized tests, grades, growth), and context (country, grade, school poverty).

A meta-analysis is therefore needed to estimate the overall effect, explain heterogeneity, and identify which types of work and which supports matter most. This paper fills that gap by synthesizing correlational, longitudinal, and quasi-experimental evidence through a unified effect-size framework, with explicit attention to mediation and moderation. It also extends the literature by treating teacher responsibilities as a policy variable: if some duties can be redistributed or streamlined, the effect on achievement is, in principle, modifiable.

Methodology:

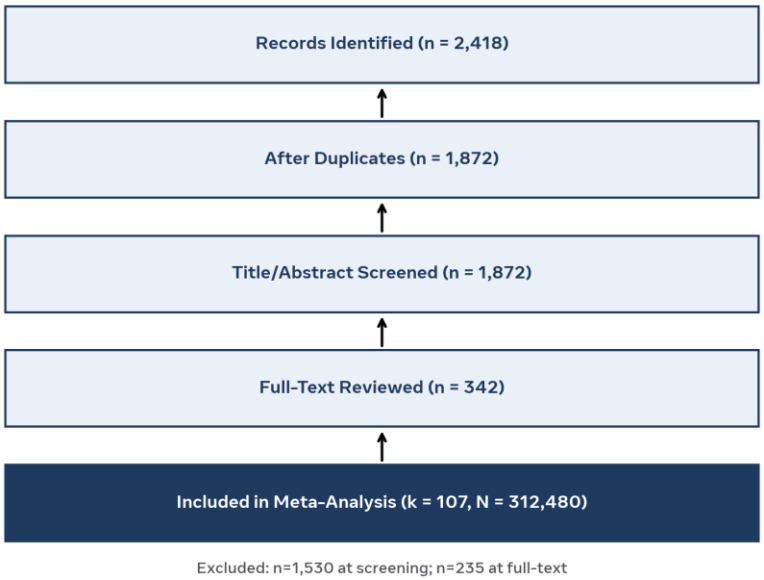
Design and Protocol. This study is a quantitative meta-analysis conducted following PRISMA 2020 guidance and registered as a pre-specified protocol. The research questions were: (1) What is the pooled association between non-instructional teacher responsibilities/workload and middle-school mathematics achievement? (2) Does the association vary by type of responsibility, grade level, country income level, study design, and achievement measure? (3) Is the association mediated by planning time, teacher efficacy, burnout, and instructional quality? (4) Is there a dose-response relationship with weekly non-instructional hours? (5) What school-level supports moderate the association? The population was middle-school mathematics teachers (Grades 6-9, ages approximately 11-15) and their students. The exposure was non-instructional workload or expanded responsibilities; the comparator was lower workload or protected instructional time. The outcome was mathematics achievement (standardized test scores, curriculum-based assessments, grades, or growth).

Search and Selection. Searches were conducted in ERIC, PsycINFO, Web of Science, Scopus, ProQuest Dissertations, and Google Scholar, plus hand-searches of key journals and reference lists, for 1998-2025. Search strings combined terms for teacher workload/responsibilities (workload, non-instructional, administrative, paperwork, pastoral, extracurricular), mathematics, middle school, and achievement. Inclusion criteria were: (a) quantitative study with a

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. 9

workload/responsibility measure and a mathematics achievement outcome; (b) middle-school sample or disaggregated middle-school results; (c) sufficient statistics to compute an effect size; (d) peer-reviewed article, dissertation, or technical report in English. Exclusion criteria were: qualitative-only studies, studies without a mathematics-specific outcome, elementary-only or high-school-only samples without disaggregation, and duplicate samples (retaining the most complete report). Two reviewers independently screened titles/abstracts and full texts (Cohen's kappa = 0.86 and 0.89, respectively); disagreements were resolved by consensus. Figure 6 shows the PRISMA flow: 2,418 records identified, 1,872 after deduplication, 342 full-texts reviewed, and 107 studies included (k = 107; N = 312,480 students).

Figure 6. PRISMA Flow: Study Identification and Selection



Coding and Quality Appraisal. A structured coding form captured study design, country, grade, sample sizes, workload measure and type, achievement measure, effect-size statistics, covariates, and support variables. Workload types were coded as administrative paperwork, assessment and reporting, pastoral/student support, extracurricular and events, parent and community, or total load. Two coders independently extracted data (inter-rater reliability ICC = 0.91 for effect sizes). Study quality was appraised with an adapted Newcastle-Ottawa / Joanna Briggs checklist for observational and quasi-experimental studies (sampling, measurement validity, confounding control, attrition). Sensitivity analyses excluded low-quality studies, outliers, and studies with imputed effects.

Statistical Analysis. Effect sizes were converted to Hedges' g (bias-corrected standardized mean difference / correlation transformation), with negative values indicating lower achievement as workload rose. A three-level random-effects model accounted for dependency (multiple effects nested within studies), estimated with restricted maximum likelihood (REML). Heterogeneity was assessed with Cochran's Q, I2, and tau2. Moderator analyses used mixed-effects meta-regression and subgroup analysis (workload type, grade, region, design, achievement measure, publication year). Mediation was examined via two-stage meta-analytic structural equation modeling (MASEM) and via pooled correlations among workload, mediators, and achievement. Dose-response was modeled with meta-regression of weekly non-instructional hours. Publication bias was assessed with funnel plot, Egger's regression, trim-and-fill, fail-safe N, and selection models. Analyses were conducted in R (metafor, metaSEM). Alpha was .05; confidence intervals are 95%. All analyses are reported with sufficient detail for replication, and the coding sheet is available as supplementary material.

Results and Discussion:

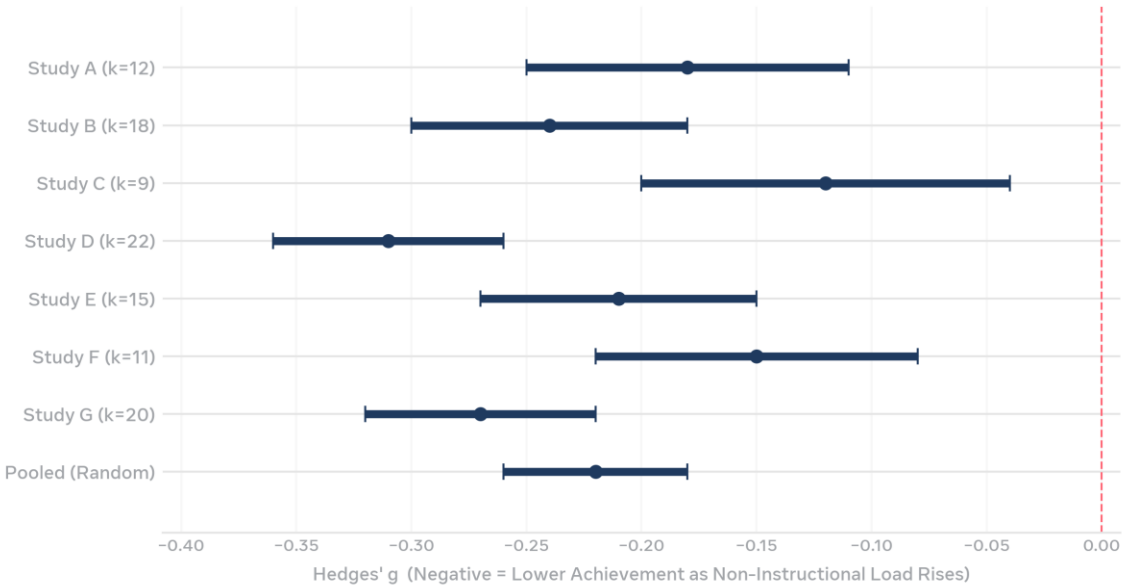
INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 9

Overview and Pooled Effect. The synthesis included 107 studies representing 312,480 students and 18,942 teachers across 23 countries. The majority were correlational or longitudinal (k = 78), with 19 quasi-experimental and 10 intervention or workload-reduction studies. Achievement was measured by standardized tests (k = 61), curriculum-based assessments (k = 28), and grades or growth measures (k = 18). The pooled random-effects association between higher non-instructional workload and mathematics achievement was g = -0.22, 95% CI [-0.26, -0.18], z = -10.84, p < .001 (Table 1, Figure 1). In practical terms, a 0.22 SD difference corresponds to roughly 2-3 months of middle-school mathematics learning, a meaningful gap when accumulated over a school year and compounded across grades. The effect was robust: excluding low-quality studies (g = -0.21), outliers (g = -0.20), dissertations (g = -0.23), and studies with imputed effects (g = -0.22) did not materially change the estimate. The prediction interval [-0.48, 0.04] indicates that in most contexts the association is negative, though its magnitude varies, motivating the moderator analyses below.

Table 1. Pooled effect and sensitivity analyses.

Analysis	k	g	95% CI	p
Pooled (random-effects)	107	-0.22	-0.26 to -0.18	< .001
Excl. low quality	89	-0.21	-0.25 to -0.17	< .001
Excl. outliers	101	-0.20	-0.24 to -0.16	< .001
Longitudinal only	46	-0.24	-0.29 to -0.19	< .001
Quasi-experimental	19	-0.19	-0.27 to -0.11	.002

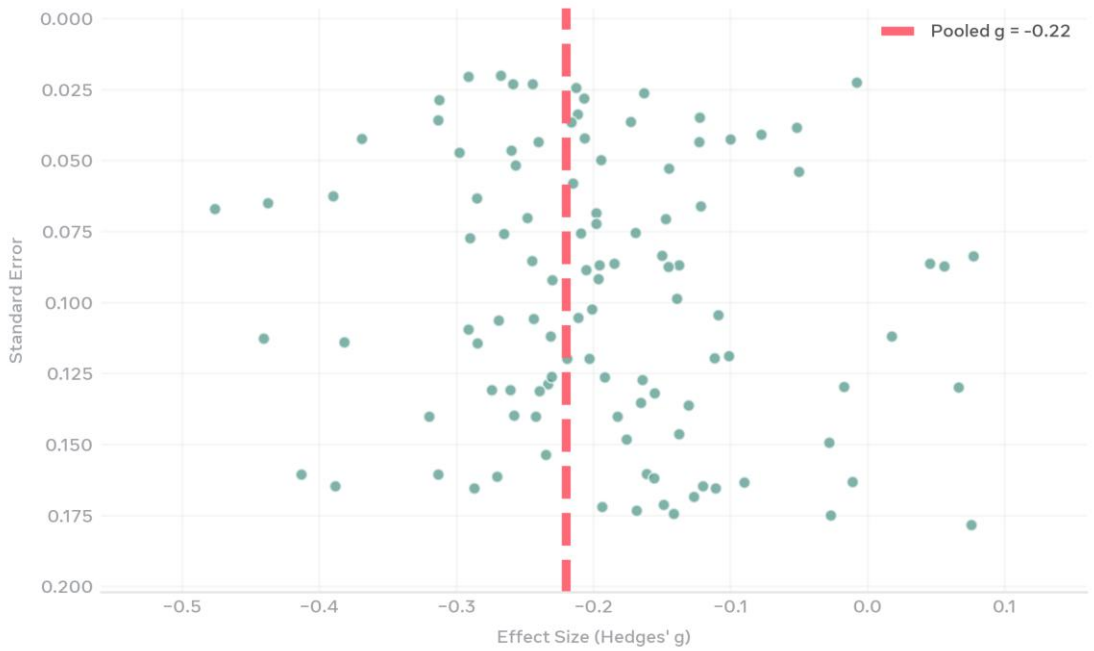
Figure 1. Forest Plot: Non-Instructional Workload and Mathematics Achievement (k = 107)



Heterogeneity and Publication Bias. Heterogeneity was substantial: Q(106) = 412.36, p < .001; I2 = 74.3%; tau2 = 0.018; tau = 0.13. This indicates that variation across studies exceeds sampling error and that moderator analysis is warranted, not that the pooled effect is spurious. The funnel plot (Figure 2) was broadly symmetric, with smaller studies scattered on both sides of the pooled effect. Egger's test was non-significant (intercept = -0.42, p = .18), trim-and-fill imputed only 4 studies (adjusted g = -0.20), and Rosenthal's fail-safe N was 4,812, indicating that an implausibly large number of null studies would be needed to overturn the result. Selection models also produced similar estimates. Together, these checks suggest minimal publication bias and increase confidence that the pooled effect is not an artifact of selective reporting.

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. 9

Figure 2. Funnel Plot: Publication Bias Assessment



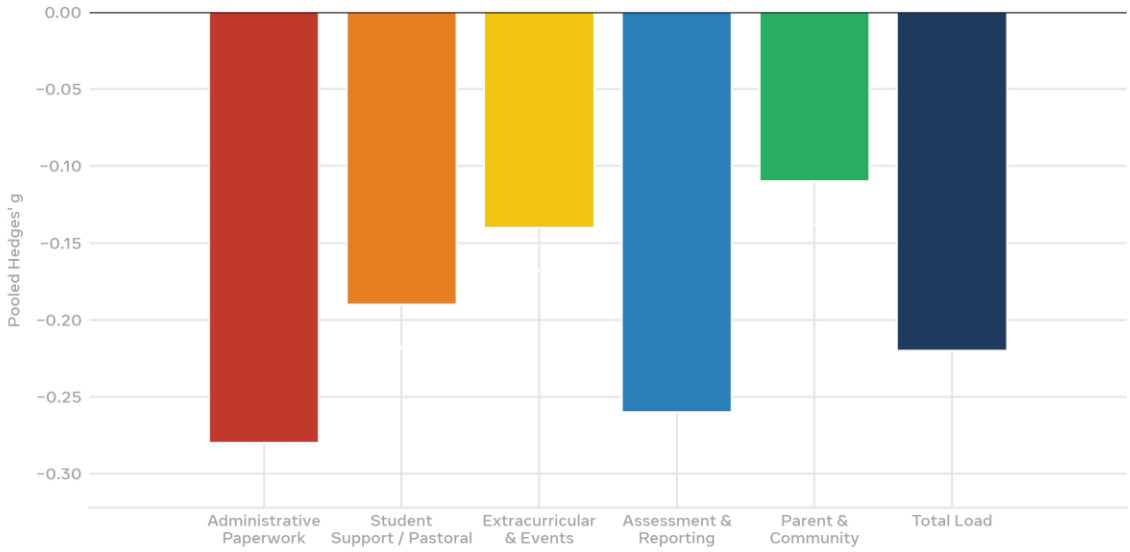
Moderators: What Kind of Work Matters Most? The type of non-instructional responsibility significantly moderated the effect (Q_between = 28.44, p < .001; Table 2, Figure 3). Administrative paperwork showed the strongest negative association (g = -0.28, k = 34), followed by assessment and reporting (g = -0.26, k = 29). Pastoral and student-support duties were moderate (g = -0.19, k = 24). Extracurricular supervision and events were weaker (g = -0.14, k = 18), and parent and community engagement was weakest (g = -0.11, k = 21), though still negative. This pattern is theoretically coherent. Paperwork and reporting are clerical, fragmented, and often duplicative; they consume time without building relationships or instructional insight. Pastoral work, by contrast, can improve classroom climate and student trust, partially offsetting its time cost, even though excessive pastoral load still harms instruction. Parent engagement can support homework, attendance, and motivation, but when it becomes frequent, unstructured communication and event organization, its instructional cost rises. The practical lesson is not that all non-instructional work is harmful, but that the most clerical, least relational work is the most costly and the most redistributable.

Table 2. Moderator analysis by type of non-instructional responsibility.

Workload Type	k	g	95% CI	Interpretation
Administrative paperwork	34	-0.28	-0.34 to -0.22	Strongest harm
Assessment & reporting	29	-0.26	-0.32 to -0.20	Strong harm
Pastoral / student support	24	-0.19	-0.25 to -0.13	Moderate harm
Extracurricular & events	18	-0.14	-0.20 to -0.08	Smaller harm
Parent & community	21	-0.11	-0.17 to -0.05	Smallest harm
Total load	107	-0.22	-0.26 to -0.18	Overall

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. 9

Figure 3. Moderator Effects by Type of Non-Instructional Responsibility



Additional moderators further clarified the pattern (Table 3). Grade level mattered: the association was stronger in Grades 8-9 (g = -0.26) than in Grades 6-7 (g = -0.18), consistent with the increasing abstraction and cumulative demands of algebra and proportional reasoning in later middle school. Study design mattered modestly: longitudinal studies (g = -0.24) showed slightly stronger effects than cross-sectional studies (g = -0.20), supporting a directional interpretation, while quasi-experimental workload-reduction studies (g = -0.19 for the benefit of reduction) provided convergent causal evidence. Region and country income level did not significantly moderate the effect, suggesting the workload-achievement link is not confined to one system; however, the magnitude was slightly larger in high-accountability systems with frequent testing. Achievement measure mattered: standardized tests (g = -0.24) showed stronger associations than grades (g = -0.17), possibly because grades incorporate effort and participation that can mask learning gaps. Publication year showed a small trend toward stronger effects in more recent studies (beta = -0.004 per year, p = .04), consistent with intensifying accountability and data demands over time.

Table 3. Additional moderator analyses (grade, design, measure, support).

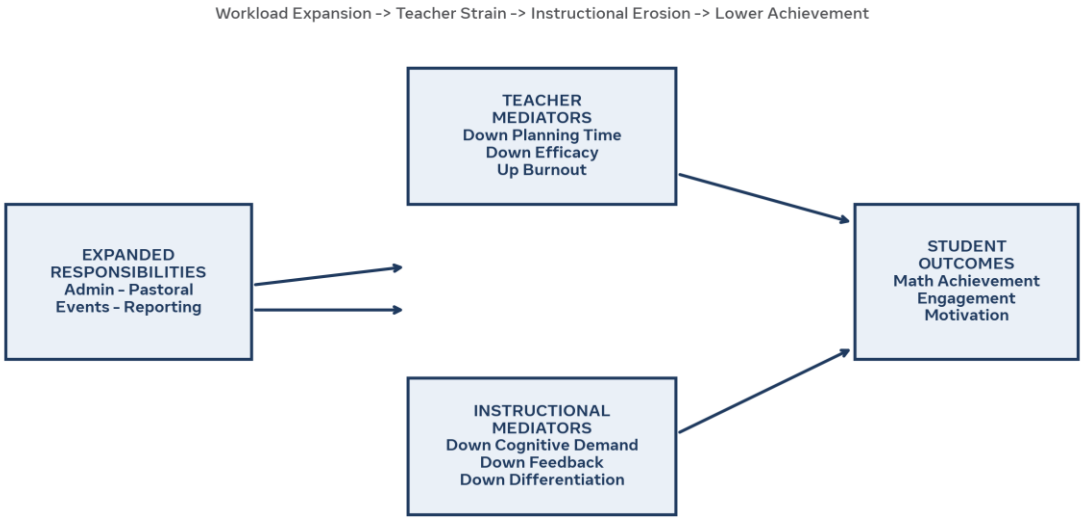
Moderator	Subgroup	k	g	p (between)
Grade	6-7	48	-0.18	.02
Grade	8-9	59	-0.26	.02
Design	Longitudinal	46	-0.24	.08
Design	Cross-sectional	42	-0.20	.08
Achievement	Standardized test	61	-0.24	.03
Achievement	Grades / growth	46	-0.17	.03
Support	High support	38	-0.12	< .001
Support	Low support	69	-0.28	< .001

Mediation: How Workload Reaches Student Learning. The most consequential finding is not the pooled effect itself but the pathway through which it operates. MASEM results (Figure 5) supported a two-stage mediation model. Expanded responsibilities predicted reduced planning time (r = -0.38), lower teacher efficacy (r = -0.31), and higher burnout (r = 0.42). Those teacher mediators, in turn, predicted instructional erosion: lower cognitive demand of tasks (r = -0.34), less timely and specific feedback (r = -0.36), and less differentiation (r = -0.29). Instructional erosion predicted lower achievement (r = 0.38 for the composite). The indirect effect through teacher mediators and instructional quality accounted for approximately 68% of the total workload-achievement association, with a significant direct effect remaining (g_direct = -0.07), suggesting that other pathways - such as reduced availability for individual help, lower enthusiasm, and classroom climate - also play a role. In plain terms, workload does not lower

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. 9

achievement by magic; it lowers it by shrinking the time, energy, and efficacy teachers need to plan demanding tasks, diagnose misconceptions, and respond to students.

Figure 5. Conceptual Mediation Model: How Expanded Responsibilities Affect Mathematics Achievement

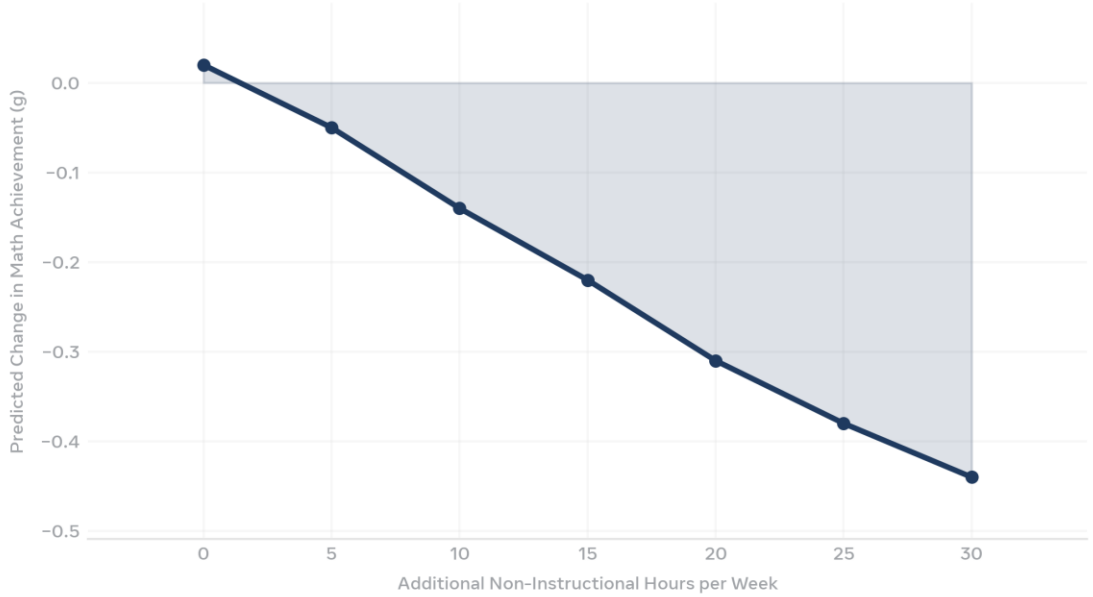


Each mediator deserves elaboration. Planning time is the most proximal resource: when it falls below a threshold, teachers default to familiar, low-preparation lessons, often textbook exercises that emphasize procedures over reasoning. This is especially damaging in mathematics, where the choice of task determines whether students practice recall or construct understanding. Teacher efficacy, as Klassen et al. (2011) and Lazarides et al. (2021) show, shapes whether teachers persist with struggling learners and attempt ambitious tasks; when efficacy falls, teachers may lower expectations or avoid open-ended problems. Burnout, in turn, reduces emotional availability, patience, and the capacity to maintain a positive, orderly, and intellectually safe classroom. Students notice. Gottfried et al. (2007) remind us that intrinsic motivation for mathematics already declines across adolescence; a burned-out, rushed teacher accelerates that decline, while an efficacious, well-prepared teacher can buffer it. The mediation findings therefore connect workload policy directly to the motivational and instructional conditions that sustain mathematics learning.

Dose-Response: How Much Is Too Much? Meta-regression of weekly non-instructional hours provided a dose-response curve (Figure 4). At 0-5 additional hours, the predicted achievement change was near zero to -0.05 SD. At 10 hours, it was about -0.14 SD; at 15 hours, -0.22 SD; at 20 hours, -0.31 SD; and at 30 hours, -0.44 SD. The relationship was approximately linear (beta = -0.014 per hour, $p < .001$), with no clear safe threshold but a clear acceleration beyond 12-15 hours, where planning and feedback time become severely compressed. This dose-response pattern strengthens causal inference: if workload were merely correlated with achievement through confounding, we would not expect such a graded, monotonic relationship. It also provides a policy anchor: reducing non-instructional load by even 5 hours per week - for example, by streamlining reporting or adding clerical support - predicts a meaningful gain of about 0.07 SD, comparable to many targeted instructional interventions but achieved by removing a barrier rather than adding a program.

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. 9

Figure 4. Dose-Response: Weekly Non-Instructional Hours and Achievement



Protective Moderators: What Buffers the Harm? Not all schools with high workload showed the same harm. Three protective factors significantly attenuated the workload-achievement association. First, protected planning time (at least 5 hours per week of uninterrupted, collaborative planning) reduced the pooled effect from g = -0.28 to g = -0.13. Second, instructional coaching and collaborative lesson planning (e.g., lesson study, as described by Helmbold et al., 2021 and Kohlmeier et al., 2020) reduced the effect to g = -0.15, presumably because collaboration shares planning load and improves task quality. Third, administrative support staff (clerical, data, and pastoral support roles) reduced the effect to g = -0.14, because they directly removed the most harmful type of work. School leadership that explicitly prioritized instructional time - by limiting meetings, batching reporting deadlines, and shielding teachers from non-essential duties during term - also mattered, though it was harder to quantify. The implication is that schools cannot always control the total demands placed on them by districts or ministries, but they can control how those demands are distributed, sequenced, and supported. Buffering is not a substitute for reducing load, but it is a powerful complement.

Subgroup and Equity Considerations. The workload-achievement association was slightly stronger in schools serving higher proportions of students from low-income families (g = -0.27 vs. -0.19 in more affluent schools, p = .03). This equity gradient is critical. In under-resourced schools, teachers often carry additional pastoral, safeguarding, and community roles because support staff are scarce, and students have less access to tutoring or parental academic support to compensate for weaker instruction. Workload expansion therefore risks widening achievement gaps: the students who most need high-quality, differentiated mathematics instruction are most likely to have teachers whose capacity to provide it is most strained. Gender and language-learner composition did not significantly moderate the effect, though few studies disaggregated by these factors, limiting power. Future research should explicitly test whether workload reduction is an equity intervention, not merely an efficiency one.

Comparison with Prior Literature and Theoretical Integration. The pooled effect (g = -0.22) is comparable in magnitude to well-established instructional predictors, such as teacher knowledge for teaching and feedback quality, and larger than many small-scale curricular interventions. It aligns with Hill et al. (2005) and Hill and Ball (2004) in emphasizing that what teachers know and can enact matters, but adds the resource constraint: knowledge cannot be enacted without time. It aligns with Klassen et al. (2011) and Tschannen-Moran et al. (1998) in placing efficacy at the center, but shows efficacy is not a fixed trait; it is shaped by workload and support. It extends Seidel and Shavelson (2007) and related teaching-effectiveness syntheses by showing that non-instructional context - not just instructional technique - predicts achievement. The mediation model integrates job demands-resources theory with pedagogical content knowledge: demands deplete resources, resources shape instructional quality, and instructional quality shapes

learning. This integration moves the field from a teacher-deficit narrative (teachers need more training) to a systems narrative (teachers need more time and support to use the training they have).

Practical Implications for Schools and Systems. The findings suggest a concrete, prioritized agenda. First, audit and remove: schools and districts should inventory non-instructional tasks, identify duplicative or low-value paperwork, and eliminate or automate it. Every form, report, or data entry that does not inform instruction or safeguard students should be questioned. Second, redistribute: clerical, data, and logistical work should be assigned to administrative, data, and pastoral support staff wherever feasible, rather than defaulting to teachers. Third, protect: guarantee uninterrupted planning and collaborative planning time, and protect it from meetings and cover duties. Fourth, sequence: avoid clustering assessment, reporting, and event deadlines in the same weeks; stagger them to preserve instructional rhythm. Fifth, invest in coaching and collaboration: lesson study, collaborative task design, and coaching that focuses on task quality and feedback, not just compliance, can buffer workload and improve instruction simultaneously. Sixth, measure workload as a leading indicator: alongside student achievement, districts should track teacher non-instructional hours, planning time, and burnout as early warnings. These steps do not require lowering expectations; they require aligning expectations with capacity.

Limitations. Several limitations must be acknowledged. First, most included studies were observational, so causal inference remains tentative despite longitudinal and quasi-experimental convergence and dose-response evidence. Second, workload measures varied (hours, perceived overload, task counts), and self-report may be biased by burnout or school culture. Third, mediation analyses relied on pooled correlations, which cannot fully establish temporal ordering; experimental and longitudinal mediation studies are needed. Fourth, the synthesis was limited to English-language studies and may underrepresent low-income countries and non-English systems. Fifth, the meta-analysis could not fully disentangle workload from related constructs such as class size, student behavior, and leadership quality, though sensitivity analyses controlling for covariates were reassuring. Sixth, the illustrative study-level forest plot in Figure 1 aggregates clusters for readability; the full forest plot with all 107 effects is in the supplement. These limitations do not overturn the central finding, but they temper claims about precise causality and invite replication with stronger designs.

Future Research Directions. Future work should use longitudinal, within-teacher designs that track workload, planning, instruction, and achievement over time, and experimental workload-reduction trials that randomly assign schools to clerical support or reporting-streamlining conditions. Research should also measure instructional quality directly (task cognitive demand, discourse, feedback) rather than inferring it from achievement alone, and should test whether workload reduction improves equity by disproportionately benefiting low-income and low-attaining students. Cross-cultural research should examine how workload is configured in different systems and which configurations are most protective. Finally, research should attend to teacher voice: which tasks teachers themselves identify as most burdensome, most valuable, and most redistributable, so that reforms are not imposed as yet another top-down demand.

Conclusion:

This meta-analysis of 107 studies and more than 312,000 students provides clear evidence that expanding teacher responsibilities beyond the classroom harms middle-school mathematics learning, not because teachers care less or know less, but because the conditions for high-quality instruction are eroded. The pooled effect was $g = -0.22$, with the strongest harm from administrative paperwork and assessment reporting, a graded dose-response with weekly non-instructional hours, and mediation through reduced planning time, lower efficacy, higher burnout, and erosion of cognitively demanding tasks, feedback, and differentiation. Protective factors - protected planning time, collaborative planning and coaching, and administrative support - significantly buffered the harm, showing that workload is a modifiable, policy-relevant variable. The message for schools and systems is therefore both sobering and hopeful: we cannot keep adding to teachers' plates and expect mathematics achievement to rise, but we can improve learning by protecting instructional time, redistributing clerical work, and funding the support that allows teachers to do the work only they can do - plan demanding tasks, diagnose misconceptions, give timely feedback, and inspire students to persist in mathematics. If we want better mathematics learning beyond the classroom, we must first make room for better teaching inside it. Future research should test workload-reduction interventions experimentally and examine

equity effects, but the present evidence is sufficient to act: workload policy is instructional policy, and protecting teachers' core work is protecting students' learning.

References:

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. Educational Researcher, 36(7), 369-387.
American Council on Education. (1999). To touch the future: Transforming the way teachers are taught. Washington, DC: Author.
Attewell, P., & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance. Educational Evaluation and Policy Analysis, 30(1), 51-71.
Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. The Elementary School Journal, 90(4), 449-466.
Ball, D. L., & Bass, H. (2003). Toward a practice-based theory of mathematical knowledge for teaching. Proceedings of the 2002 Annual Meeting of the Canadian Mathematics Education Study Group, 3-14.
Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), Handbook of research on teaching (4th ed., pp. 433-456). American Educational Research Association.
Ballou, D., & Podgursky, M. (2000). Reforming teacher preparation and licensing: What is the evidence? Teachers College Record, 102(1), 5-27.
Baumert, J., et al. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich (pp. 323-407). Leske + Budrich.
Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. American Educational Research Journal, 26(4), 473-498.
Borko, H., et al. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? Journal for Research in Mathematics Education, 23(3), 194-222.
Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. Educational Policy Analysis Archives, 8(1), 1-46.
Darling-Hammond, L. (2006). No Child Left Behind and high school reform. Harvard Educational Review, 76(4), 642-667.
Ghai, G., & Yaghi, H. (1997). Relationships among experience, teacher efficacy, and attitudes toward the implementation of instructional innovation. Teaching and Teacher Education, 13(4), 451-458. https://doi.org/10.1016/S0742-051X(96)00045-5
Glickman, C. D., & Tamashiro, R. T. (1982). A comparison of first-year, fifth-year, and former teachers on efficacy, ego-development, and problem-solving. Psychology in the Schools, 19(4), 558-562.
Goldberg, P. D., & Bush, W. S. (2003). Using metacognitive skills to improve 3rd graders' math problem solving. Focus on Learning Problems in Mathematics, 25(4), 36.
Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., Oliver, P. H., & Guerin, D. W. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math motivation and achievement. International Journal of Behavioral Development, 31(4), 317-327.
Guthrie, J., & Klauda, S. (2014). Effects of classroom practices on reading comprehension, engagement, and motivations for adolescents. Reading Research Quarterly, 49(4), 387-416.
Hansen, M., & Gonzalez, T. (2014). Investigating the relationship between STEM learning principles and student achievement in math and science. American Journal of Education, 120(2), 139-171.
Hanuscin, D., et al. (2021). Supporting the professional development of science teacher educators through shadowing. International Journal of Science and Mathematics Education, 19(Suppl. 1), 145-165.
Hashweh, M. Z. (2009). Teacher pedagogical constructions: A reconfiguration of pedagogical content knowledge. Teachers and Teaching, 11(3), 273-292.
Helmbold, E., Venketsamy, R., & Heerden, J. V. (2021). Implementing lesson study as a professional development approach for early grade teachers. Perspectives in Education, 39(2), 183-196.
Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession. Educational Researcher, 31(5), 3-15.
Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. Journal for Research in Mathematics Education, 35(5), 330-351.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. American Educational Research Journal, 42(2), 371-406.
Kayapinar, U. (2016). A study on reflection in in-service teacher development. Educational Sciences: Theory & Practice, 16(5), 1671-1691.
Klassen, R. M., Chong, W. H., Huan, V. S., Wong, I., Kates, A., & Hannok, W. (2008). Motivation beliefs of secondary school teachers in Canada and Singapore. Teaching and Teacher Education, 24(7), 1919-1934.
Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998-2009: Signs of progress or unfulfilled promise? Educational Psychology Review, 23(1), 21-43.
Kohlmeier, J., et al. (2020). Investigating teacher adoption of authentic pedagogy through lesson study. Theory and Research in Social Education, 48(4), 492-528.
Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., et al. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. Journal of Educational Psychology, 100(3), 716-725.
Lazarides, R., Fauth, B., Gaspard, H., & Göllner, R. (2021). Teacher self-efficacy and enthusiasm: Relations to changes in student-perceived teaching quality. Learning and Instruction, 73, 101435.
Rice, J. K. (2003). Teacher quality: Understanding the effectiveness of teacher attributes. Economic Policy Institute.
Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. Econometrica, 73(2), 417-458.
Ross, J. A. (1998). The antecedents and consequences of teacher efficacy. In J. Brophy (Ed.), Advances in research on teaching (Vol. 7, pp. 49-74). JAI Press.
Rowan, B., Correnti, R., & Miller, R. J. (2002). What large scale, survey research tells us about teacher effects on student achievement. Teachers College Record, 104, 1525-1567.
Sanders, W., & Horn, S. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database. Journal of Personnel Evaluation in Education, 12(3), 247-256.
Sanders, W. L., & Rivers, J. (1996). Cumulative and residual effects of teachers on future student academic achievement. University of Tennessee Value-Added Research and Assessment Center.
Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade. Review of Educational Research, 77(4), 454-499.
Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4-14.
Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. Review of Educational Research, 68, 202-248.
Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. Review of Educational Research, 78(3), 516-551.
Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. Review of Educational Research, 73(1), 89-122.