



A Meta-Analysis of the Experimental Evidence Linking Mathematics and Science Professional Development Interventions to Teacher Knowledge, Classroom Instruction, and Student Achievement

Kathleen Lynch
University of Connecticut

Kathryn Gonzalez
Mathematica Policy
Research

Heather Hill
Harvard University

Ramsey Merritt
Harvard University

Despite evidence that teacher professional development interventions in mathematics and science can increase student achievement, our understanding of the mechanisms by which this occurs – particularly how these interventions affect teachers themselves, and the extent to which teacherlevel changes predict student learning – remains limited. The current meta-analysis synthesizes 46 experimental studies of PK-12 mathematics and science professional development interventions to investigate how these interventions affect teacher-level outcomes, including knowledge and classroom instruction, and how these impacts relate to intervention effects on student achievement. Compared with controls, treatment group teachers in PK-12 mathematics and science PD intervention studies demonstrated stronger performance on teacher-level outcomes (pooled average impact estimate: +0.52 SD). Programs with larger impacts on teacherlevel outcomes also tended to have significantly larger mean impacts on student achievement. We discuss implications for future research and practice.

VERSION: May 2025

Suggested citation: Lynch, Kathleen, Kathryn E. Gonzalez, Heather C. Hill, and Ramsey Merritt. (2025). A Meta-Analysis of the Experimental Evidence Linking Mathematics and Science Professional Development Interventions to Teacher Knowledge, Classroom Instruction, and Student Achievement. (EdWorkingPaper: 24 -1023). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/r79z-tf23>

A Meta-Analysis of the Experimental Evidence Linking Mathematics and Science Professional
Development Interventions to Teacher Knowledge, Classroom Instruction, and Student
Achievement

Kathleen Lynch¹, Kathryn Gonzalez², Heather Hill³, and Ramsey Merritt³

¹Department of Educational Psychology, Neag School of Education, University of Connecticut

²Mathematica Policy Research

³Graduate School of Education, Harvard University

Author Note

Correspondence concerning this article should be addressed to Kathleen Lynch, Neag School of Education, Charles B. Gentry Building, 249 Glenbrook Road, University of Connecticut, Storrs, Connecticut. Email: kathleen_lynch@uconn.edu.

Abstract

Despite evidence that teacher professional development interventions in mathematics and science can increase student achievement, our understanding of the mechanisms by which this occurs – particularly how these interventions affect teachers themselves, and the extent to which teacher-level changes predict student learning – remains limited. The current meta-analysis synthesizes 46 experimental studies of PK-12 mathematics and science professional development interventions to investigate how these interventions affect teacher-level outcomes, including knowledge and classroom instruction, and how these impacts relate to intervention effects on student achievement. Compared with controls, treatment group teachers in PK-12 mathematics and science PD intervention studies demonstrated stronger performance on teacher-level outcomes (pooled average impact estimate: +0.52 SD). Programs with larger impacts on teacher-level outcomes also tended to have significantly larger mean impacts on student achievement. We discuss implications for future research and practice.

Keywords: Professional development, mathematics, science, meta-analysis

**A Meta-Analysis of the Experimental Evidence Linking Mathematics and Science
Professional Development Interventions to Teacher Knowledge, Classroom Instruction,
and Student Achievement**

Strengthening children’s understanding and skills in mathematics and science is a major focus of many nations’ education systems (e.g., National Academies of Sciences, Engineering, and Medicine [NASEM], 2020; OECD, 2007). Given concerns about students’ preparation for a labor market that increasingly values higher-level quantitative and communications skills (Deming, 2017) and the urgent need in many countries to make mathematics and science learning opportunities more rigorous and efficacious (Chmielewski & Reardon, 2016), researchers and policymakers have made significant investments in improving instructional quality in mathematics and science classrooms. A core policy investment by which many nations seek to improve the quality of PK-12 mathematics and science teaching is the provision of teacher professional development (PD) (Desimone, 2009).

In the United States, the country where most of the experimental research on this topic has been conducted, policy logic models posited that improvements in student learning would operate through changes in teachers and teaching (see, e.g., Bethell, 2016; National Research Council [NRC], 2007; Smith & O’Day, 1990; Wilson, 2008). U.S. policymakers in particular targeted two areas: perceived constraints in teachers’ knowledge of the content to be taught, and perceived weaknesses in classroom instruction. Pertaining to teachers’ knowledge, for example, the National Academies’ (2010) report *Rising Above the Gathering Storm, Revisited* noted that pre-service teachers in the U.S. ranked seventh among 15 nations on an international test of mathematics content knowledge at the lower secondary level. Researchers further noted that elementary science teachers often possess important gaps in their content and pedagogical

content knowledge backgrounds, gaps which contribute to challenges in portraying science content to students (Abell, 2007). Meanwhile, in the domain of classroom instruction, nationally representative observational studies concluded that mathematics instruction in U.S. eighth grade classrooms was heavily focused on review of previously learned procedures, and offered students limited opportunities to draw meaningful connections across mathematical concepts (Hiebert et al., 2005), while instruction in U.S. science classrooms featured predominantly lectures and was judged by observers as unlikely to foster student understanding of key science content and the skills needed to conduct scientific inquiry (Banilower et al., 2006).

Perceived constraints in teachers' content knowledge and classroom instruction led scholars and policymakers to design policy logic models theorizing that teachers' participation in PD would influence teachers' knowledge, skills, and instructional practice. Ultimately, these changes were expected to catalyze improved student learning outcomes (Ball & Cohen, 1996; Carlson et al., 2019; Philipp, 2007; for similar logic models in a broader range of countries, see, e.g., Bethell, 2016; Kärkkäinen & Vincent-Lancrin, 2013).

In line with these logic models, in the past two decades, researchers have developed a broad range of teacher PD interventions that target facets of teachers' knowledge, classroom instruction, or both. Beginning in the early 2000s, the Institute for Education Sciences and National Science Foundation's increasing emphases on funding randomized experiments produced a wealth of new studies of teacher PD using designs supporting causal inference. Research syntheses indicate that studies of PD interventions tend to show positive average impacts on students' mathematics and science learning, although with considerable heterogeneity in study outcomes (e.g., Blank & De las Alas, 2009; Lynch et al., 2019; Scher & O'Reilly, 2009; Slavin et al., 2009).

Yet despite the growing evidence base documenting the potential for PD interventions to bolster student achievement in mathematics and science, the teacher-level knowledge and practice mechanisms by which these investments operate remain not well understood. One limitation lies in prior scholarship's inability to empirically test different elements of policymakers' logic model, namely that PD would generate growth in teacher knowledge, improvements in teachers' instruction, and that changes in one or both would predict improvements in student achievement. A second limitation lies in the existing literature's small number of efforts to explain heterogeneity in programs' impacts on teachers, i.e., testing whether we can isolate specific program foci or characteristics that predict stronger teacher impacts.

To remediate this gap in the literature, we conducted a comprehensive meta-analysis of contemporary experimental studies of teacher PD to empirically test several important, yet understudied, components of the core policy logic model for PD in mathematics and science outlined above. In prior work (Lynch et al., 2019), we synthesized STEM classroom interventions' impacts on student-level achievement outcomes. In the current study, we directly build on and extend this prior work by synthesizing two decades of evidence on the impacts of mathematics and science PD on teacher-level outcomes, including knowledge and classroom instruction. We also investigated whether key program content and contextual factors predicted improved teacher outcomes. We then linked interventions' impacts on teacher-level outcomes to impacts on measures of student learning, thereby evaluating the extent to which key elements of a core logic model supporting policy investments into instruction are borne out in the contemporary research base. This study addresses the following research questions:

[RQ1]. What is the causal impact of mathematics and science PD interventions on in-service teachers' outcomes, including knowledge and classroom instruction?

[RQ2] Do specific features and foci of mathematics and science PD interventions predict stronger impacts on in-service teachers' outcomes, including knowledge and classroom instruction?

[RQ3]. Are PD program-induced changes in in-service teachers' outcomes, including knowledge and classroom instruction, linked to improvements in student achievement?

To address these questions, we used exclusively randomized experiments, a 'gold standard' research design that supports causal inference and mitigates many threats to internal validity.

Literature Review

Policymakers and researchers in many countries globally have prioritized improving mathematics and science student outcomes for the past several decades (AAAS, 1990; NGA, 2007; NRC, 2007; OECD, 2007). Many of these efforts called for more mathematics and science literacy among high school and college graduates, with dual attention to meeting employers' needs and fueling innovation in the technology sector. These concerns about workforce readiness dovetailed, chronologically, with widespread dissemination in the field of updated guidance for how students learn content (Bransford et al., 2000), with the result that most mathematics- and science-related policy efforts indicated a preference for conceptual understanding and application of knowledge over rote memorization and basic skills.

Government agencies and researchers advocating for educational reforms often argued that improving student outcomes in mathematics and science would require a coordinated strategy aimed at changing classroom instruction. Advocates of systemic and standards-based reforms (e.g., Smith & O'Day, 1990) recommended a structure widely adopted by policymakers and funders (Knapp, 1997): standards, assessments, and accountability would create policy guidance and incentives for instructional improvement. Within this structure, PD interventions

would increase teachers' and schools' capacity to offer improved instruction.

What Do We Know about the Effects of Teacher Professional Development in Mathematics and Science?

An influential stream of scholarship on teacher PD in the 2000s developed a framework that posited that effective teacher PD should contain several common elements: content focus, active learning, coherence, sufficient duration, and collective participation (Desimone, 2009; Garet et al., 2001). This framework was generative to the field; however, several federally commissioned large-scale studies that aimed to test PD containing these elements had disappointing results (e.g., Garet et al., 2008; 2011; 2016). This raised questions about potential refinements to the field's conception of effective practices for teacher PD, and influenced new scholarly efforts to conduct and synthesize a broad range of PD evaluations. We review results from these efforts below. We organize this review by first summarizing the evidence for impacts on students—looking at the degree to which PD interventions of the types studied in the literature meet their ultimate goal of improving student achievement. We then review evidence on the mechanisms by which PD achieves its results—looking at the impacts on teachers' knowledge and classroom instruction.

Impacts on Students

Several reviews have examined the impacts of mathematics and science teacher PD on student achievement. In prior work (Lynch et al., 2019), we conducted a related meta-analysis of the student-level impacts of teacher professional development and curriculum improvement interventions in PK-12 mathematics and science, finding a pooled mean effect size on student achievement of 0.21 SD. Stronger results were observed in programs that included summer workshops; assisted teachers in using curriculum materials; emphasized improving teachers'

content knowledge, pedagogical content knowledge, and/or understanding of how students learn; and included convenings of teachers to discuss implementation. Taylor et al. (2018) reviewed the student learning impacts of several types of educational programs in elementary and secondary science, including curricula and software programs; teacher professional development; and new instructional techniques. Interventions could be conducted in either schools or research labs, and delivered either by teachers or other providers (e.g., researchers). The authors found that these interventions produced a pooled mean effect size of 0.49 SD on student achievement outcomes, and that researcher-developed tests showed larger mean effects than did standardized assessments.

Albeit not a meta-analysis, Kennedy (2016) reviewed the literature using graphical comparisons of studies, and concluded that programs that focused on a number of different instructional issues, including improving student behavior, bolstering student participation, making students' thinking and problem-solving visible, and presenting the content of the curriculum all appeared similarly likely to improve student achievement. Programs that fostered teachers' intellectual involvement with the material and featured voluntary, rather than mandatory, enrollment of teachers also tended to be more effective. Other authors have reviewed subsets of these literatures (in elementary mathematics, Pellegrini et al., 2021; in secondary science, Cheung et al., 2017; in elementary science, Slavin et al., 2014), or pooled math and science programs with those focusing on a range of content areas (e.g. literacy skills, economics, etc., Sims et al., 2023), generally finding positive student achievement effects.

Evidence on the Mechanisms by Which Teacher Professional Development Achieves its Effects: Impacts on Teachers' Knowledge and Classroom Instruction

Our first research question examines the impacts of PD interventions on proximal teacher outcomes, including teachers' knowledge and classroom instruction. As elaborated in the logic model below, understanding impacts on teachers' outcomes, including knowledge and classroom instruction, is important for unpacking the mechanisms by which teacher PD achieves its results. A handful of prior reviews have examined facets of teacher-level outcomes of PD interventions. We briefly review the methods and results from these reviews below, before describing the contributions of the present study to the evidence base.

We identified five prior meta-analyses that examined impacts of different types of teacher PD for a sample that included mathematics and/or science teachers on instructional practice outcomes (Egert et al., 2018 [specific to early childhood]; Garrett et al., 2019; Kowalski et al., 2020; Kraft et al., 2018 [specific to coaching]; Scher & O'Reilly, 2009); and one that included an analysis of impacts of teacher training activities on teachers' knowledge (Kowalski et al., 2020). The earliest of these, Scher and O'Reilly (2009), reviewed the literature published through 2004 on elementary and secondary in-service teacher professional development in mathematics and science. The authors reported pooled mean effects of teacher PD interventions on teacher attitudes of 0.45 SD, and effects on instructional practices of 0.63 SD. However, the authors were able to find only five studies that examined impacts on teacher attitudes, and four that examined impacts on teacher practices. They further cautioned that none were randomized trials, and many of the included studies had "significant methodological weaknesses" (p. 223), including not requiring evidence of group baseline equivalence in quasi-experiments. Also considering instructional practice, Garrett et al. (2019) examined the effects of in-service teacher professional development programs across a range of content areas (e.g., reading, math, science, social studies) on teacher practice outcomes, specifically teachers' scores on classroom

observation indicators. Pooling across these interventions, the authors found a mean effect size of 0.42 SD on observation scores.

Additionally, Kraft et al. (2018) conducted a meta-analysis of teacher coaching programs, which they operationalized as “in-service PD programs that incorporate coaching as a key feature of the model” (p. 553), conducted with PK-12 teachers teaching in a range of content areas (e.g., reading, math, science). Combining experimental and rigorous quasi-experimental studies, Kraft et al. found a pooled mean impact of 0.49 SD on teacher practice outcomes. In a meta-analysis of in-service training programs for educators in early childhood center-based care settings, Egert et al. (2018) examined experimental and quasi-experimental studies, finding pooled mean impacts on childcare classroom quality ratings of 0.68 SD.

Specific to science, Kowalski et al. (2020) conducted the one prior meta-analysis we could identify that examined outcomes on teacher knowledge. The authors meta-analyzed a set of quasi-experimental and experimental studies of science interventions for pre-service and/or in-service teachers. Included studies aimed to improve knowledge, practices, and/or attitudes; were conducted in either schools or research labs; and, for quasi-experiments, were not required to demonstrate the comparability of treatment and comparison groups at baseline. The authors found a pooled mean effect size of 0.49 SD on knowledge outcomes, and noted that effect sizes from instructional practice and attitudes outcomes were substantially smaller than those identified for knowledge outcomes.

Programmatic Features and Contexts of PD Programs in Mathematics and Science that May Moderate Effects on Teacher-Level Outcomes

Our second research question examines how the features and contexts of PD interventions may moderate PD programs’ impacts on teachers. PD interventions vary on key

characteristics that may explain variation in their impacts on teacher outcomes, including knowledge and classroom instruction. First, understanding the extent to which *PD duration* may predict the magnitude of programs' impacts on teachers' knowledge and instruction is important given the high costs of both teachers' and PD providers' time, and the costs of substitutes to replace in-service teachers participating in PD during school (Odden et al., 2002). While Desimone's model included 'sufficient duration' as a critical characteristic of high-quality PD, shorter-duration PD programs could potentially focus on more discrete skills, which could effectively contribute to incrementally building teachers' repertoires (e.g., Siegle & McCoach, 2007). Meta-analyses have yielded disparate findings about the relationship between PD duration and student learning outcomes (Kennedy, 2016; Lynch et al., 2019; Yoon et al., 2007), raising questions about the links between duration and specific proximal outcomes of in-service mathematics and science teachers' knowledge and instruction.

As well, PD interventions may or may not include the *introduction of new curriculum materials* planned for adoption in the classroom. Curriculum materials themselves may be educative for teachers, building teachers' content knowledge and supporting new instructional practices (Davis & Krajcik, 2005). Grounding teacher PD in the specific context of curricular materials may also help teachers to map PD content more directly onto their existing curriculum-connected knowledge base and future instructional interactions (e.g., Davis & Krajcik, 2005; Penuel et al., 2007), thereby increasing the potential for PD participation to catalyze knowledge growth and changes to classroom instruction. Given that prior work (Kraft et al., 2018; Lynch et al., 2019) has identified the presence of curriculum materials accompanying professional development as a predictor of larger PD program impacts on students, it is important to understand whether curriculum inclusion predicts stronger impacts on teachers' outcomes,

including classroom instruction and/or knowledge, as well.

PD interventions also vary in their programmatic foci. PD programs may or may not include an *explicit programmatic focus on improving teachers' knowledge*, which may include facets of teachers' content knowledge and/or pedagogical content knowledge. Various aspects of mathematics and science teachers' knowledge have been conceptualized as resources for instruction and student learning (e.g., Fennema & Franke, 1992; Institute of Medicine, 2010), motivating teacher knowledge development as an explicit component of some PD programs. One prior meta-analysis, Kowalski et al. (2020) did not find a significant relationship between science programs' inclusion of content deepening experiences for pre-service and/or in-service teachers and the pooled teacher outcomes they examined, although they urged caution in interpretation as nearly all of the programs they examined included content deepening.

Teacher PD programs in mathematics and science also may or may not incorporate a focus on *content-specific or content-general instructional strategies*. Content-specific instructional strategies may include, for example, mathematics-specific discussion strategies or investigation-based approaches to teaching science; while content-general instructional strategies may include, for example, strategies to improve classroom climate, incorporating frequent quizzing, and so forth. Content-general instructional strategies may complement math- and science-specific materials and approaches (e.g., Smith & O'Day, 1990), contributing to overall improvements in teachers' classroom management and practices, and potentially fostering an improved classroom climate where teachers are better able to build and practice the knowledge they are gleaned from PD participation.

A third programmatic focus is PD programs' inclusion of *content-specific formative assessment*. This focus has been of consistent interest within scholarship on teacher learning

(e.g., Black & Wiliam, 1998, 2010) and has been examined in a growing number of empirical studies (e.g., Lang et al., 2014; Schneider & Meyer, 2012). Formative assessment techniques, in which teachers employ the information they glean from assessment to strategically adjust their instruction (Black & Wiliam, 2010), have the potential to enrich both student classroom experiences and teachers' attention to student learning, suggesting that a focus on these approaches in PD could be expected to improve both teachers' knowledge and their classroom instruction.

Lastly, elements of the *contexts in which interventions are conducted* may influence the effects of PD interventions. Recognizing that PD interventions have been implemented at different scales and in varied settings, we follow the approach of other meta-analyses in PK-12 education and explore heterogeneity of intervention impacts across contexts (e.g., Kraft et al., 2018; Sheridan et al., 2019; Sirin et al., 2005; Stockard et al., 2018). Prior research suggests that smaller-scale demonstration studies may be expected to yield larger impacts compared to scaled-up versions of the programs, due to their tendency to offer more intensive resources to a limited number of teachers (Hill, 2004; Hill et al., 2019). Program impacts on instructional quality did appear smaller in the scaled-up programs reported in Kraft et al.'s (2018) meta-analysis of programs that contain coaching, and in Garrett et al.'s (2019) study of PD programs across a range of content areas, though not in Egert et al.'s (2018) meta-analysis of early childhood in-service programs.

At the setting level, schools in different geographic locales may experience different contexts and resource environments that affect the comprehensiveness with which instructional initiatives are carried through (e.g., Wilson, 2013). Meanwhile, examining the effectiveness of instructional interventions by sociodemographic characteristics of participating students is

important given the critical need to improve mathematics and science learning opportunities for all students, including students in high-poverty settings, students from communities of color, and students learning English as a new language (e.g., Oakes, 1990). We are not aware of prior meta-analyses examining these variables in the context of the influence of PD on teacher outcomes.

Can We Identify Links Between PD Impacts on Teachers' Outcomes and Gains in Student Achievement?

As described above, we hypothesized that PD interventions will affect teachers' outcomes, including knowledge and classroom instruction, and that some interventions will be more effective than others. A logical next question and the third aim of our paper is: To the extent that programs are able to improve teachers' outcomes, including knowledge and/or classroom instruction, how might this matter for impacts on student achievement? Different scholars and policy interventions have placed different degrees of emphasis on the importance of improving teachers' content knowledge versus improving their practices, raising questions about whether improvements in knowledge or practices may be more strongly related to improvements in ultimate student achievement. We briefly summarize these different perspectives below, to motivate our third analysis where, using a similar approach as others (Egert et al., 2018; Kraft et al., 2018), we test via a regression approach the relationships between PD effects on teachers and their effects on students in the logic model.

Theorizing the Importance of Improving Teachers' Knowledge to Strengthen Student Outcomes

A substantial history of education scholarship and policy has been concerned with teachers' content knowledge, and theorized that it has a major influence on students' opportunities to learn science and mathematics (e.g., Conference Board of the Mathematical

Sciences, 2012; Fennema & Franke, 1992; NASEM, 2007; National Research Council, 1996; No Child Left Behind Act, 2001; Wu, 2011). For instance, beginning in the 1980s and 1990s, a stream of research employed observations to draw attention to how constraints in teachers' subject matter knowledge in mathematics and science could lead to confusing or incorrect delivery of curricular content to students (Cohen, 1990; Heaton, 1992; Putnam et al., 1992; Stein et al., 1990). Subsequent work developed a range of quantitative measures of facets of teachers' knowledge, and worked to link teachers' scores on these measures with indicators of instructional quality and students' achievement gains (e.g., Jacob et al., 2018; Hill et al., 2005; Rockoff et al., 2011). This line of research supported the argument that improving content knowledge was likely a critical component to improving classroom instructional quality and, ultimately, student outcomes.

Aligned with these views and this scholarship, improving teachers' content knowledge in mathematics and science has been a major focus of federal education policy and funding investments in the U.S. The National Academies' influential *Rising Above the Gathering Storm* report contended that "We need to reach all K–12 science and mathematics teachers and provide them with high-quality continuing professional development opportunities—specifically those that emphasize rigorous content education" (p. 119). The U.S. federal government made major investments emphasizing teachers' content knowledge improvement. The National Science Foundation and US Department of Education spent approximately \$1.2 billion between 2002–2007 on 'math-science partnership programs' which provided content-focused training for pre-service and in-service teachers (Hill, 2011). As Kennedy (2016) observed, "There is a tendency for critics of education to press for PD that addresses primarily, or only, subject matter knowledge" (p. 971).

*Theorizing the Importance of Improving Classroom Instruction to Strengthen Student**Outcomes*

Alongside policies and interventions weighted toward improving teachers' knowledge, other scholars have placed more relative emphasis on the importance of improving teachers' classroom instruction as central to strengthening student achievement in science and mathematics. Early research in the 'process-product' literature tradition (e.g., Brophy & Good, 1986) aimed to categorize facets of teachers' instructional behaviors that were associated with better student achievement. Observational research has also drawn attention to weaknesses in classroom instruction (e.g., Hill et al., 2018). For example, in a nationally representative observational study of science classrooms, Banilower et al. (2006) found that lecture formats dominated instructional activities, with limited time allocated to hands-on activities or small group work.

Scholars' emphasis on topics such as questioning (Godbold, 1973), discussion routines (Pehmer et al., 2015), talk moves (Michaels & O'Connor, 2015), and coaching programs focused on discussing teachers' instruction using classroom observation rubrics (Kraft & Hill, 2018; Wayne et al., 2023) also place an implicit emphasis on improving teachers' practices as a central needed lever to improve student achievement. Influential popular books have also disseminated the message that improvements in teachers' practices are of major importance to bolstering student achievement (Lemov, 2021). Meanwhile, many PD programs and policy investments with a focus on instructional practices also have a dual focus on improving aspects of teachers' content knowledge, reflecting the view that changes in one area could spark or reinforce changes in another (e.g., Ball & Cohen, 1999; Clarke & Hollingsworth, 2002), although the relative emphasis placed on teachers' knowledge versus practices varies.

Given these different emphases, it remains unclear but important to know how PD-induced improvements in teachers' outcomes, including content knowledge and classroom instructional practices, may be related to improvements in student learning. Only two meta-analyses that we can locate linked teacher to student outcomes. Using a subset of studies that collected information on both classroom instruction and student achievement, Egert et al. (2018) and Kraft et al. (2018) found some evidence of positive correlations between impacts on instruction and improvements in student achievement in their meta-analyses, specific to the early childhood and teacher coaching contexts, respectively – though in Kraft et al. (2018) this correlation was not statistically significant. Neither study examined relationships between impacts on teacher knowledge and impacts on student learning, nor did they examine these relationships specifically in the context of in-service teacher PD in mathematics and science. We examine these questions in our analysis.

The Present Study

The present study aims to estimate the causal effects of PK-12 mathematics and science teacher PD interventions of the kinds studied in the experimental literature on teacher outcomes, including knowledge and classroom instruction, and to understand how these impacts correlate to student outcomes, using the most rigorous evidence available. Prior reviews, while certainly informative, have typically examined only measures of instruction, and not specifically in mathematics and science (e.g., Egert et al., 2018, in early childhood; Garrett et al., 2019, across a range of content areas; Kraft et al., 2018, specific to coaching studies across a variety of subjects). These meta-analyses largely included studies on literacy and social-emotional learning, leaving questions about mathematics- and science-specific programs' impacts on teacher outcomes unanswered. One prior meta-analysis has synthesized the impacts of educator training

programs on science educators' knowledge and practices (Kowalski et al., 2020); while quite useful, this science-specific review did not specifically examine in-service teachers, nor did it require that included quasi-experiments meet methodological criteria such as demonstrating group baseline equivalence (What Works Clearinghouse, 2022).

Thus more can be learned from the literature, by specifically examining the impacts of PD programs on in-service mathematics and science teachers using rigorous contemporary evidence from RCTs. We further probe potential explanations for observed variations in PD programs' impacts on teachers by empirically testing a set of important programmatic and contextual features of PD studies (described below).

Second, as prior reviewers have pointed out (e.g. Garrett et al., 2019; Kowalski et al., 2020), no existing reviews have empirically tested the links in the logic chain between interventions' proximal impacts on teachers and their ultimate influences on student learning. However, doing so is important. The core purpose of PD investments is student learning, and the need to understand the mechanisms that produce that learning is crucial. Therefore, the current review takes the critical step of investigating how changes in teachers' outcomes catalyzed by experimentally evaluated interventions predict improvements in student achievement. This analysis approximates as closely as possible an analytic test of the system of relationships between PD; teacher outcomes, including knowledge and instruction; and student achievement.

Logic Model for the Effects of PD Interventions

Adapting the model proposed by Garrett et al. (2019), Figure 1 provides an illustration of hypothesized pathways by which PD interventions that aim to influence teachers' knowledge and/or instruction may lead to changes in student learning. This model structures our analysis. The left panel of the Figure 1 shows an illustration of potential mechanisms by which PD

interventions are hypothesized to lead to changes in student learning. The right panel of Figure 1 illustrates the subset of mechanisms that we examine in the current analyses.

In the full model, illustrated in the left panel of Figure 1, teachers' participation in a PD may lead to changes in their knowledge (Path *a*) and/or changes in classroom instruction (Path *b*); changes in instruction may also operate through changes in knowledge (Path *f*). These changes, in turn, affect how teachers interact with students in classrooms, leading to changes in student outcomes (Paths *h* and *i*). Prior meta-analyses have primarily analyzed Paths *b* and *i* (Egert et al., Garrett et al., 2019; Kraft et al., 2019). Our data provided us a novel opportunity to examine Paths *a* and *h*.

Interventions may also affect student outcomes independent of teacher-level mediators, such as through the provision of new materials and student supports (Path *c*). Context (e.g., characteristics of the setting where the intervention was implemented; teachers' existing beliefs, experience, and attitudes) and intervention characteristics (e.g., intervention foci and features) may also moderate impacts on teacher knowledge, classroom instruction, and student outcomes (Paths *d*, *e*, and *g*).¹

Our analyses permit us to test several elements of this logic model, illustrated in the right panel of Figure 1. In our first analysis [RQ1], we directly tested the impact of mathematics and science PD interventions on teacher outcomes, specifically knowledge (Path *a*) and classroom instruction (Path *b*). Doing so addresses the first research question: *What is the causal impact of mathematics and science PD interventions on in-service teachers' outcomes, including knowledge and classroom instruction?*

¹ We note that while we are able to investigate the magnitude of observed relationships among specific variables, our meta-analysis models do not permit testing potential reciprocities in these relationships (e.g. Clarke & Hollingsworth, 2002; Goldsmith et al., 2014). As such, consistent with the tested models, we represent these pathways unidirectionally.

Given the importance of understanding factors that may contribute to variation in interventions' impacts on teacher outcomes, we also examine Paths *d* and *e*, addressing the second research question [RQ2]: *Do specific features and foci of mathematics and science PD interventions predict stronger impacts on in-service teachers' outcomes, including knowledge and classroom instruction?* In contrast to other recent meta-analyses, which had more heterogeneity in content areas and program foci, centering our analysis on mathematics and science-specific programs allows us to test the influence of key moderators on teacher outcomes specifically within these content areas.

In our third set of analyses, we examined the links between impacts on teacher outcomes, including knowledge (Path *h*) and classroom instruction (Path *i*) and improvements in student learning, addressing the third research question [RQ3]: *Are PD program-induced changes in in-service teachers' outcomes, including knowledge and classroom instruction, linked to improvements in student achievement?*

We describe our meta-analytic search procedures and analyses in the next section.

Methods

Defining Mathematics and Science PD Interventions

For the current meta-analysis, we defined mathematics and science PD interventions to include programs that aimed to improve student learning in mathematics and science via teacher PD for in-service teachers. This definition excludes interventions that do not focus on improving classroom instruction, such as out-of-school time tutoring, summer programs, or homework interventions, and those that do not involve in-service classroom teachers, such as interventions in which researchers provided instruction directly to students. Included programs ranged from those providing PD only (e.g., Dash et al., 2012; Jayanthi et al., 2017; Piasta et al., 2015), to

those that included a relatively brief PD that focused on a set of new curriculum materials intended for implementation (e.g., Miller et al., 2007; Resendez & Azin, 2009).

We only included studies that used randomized experimental designs. We restricted our sample in this way to ensure that teachers in the treatment and control groups were similar on all observed and unobserved characteristics, meaning that differences in teacher knowledge and instruction between the two groups can be attributed to the impact of the intervention. We excluded studies that formed treatment and comparison groups in other ways (for example, by comparing teachers who elected to participate in a PD program against teachers who did not) because it would be impossible to disentangle the impact of the program from other factors, such as motivation differences between the groups.

The current meta-analysis pooled mathematics and science studies for conceptual and technical reasons. First, many nations' education systems have a dual focus on improving mathematics and science education (NASEM, 2020; NRC, 2013), given the need for improved instruction in these areas and their joint importance in preparing students for the contemporary labor market. Additionally, the observation that teachers are less likely to have specialized training in these subjects as compared with ELA/reading and other content areas (e.g., NASEM, 2007; 2020; 2022) has led many scholars and policymakers to hypothesize unique kinds of PD needs for teachers in mathematics and science. These considerations have led to the funding of numerous initiatives aimed at improving instruction in both mathematics and science (e.g., the United States' Title II's Math-Science Partnerships [Hill et al., 2019] and National Science Foundation's Systemic Initiatives program [Hamilton et al., 2003]), and motivated earlier meta-analyses that considered instructional interventions in both mathematics and science (Blank & de las Alas, 2009; Scher & O'Reilly, 2009). Prior research has argued that notwithstanding their

obvious disciplinary distinctiveness, mathematics and science share elements of overlapping subject matter cultures and intersecting visions for student learning;² in turn, these historically have informed joint education policy and federal grant funding initiatives pertaining to instruction in both content areas (Knapp, 1997). Second, mathematics and science PD interventions tend to share similar approaches to improvement – e.g., the pairing of PD with curriculum materials, or providing information on student learning as a means to changing teachers’ practice. Including both mathematics and science studies increases our statistical power to test the efficacy of these approaches. Third, our dual focus on mathematics and science allowed us to restrict the review to include only studies using experimental designs that permit causal inference while maintaining sufficient statistical power to conduct a quantitative meta-analysis.

Search and Screening Procedures

The procedures employed are similar to those we used in a prior meta-analysis focused on student-level impacts of STEM classroom interventions (Lynch et al., 2019). We applied the following search procedures to capture relevant published and unpublished experimental studies of mathematics and science PD interventions. To be included in the meta-analysis, studies had to meet the following criteria relating to study design, intervention, sample, and outcomes: (1) Include students in grades PK-12; (2) Focus on classroom-level instructional improvement in mathematics and/or science through professional development, which may or may not be coupled with the introduction of novel curriculum materials; (3) Employ a randomized

² As Knapp (1997, p. 232) wrote of federally supported systemic reforms of the 1990s time period, “The current generation of systemic reforms presumes an elaborate, ambitious vision of mathematics and science learning, one that emphasizes the learner's understanding of central ideas and processes; application of ideas and skills to nonroutine, complex problems; in-depth immersion in important themes and topics rather than exhaustive coverage of material; active mental (and often physical) engagement with scientific phenomena; and exposure to authentic, real-world phenomena, problems, and scientific activities.”

experimental design comparing a treatment group to a business as usual control group; (4) Be published in 2001 or later, coinciding with the passage of the No Child Left Behind Act, which contributed to catalyzing an increased focus on effectiveness research in education; (5) Be written in English; and (6) Report sufficient data to calculate one or more effect sizes for both teacher and student outcomes. Examples of excluded studies included those focused on postsecondary education (e.g., Rust, 2011; Sullins et al., 2010) or on subject areas other than mathematics or science (e.g., Connor et al. 2013; Farver et al., 2009); and studies that did not use random assignment. We focused specifically on studies that include both teacher and student outcomes because they are directly aligned with the intended impacts of PD interventions in our theoretical model. The publication year span that we employed for study inclusion is similar to that used by the What Works Clearinghouse, which generally declines to review studies that are more than two decades old due to significant cross-time shifts in the educational landscape (WWC, n.d.).

We searched in several channels. First, we examined the reference lists of previous research syntheses for studies that examined the topics of teacher PD and curriculum improvement in mathematics and science (e.g., Blank et al., 2008; Cheung et al., 2017; Garrett et al., 2018; Gersten et al., 2014; Kennedy, 2016; Kowalski et al., 2020; Scher & O'Reilly, 2009; Slavin & Lake, 2008; Slavin et al., 2009; Slavin et al., 2014; Taylor et al., 2018; Yoon et al., 2007). We next performed electronic library database searches using Academic Search Premier, Ed Abstracts, EconLit, ERIC, PsycINFO, and ProQuest Dissertations & Theses, for the period 2004 to 2024. As in Lynch et al. (2019), these searches were performed using subject-related search terms adapted from Yoon et al. (2007) and methodology-related search terms adapted

from Kim and Quinn (2013).³ We also examined relevant websites, including American Institutes for Research, Empirical Education, Inter-American Development Bank, Mathematica, MDRC, Regional Education Labs, WWC, the World Bank, and Society for Research on Educational Effectiveness (SREE) conference abstracts. Finally, we searched the NSF Community for Advancing Discovery Research in Education and IES websites, and developed a list of all potentially relevant mathematics and science grant awards from the years 2002 to 2020. We searched electronic databases and the Web to identify relevant studies produced based on these grants, and contacted study PIs to request reports if we could not identify publications with impact results from their grants. For the current review, we included the results of an initial round of searches we conducted for Lynch et al. (2019), then we extended the initial database of studies via an additional search for reports published through 2024. Via these search streams, we located 12,086 records from database searches and 1,402 records from other sources. Following the removal of duplicates, this yielded 10,777 studies.

Second, raters screened the titles and abstracts of these studies for whether or not they were conducted at the PK-12 level, reported information on student outcomes, and related to

³ The search parameters used to conduct electronic database searches are as follows: (“professional development” OR “faculty development” OR “Staff development” OR “teacher improvement” OR “inservice teacher education” OR “peer coaching” OR “teachers’ institute*” OR “teacher mentoring” OR “Beginning teacher induction”; “teachers’ Seminar*” OR “teachers’ workshop*” OR “teacher workshop*” OR “teacher center*” OR “teacher mentoring” OR curriculum OR instruction*) AND (“Student achievement” OR “academic achievement” OR “mathematics achievement” OR “math achievement” OR “science achievement” OR “Student development” OR “individual development” OR “student learning” OR “intellectual development” OR “cognitive development” OR “cognitive learning” OR “Student Outcomes” OR “Outcomes of education” OR “educational assessment” OR “educational measurement” OR “educational tests and measurements” OR “educational indicators” OR “educational accountability”) AND (“*experiment*” OR “control*” OR “regression discontinuity” OR “compared” OR “comparison” OR “field trial*” OR “effect size*” OR “evaluation”) AND (“Math*” OR “*Algebra*” OR “Number concepts” OR “Arithmetic” OR “Computation” OR “Data analysis” OR “Data processing” OR “Functions” OR “Calculus” OR “Geometry” OR “Graphing” OR “graphical displays” OR “graphic methods” OR “Science*” OR “Data Interpretation” OR “Laboratory Experiments” OR “Laboratory Procedures” OR “Experiment*” OR “Inquiry” OR “Questioning” OR “investigation*” OR “evaluation methods” OR “laboratories” OR “biology” OR “observation” OR “physics” OR “chemistry” OR “scientific literacy” OR “scientific knowledge” OR “empirical methods” OR “reasoning” OR “hypothesis testing”).

mathematics and/or science content and/or instructional strategies. During this phase, raters identified 840 studies as meeting the preliminary relevance criteria. These studies then underwent full-text screening, in which raters evaluated each study's full text to determine whether it met the full criteria described above, including requiring a randomized experimental research design, reporting impacts on both teacher and student outcomes, and being published within the review period. See Figure 2 for a PRISMA diagram.

Analytic Sample

The final meta-analytic sample includes 46 studies contributing 200 effect sizes for teacher outcomes and 126 effect sizes for student achievement outcomes. These include separate effect sizes for each study-reported assessment, treatment contrast, and teacher sample. Because we examine interventions' impacts on all teacher outcomes as well as on teacher outcomes disaggregated by type, we also categorized each teacher outcome as either a teacher knowledge or classroom instruction outcome. Following others (Kowalski et al., 2020), while during initial coding we first classified teacher knowledge measures as either content knowledge or pedagogical content knowledge (PCK), we consolidated these measures for the analysis because PCK was measured in too few studies to permit separate analysis, thus it was preferable to pool it with the other knowledge measures rather than discarding the data altogether (Kowalski et al., 2020). We further argue that while our approach of categorizing outcomes as knowledge and practice may not be the only possible approach, it is a reasonable one, based on the way these variables are reported in empirical studies in our dataset. For all analyses, we first considered all teacher outcomes simultaneously, and then considered each group of outcomes separately.

Study Coding

We developed codes for meta-analytic moderator analyses based on a review of the

existing literature on teacher PD, including prior research syntheses (e.g., Kennedy, 2016; Kraft et al., 2018; Scher & O'Reilly, 2009). The codes examined in the meta-analysis are described below. To complete study coding, after establishing interrater reliability at the start of the coding process (i.e., 80% agreement), study authors and trained research assistants coded full-text studies in pairs. Each researcher in the pair first coded the study independently, then pairs met to resolve all coding disagreements through discussion (for more details about code development, see Lynch et al., 2019).

We grouped potential moderators into three categories: (1) *intervention features*; (2) *PD focus*; and (3) *contextual features*. We first coded studies for two *intervention features*, which indexed (1) whether programs combined PD with curriculum materials or provided PD only; and (2) PD duration, measured in contact hours.

Second, we coded each study for evidence that the PD intervention included an explicit *PD focus* on (a) improving teacher knowledge, which could include facets of content knowledge, pedagogical content knowledge, and/or knowledge of how students learn; (b) content-specific and/or content-generic instructional strategies; and/or (c) content-specific formative assessment. These codes allowed us to examine whether programmatic focus on each of these areas moderated program impacts on teacher outcomes. Note that, as is often the case in meta-analytic moderator analysis, a given intervention could include more than one of these foci, for example, foci on both improving teacher knowledge and content-specific formative assessment.

Lastly, we examined how the impacts of PD interventions varied based on the *contextual features* of the intervention, including intervention scale, setting type, and characteristics of the student sample. To do so, we coded interventions in four ways. First, we coded for teacher sample size as a measure of intervention scale. Second, we coded for whether the intervention

took place in an urban versus suburban or rural school setting, recognizing that the geographical environments in which schools are situated are contexts that may contribute to variation in the local resources available for PD interventions' implementation. Third, we coded for demographic and socioeconomic characteristics of the student sample (e.g., the percent of low income or free or reduced-price lunch-eligible students, the percent of emergent bilingual students/students learning English as a new language, and the percent of students of color), recognizing the importance of examining how interventions impact students of color and other underrepresented groups historically marginalized in mathematics and science learning (Martin, 2009; Robinson et al., 2016). Fourth, we coded for student grade level to understand how intervention impacts differed across preschool, elementary, and secondary grades.

Effect Size Calculation

We computed standardized mean difference effect sizes for impacts on teacher outcomes and student achievement outcomes using Hedges's g :

$$g = J \times \frac{(Y_E - Y_C)}{S^*}$$

Where \underline{Y}_E indicates the average treatment group outcome, \underline{Y}_C indicates the average control group outcome, S^* indicates the pooled within-group standard deviation, and J is a correction factor to avoid small sample bias.

We computed effect sizes based on author-reported effect sizes, raw means and standard deviations, and other author-reported results, using the *Comprehensive Meta Analysis* software package for most instances. Where possible, we calculated effect sizes that were adjusted for covariates (e.g., pretest scores). The following decision rules were applied to compute effect sizes: For cases in which a study presented a standardized mean difference effect size (e.g., Cohen's d), the author-reported effect sizes were converted to Hedges's g . In cases where the

study did not report a standardized mean difference effect size but did present a covariate-adjusted mean difference (e.g., a coefficient from a regression model) and unadjusted standard deviations, a standardized mean difference effect size was computed and converted to Hedges's *g*. In cases where the study did not report adjusted mean differences, effect sizes derived from posttest means and standard deviations were computed. If this information was not available, we computed effect sizes from other results (e.g., results of ANOVAs). We applied the same decision rules to compute effect sizes for teacher outcomes and student outcomes.

Impacts of PD Interventions on Teacher Outcomes

We estimated meta-regression models to examine the impacts of PD interventions on teacher-level outcomes, both pooled and separately for knowledge and classroom instruction outcomes. Many of the included studies yielded multiple effect sizes for impacts on teacher-level outcomes. We expect there are dependencies among effect sizes within our data, which violates the statistical assumptions required for the use of traditional meta-analytic methods. For example, some studies examined multiple outcome measures for one underlying construct or multiple related constructs for the same sample. We expect the sampling errors of these effect sizes to be correlated within the same study (i.e., correlated effects). In other cases, we expect multiple effect sizes within the same study to be independent—for example, because the study reports effect sizes for multiple participant samples or based on evaluations of multiple treatments. These types of dependencies are referred to as hierarchical effects (Tanner-Smith & Tipton, 2014).

As our sample of effect sizes is likely to include both types of dependences, we use a correlated and hierarchical effects (CHE) model with robust variance estimation (RVE) that combines both dependence structures. The CHE approach allows for both between-study

heterogeneity and within-study heterogeneity in true effect sizes, and for correlations between effect sizes from the same study (Pustejovsky & Tipton, 2021). This approach has been adopted in recent meta-analyses (e.g., Atit et al., 2022; Vembye et al., 2024; Waheed, 2023; Wijnia et al., 2024), and allows us to include multiple effect sizes from the same study and avoid the information loss that would arise from excluding effect sizes or calculating mean effect sizes within each study. The use of CHE with RVE guards against model misspecification given the complexities inherent in datasets with dependencies among effect sizes (Tipton & Pustejovsky, 2015). We used the *clubSandwich* (Pustejovsky, 2020) and *metafor* (Viechtmbaur, 2020) packages in R to estimate all CHE models and used the recommended value for the assumed correlation between effect sizes of 0.80 (Tanner-Smith & Tipton, 2014).

We first estimated separate unconditional CHE models to estimate the overall impact on all teacher outcomes, including teacher knowledge and classroom instruction outcomes simultaneously. We then estimated separate unconditional CHE models to estimate overall impacts on teacher knowledge and classroom instruction outcomes separately. An advantage of the CHE model is that it estimates the between-studies variance component, which we report as a measure of between-study heterogeneity in effect sizes (Tanner-Smith et al., 2016).

To examine moderators of program impact, we first examined whether *intervention features* moderated program impacts on pooled teacher outcomes, as well as classroom instruction outcomes and teacher knowledge outcomes, by fitting conditional CHE models to examine whether impacts differed for interventions that provided PD and new curriculum materials, rather than PD only. Second, we fit additional conditional models to examine whether impacts differ based on PD duration by including the number of PD contact hours as a moderator.

We then fit conditional CHE models to examine whether *PD foci* moderated study impacts on pooled teacher outcomes, as well as classroom instruction outcomes and teacher knowledge outcomes. Specifically, we fit a series of models that included indicators for whether the PD focused on aspects of *teacher knowledge* (teacher content knowledge, pedagogical content knowledge, and/or teacher knowledge of how students learn) as well as indicators for whether the PD focused on aspects of *classroom instruction* (generic instructional strategies and content-specific formative assessment).

Finally, we fit conditional CHE models to examine whether impacts differed based on *features of the PD study contexts*. These characteristics included the size of the teacher sample and various student and school characteristics, including student income (percent low-income or eligible for free or reduced-price lunch; whether a majority of students were low-income or eligible for free or reduced-price lunch); student emergent bilingual status/status as learners of English as a new language; student race/ethnicity; and school district urbanicity (urban vs. suburban or rural). As patterns of missing data varied across student and school characteristics, we estimated separate models that considered each characteristic separately.

Conditional models also featured additional study characteristics as covariates: whether effect sizes were adjusted for covariates and whether interventions focused on mathematics or science. There was within-study variability in some moderators and covariates (e.g., contact hours varied across multiple treatment-control contrasts); in these cases we followed the recommended approach of including the study-level mean (Tanner-Smith & Tipton, 2014).

Linking Impacts on Teacher and Student Outcomes

To examine the associations between intervention impacts on teacher-level outcomes and intervention impacts on student-level outcomes, we adapted the approaches used in recent meta-

analyses (e.g., Egert et al., 2018; Kraft et al., 2018). First, we calculated mean effect sizes for impacts on teacher outcomes and student achievement for each treatment-control contrast in our sample. To examine whether impacts on teacher knowledge and classroom instruction are separately associated with impacts on student achievement, we calculated three mean effect sizes for impacts on teacher outcomes: (1) mean effect sizes for impacts on all teacher outcomes; (2) mean effect sizes for teacher knowledge; and (3) mean effect sizes for classroom instruction.

We examined the links between intervention impacts on teacher outcomes and intervention impacts on student outcomes, rather than the correlations between teacher and student outcomes, for two reasons. First, correlations between teacher scores on knowledge and/or classroom instruction measures and student achievement outcomes were not reported by most studies, and estimates derived from prior literature may not be applicable to the current context of contemporary RCTs of PD interventions in STEM. Second, by looking at the correlation between intervention-induced changes in teacher and student outcomes (i.e., correlations between the intervention impact estimates for teacher and student outcomes, respectively), we get as close as possible, given the limitations of the data, to understanding how exogenous changes in teacher knowledge and classroom instruction affect student achievement outcomes. Implicit in our approach is the assumption that intervention impacts on student outcomes occur only through changes in teacher-level outcomes, which may not hold if interventions have effects on other, unobserved factors that shape student achievement. Nevertheless, this approach is preferable to a meta-analysis of correlations between teacher and student outcomes, as these correlations would likely be confounded by a number of teacher, student, and contextual factors.

Some studies ($k = 9$) in our sample reported impacts based on multiple treatment-control

contrasts. In the presence of within-study, between-treatment variation in teacher and student effect sizes, aggregating effect sizes to the study level could obscure the links between impacts on teachers and students. Therefore, we calculated mean effect sizes at the treatment level rather than at the study level to more directly test whether programs that supported improvements in teacher outcomes also increased student achievement. We also confirmed that we obtained similar results if we aggregated effect sizes to the study level rather than the treatment level.

After calculating treatment-level mean effect sizes for teacher and student outcomes, we then estimated a series of regression models predicting mean impacts on student outcomes as a function of mean impacts on teacher outcomes. We estimated three separate models, including each of the treatment-level mean effect sizes for impacts on teacher outcomes described above as predictors. These models were weighted by the average of the inverse effect size variances for teacher outcomes and featured study characteristics as covariates, including whether effect sizes were adjusted for covariates and whether interventions focused on mathematics or science.⁴

As in other studies that have examined the link between intervention impacts on teacher-level and student-level outcomes (e.g., Egert et al. 2018; Kraft et al. 2018), the data we compiled from primary studies allow us to examine whether interventions that supported teacher knowledge and classroom instruction also tended to have larger impacts on student achievement. However, this analysis does not allow us to determine whether this association is causal in nature – that is, that improvements in teacher knowledge and instruction caused an improvement in student achievement – or because such interventions impacted other, unobserved mediators that supported student achievement and are correlated with knowledge or practice.

Results

⁴ As in our models examining impacts on teacher outcomes, we used the study-level average value of covariates in cases where there was within-study variability in the values of covariates.

Study Characteristics

In Table 1, we present descriptive information about the PD interventions and studies included in the meta-analytic sample. The sample included 46 studies with 57 treatment-control contrasts, yielding a total of 200 teacher outcomes effect sizes. Among these studies, 25 (54 percent) featured both PD and new curriculum materials; 21 studies (46 percent) featured PD only. A majority focused on mathematics (28 studies; 61 percent) and roughly one third focused on science (16 studies; 35 percent); two studies focused on both mathematics and science (4 percent). Studies included a mix of grade levels, ranging from preschool through high school. Interventions included an average of 57 PD contact hours; in a majority of studies, intervention activities took place over two or more semesters (32 studies; 74 percent). All studies were randomized controlled trials.

Table 1 also shows the foci of included interventions. A majority of studies included professional development focused on at least one aspect of teacher knowledge, including improving teacher content knowledge or pedagogical content knowledge, or knowledge of how students learn (32 studies; 70 percent). The majority of studies also focused on at least one aspect of classroom instruction. Nearly all focused on content-specific instructional strategies (45 studies; 98 percent); therefore, we were unable to examine this feature as a moderator. A smaller proportion of studies included a focus on content-specific formative assessment (8 studies; 17 percent) and on generic instructional strategies (7 studies; 15 percent).

Table 1 also describes the contexts in which studies were conducted. On average, 62 percent of students were identified as free- or reduced-price lunch eligible or low-income, 16 percent were emergent bilingual students/students learning English as a new language, and 60 percent were students of color, among studies that reported this information. We did not screen

studies based on the country in which they were conducted. However, 45 of the 46 studies that met the review's inclusion criteria were conducted in the U.S; one study was conducted in the Philippines (San Antonio et al., 2011).

The studies in our sample reported on a mix of teacher knowledge and classroom instruction outcomes. As shown in Table 1, 55 effect sizes (28 percent) captured impacts on some aspect of teacher knowledge in science or mathematics. The other 145 effect sizes (73 percent) captured impacts on classroom instruction. These included impacts on both observational and self-report measures of classroom instruction. Most effect sizes were based on intervenor-developed measures (150 effect sizes; 75 percent), although some were based on standardized (41 effect sizes; 21 percent) or other (9 effect sizes; 5 percent) measures.

Overall Average Impacts on Teacher Outcomes

Table 2 presents the results of estimating unconditional CHE models examining study impacts on teacher outcomes. Pooling across studies, the average weighted impact on all teacher outcomes was $0.52 SD$ ($p < .001$). The prediction interval based on the estimated between-study variance in impacts indicates that true effect sizes would be expected to range from $0.18 SD$ to $0.87 SD$, leading to questions about the factors that may explain the observed variability.

When we considered teacher knowledge and classroom instruction separately, we found studies yielded positive, similarly sized mean impacts on each group of outcomes. We found an average weighted impact on teacher knowledge of $0.52 SD$ ($p < .001$) among the 22 studies that reported this information. Among the 36 studies with information on classroom instruction, the average weighted impact on these outcomes was $0.49 SD$ ($p < .001$). Prediction intervals also indicated substantial between-study heterogeneity in impacts.

Our primary analysis also included evaluations of both mathematics and science

interventions. As an exploratory analysis, we also examined overall impacts on teacher outcomes for mathematics and science studies separately, and found impacts were generally similar for studies in both groups (Table S1; online only). Among studies focused on interventions in mathematics, we found a mean pooled effect size of 0.51 *SD* ($p < .001$) on all teacher outcomes, 0.44 *SD* ($p < .001$) on teacher knowledge, and 0.52 *SD* ($p < .001$) on classroom instruction. Among science-focused studies, the mean weighted impact estimate was 0.55 *SD* ($p < .001$) on all teacher outcomes, 0.66 *SD* ($p < .05$) on teacher knowledge, and 0.47 *SD* ($p < .01$) on classroom instruction.

Intervention Characteristics and Contextual Factors that Moderate Program Impacts on Teacher Outcomes

Next, we turn to programmatic and contextual factors that moderated impacts on teacher outcomes. We first examined *intervention features*, including whether the program combined professional development with new curriculum materials (intervention type) and teachers' time investments (PD duration). We did not find a significant difference in impacts on teacher outcomes between studies that provided both PD and new curriculum materials, as compared to PD only, regardless of whether we considered all teacher outcomes combined or knowledge and classroom instruction separately (see Table 3). We did not observe a statistically significant association between the number of PD contact hours and teacher outcomes, nor did we observe a significant association when we replaced the continuous measure of PD contact hours with an indicator for whether the number of PD contact hours was above the sample median.

Second, we examined the association between *PD focus* and teacher-level outcomes (see Table 4). Studies that included a focus on improving teachers' knowledge had somewhat larger mean impacts on combined teacher-level outcomes, and specifically on classroom instruction

outcomes, as compared with interventions that lacked this focus (a difference of 0.18 SD, $p < .10$, for combined outcomes, and of 0.22 SD; $p < .10$, for classroom instruction outcomes). We also found that interventions that included a focus on content-specific formative assessment had larger impacts on combined teacher-level outcomes, and specifically on classroom instruction outcomes, relative to interventions that lacked this focus (a difference of 0.19 SD, $p < .10$, for combined outcomes, and of 0.27 SD; $p < .10$, for classroom instruction outcomes).

Programmatic focus on generic instructional strategies did not significantly moderate impacts on any of the examined categories of teacher-level outcomes. None of the programmatic foci we examined significantly moderated impacts on teacher knowledge outcomes; however, caution is warranted in interpretation of the knowledge outcome models as both the study sample size and variation in the incidence of the moderators were limited ($k = 22$ studies; of which 82% included a knowledge focus; 14% included a focus on content-specific formative assessment; and 23% included a focus on generic instructional strategies).

Finally, we examined whether study contexts moderated intervention impacts on teacher outcomes. We did not observe significant relationships between intervention impacts and any of the measured contextual features, including teacher sample size, demographic and socioeconomic characteristics of the student sample, whether studies were implemented in urban as compared to rural or suburban districts, and grade level (see Tables S2-S4; online only).

Linking Impacts on Teacher and Student Outcomes

Table 5 presents the results of estimating weighted regressions that tested the associations between intervention impacts on overall teacher-level outcomes, and on teacher knowledge and classroom instruction outcomes, and improvements in student achievement. When we pooled all teacher-level outcomes (model 1), we observed a positive, statistically significant association

between treatment-level mean impacts on teacher outcomes and treatment-level mean impacts on student achievement outcomes. Specifically, we found that a 1 *SD* increase in teacher-level outcomes is associated with a 0.18 *SD* improvement in student achievement (see Table 5).

We next fit separate models to consider relationships between interventions' impacts on teacher knowledge (model 2) and classroom instruction (model 3), respectively, and impacts on student achievement. As shown in model 3, a 1 *SD* improvement in classroom instruction was associated with a 0.24 *SD* increase in student achievement, and this relationship was significant. This positive association suggests that interventions that improved classroom instruction also tended to promote student achievement. However, we note that this is a correlational association, and not a causal link. This positive association may be observed because improvements in teacher practice led to improvements in student achievement; alternatively, interventions that supported instruction may have also affected other, unobserved mediators that promoted student achievement. Meanwhile, the association between mean impacts on teacher knowledge and mean impacts on student achievement was also positive in sign, albeit smaller in magnitude and not statistically significant (with a 1 *SD* increase in teacher knowledge associated with a non-significant 0.08 *SD* increase in student achievement; see model 2).

Publication Bias

In all systematic reviews there exists the possibility of publication bias among available studies. We took three approaches to examine this issue in the present sample of studies. Following conventional practices in meta-analysis we first inspected funnel plots (showing effect sizes against their standard errors) to explore whether there is visual evidence of publication bias. The rationale for this approach is that smaller studies (with larger standard errors) have less precision, therefore are less likely to yield statistically significant results, and therefore are more

likely to be affected by publication bias. Asymmetrical patterns in the funnel plots indicate potential publication bias. We examined three funnel plots, including plots for all teacher outcomes, teacher knowledge only, and classroom instruction only. We observed asymmetry in all three funnel plots, providing visual evidence of possible publication bias (see Figures S1 to S3; online only). This pattern appears somewhat more pronounced for teacher knowledge relative to classroom instruction.

Then, we conducted statistical tests for publication bias using Egger's regression test. This method is widely used in meta-analysis to assess potential publication bias by statistically testing for asymmetry in the funnel plot. This method performs a linear regression of the standardized effect sizes on their standard errors, weighted by precision (the inverse of the effect size variance), testing the null hypothesis of a zero intercept (i.e., that there is not publication bias). A rejection of the null hypothesis indicates evidence of publication bias (e.g., Egger et al., 1997). Given the structure of our data which includes multiple effect sizes per study, we performed two versions of Egger's test. First, we aggregated effect sizes and effect size standard errors to the study level by calculating the average effect size and average effect size standard error across all effect sizes in each study. We then regressed the standard normal deviation, i.e., the effect size divided by its standard error, on effect size standard error's inverse. Second, we used a modified approach that regresses the effect sizes on their standard errors, weighted by precision, which yields equivalent results as the traditional Egger's test (Rothstein et al., 2005), but allows us to retain multiple effect sizes per study. Specifically, we added the effect size standard error as a moderator to the unconditional CHE model. If the standard errors of the effect sizes predict the magnitude of the effect sizes, this similarly indicates evidence of publication bias. We conducted both tests for all teacher outcomes, teacher knowledge only, and classroom

instruction only. Results are consistent with the presence of publication bias, and suggest that this may be more pronounced among studies that examined teacher knowledge (see Tables S5 and S6; online only).

Finally, we compared mean impacts on knowledge and instructional outcomes for peer-reviewed versus non-peer-reviewed studies (see Table S7; online only). For this test, we estimated conditional CHE models that included an indicator for whether the study was peer-reviewed (including IES-approved publications and NCEE reports) as a moderator. We did not observe a statistically significant difference in mean impacts between peer-reviewed and non-peer reviewed studies. However, the direction of results suggests that peer-reviewed studies had more positive impacts, consistent with the results described above suggesting the presence of publication bias in our sample.

Results of the three approaches described above are not entirely consistent but suggest that there may be publication bias in our sample. Findings are consistent with a recent meta-analysis examining the impact of professional development on instructional practice which also found evidence of publication bias (Garrett et al., 2019). This points to the importance of searching the grey literature when capturing research in this domain and, as a field, encouraging the publication of null findings.

Sensitivity Checks

Overall Average Impacts on Teacher Outcomes

We conducted several sensitivity checks to test the robustness of our findings to different sample and model specifications. First, we examined whether study quality issues could bias the results. As we include only RCTs, the main threat is attrition. Teachers frequently exit studies due to moving between schools, grade levels, and other reasons—particularly for multi-year

studies which comprise a substantial proportion of our sample. A high degree of attrition and/or differential attrition between treatment and control groups could lead to bias in estimates of intervention impacts, even in the context of an RCT. We coded for attrition in our sample in two ways: overall attrition at the cluster *or* student level (defined as attrition of 20% or more) and differential attrition between treatment and control groups (defined as differential attrition of 10% or more).⁵ We estimated our unconditional CHE model after restricting the sample in two ways: excluding studies with high overall attrition or high differential attrition, and excluding studies with high overall attrition *and* high differential attrition. For both, results were similar to our main results with the full sample (see Table S8; online only).

Additionally, effect sizes for impacts on classroom instruction represented a combination of self-report and observational measures of instructional practice. To determine whether this mix of self-report and observational outcomes could influence our estimate of the overall average impact on classroom instruction, we first replicated our unconditional CHE model after restricting the sample first to effect sizes based on self-report measures of instructional practice, and then to effect sizes based on observational measures of classroom instruction. Results indicate that the overall average impacts were somewhat larger for observational compared to self-reported measures of instructional practice (see Table S9; online only).

Effect sizes also represent a mix of standardized, intervenor-developed, and other types of outcome measures. Therefore, we also replicated our unconditional model after restricting the

⁵ We note that the differential attrition value (10%) was selected because it was approximately the cutoff in the WWC 3.0 manual, which was the manual in place when we established the coding protocol and commenced coding. Specifically, it was approximately the cutoff for cases in which “attrition is expected to result in an acceptable level of bias” under optimistic assumptions for the studies with a level of overall attrition we expected to see in the sample of PK-12 studies (i.e., overall attrition rates from 0 to 20%). The overall attrition threshold we applied (20%) was the maximum value for a study to retain the designation “low attrition” under our applied threshold level of differential attrition. Thus, a study that met both of our criteria would be expected to meet WWC standards for attrition based on the manual in place when we conducted coding.

sample to effect sizes based on standardized or intervenor-developed outcome measures, and after restricting the sample to effect sizes based on intervenor-developed outcome measures. We could not examine impacts on standardized or other outcome measures separately due to the small number of effect sizes in each category. Results indicate that excluding these different groups of outcome measures did not substantially affect the magnitude of estimated impacts on classroom instruction or teacher knowledge (see Table S10; online only).

Linking Impacts on Teacher and Student Outcomes

We also examined the sensitivity of our findings regarding the links between impacts on teacher and student outcomes to model specification. We confirmed that we received similar results when we examined unweighted associations between intervention impacts on teacher and student outcomes (see Table S11; online only), and when we examined weighted associations between impacts on teacher and student outcomes that used mean effect sizes aggregated to the study level rather than to the treatment-contrast level (see Table S12; online only).

Discussion

In this study, we meta-analyzed experimental research on the impacts of mathematics and science teacher professional development programs on teachers' outcomes, including knowledge and classroom instruction; examined the extent to which programmatic and contextual moderators explained variation in programs' impacts on teacher-level outcomes; and probed the extent to which improvements to teacher outcomes ultimately predicted improvements in students' mathematics and science achievement. This analysis represents a fuller test of recent policy initiatives' logic model than has occurred to date.

To summarize the primary findings, in-service mathematics and science teacher PD interventions of the types examined in the experimental literature base had mean positive

impacts on teacher outcomes, both overall and disaggregated by measures of knowledge and classroom instruction. Moderator analyses suggested that PD programs with a focus on improving teachers' knowledge, and those that included a focus on content-specific formative assessment, had stronger mean impacts on pooled teacher-level outcomes and on classroom instruction outcomes relative to programs that lacked these foci. Improvements in teacher-level outcomes, in turn, were associated with improvements in student achievement. We discuss the findings in more detail below.

Overall Impacts on Teacher Outcomes

As estimated from the available random assignment research, mathematics and science PD interventions that measured both teacher and student outcomes had mean positive impacts on teacher outcomes of 0.52 *SD*.

This included positive mean effects on measures of teacher knowledge, a key component of recent mathematics and science education policy initiatives' logic model (e.g., NRC, 2007; 2013). The average weighted impact on teacher knowledge was 0.52 *SD*. Using a calculation converting standard deviations to percentile ranks, following the approach described in Lipsey et al. (2012) to translate effect sizes to more readily interpretable forms, an average teacher who was in the treatment group would be expected to rank approximately 20 percentile points higher than an average teacher in the control group on mean indicators of teacher knowledge.

Interventions also had positive mean effects on measures of classroom instruction, with a pooled mean weighted impact estimate of 0.49 *SD*. Expressed in terms of percentiles, an average teacher in the treatment group would be expected to rank approximately 19 percentile points higher on measures of classroom instruction than an average teacher in the control group. The magnitudes of the pooled impacts on instruction are in the same range as those detected in recent

syntheses of the impacts of instructional interventions on observer-rated teaching practice (e.g., 0.42 *SD* in Garrett et al. 2019; 0.49 *SD* in Kraft et al. 2018). Overall, the magnitudes of these estimates imply that PD interventions of the types evaluated cause notable improvements to a range of teacher-level outcomes.

Associations of Program Features and Foci with Teacher Outcomes

Programs have finite time to spend with teachers, meaning that evidence regarding the specific program features and foci associated with impacts on teacher outcomes can provide insights into potentially valuable facets of program design. We thus sought to quantitatively test potential mechanisms that may explain heterogeneity in programs' efficacy. Unlike prior reviews, we were able to examine potential moderators of PD programs' impacts on teacher outcomes, including knowledge and classroom instruction, specifically for in-service mathematics and science teachers. We point out that caution is necessary in the interpretation of moderator tests due to the power constraints inherent in the small sample sizes available for moderator analyses. Additionally, like all meta-analytic moderator analyses, the moderator analyses presented in the manuscript are inherently correlational, not causal, in nature (e.g., Borenstein et al., 2021). As such, it is always possible other variables may explain observed patterns of findings. Thus meta-analytic moderator analyses serve to identify potentially noteworthy trends across studies that may warrant follow-up.

With respect to impacts on classroom instruction, moderator analyses suggested that interventions that included a focus on strengthening teacher knowledge tended to have stronger impacts on pooled teacher-level outcomes, and specifically on classroom instruction outcomes, on average, as compared to interventions that lacked this focus. These differences were marginally significant ($p < .10$). As nearly all of the interventions in the sample also included a

focus on strengthening content-specific instructional practices, it may be the case that this dual focus on knowledge-building and content-specific instructional practices is especially supportive of teachers' pedagogical skill growth, as knowledge gains gleaned from the content-deepening portion of the PD feed into practice development (e.g., Clarke & Hollingsworth, 2002), although we do not find consistent evidence for this in the teacher knowledge models. Alternatively, programs with this combined focus may simply be stronger programs taking a more comprehensive approach to teacher learning, and hence more efficacious at improving practice. However, the moderator analysis does suggest a positive role for a PD focus on content knowledge as a mechanism linked to better than typical impacts on instruction. This points to the value of future research parsing the level and type of content knowledge focus in PD that may be especially generative to teachers' practice development.

The moderator analyses also indicated that interventions' impacts on teacher-level outcomes, and specifically on classroom instruction outcomes, tended to be larger among programs that included a focus on content-specific formative assessment, as compared with programs that lacked this feature. While not causal, it is plausible that the inclusion of an explicit focus on instructional adjustments based on evidence of students' learning needs, which is characteristic of formative assessment, could potentially have benefitted teachers' instructional decision-making and demonstrated classroom practices (Palm et al., 2017). Again, given the number of interventions that included this focus was small, this finding is not definitive but suggestive of a relationship worth noting, and worthy of future follow-up.

Turning to teacher knowledge, we found that effect size magnitudes were not significantly predicted by the key intervention features that we analyzed, including intervention type, duration of the PD, or PD programmatic focus. Rather, it appears to be the case that a

variety of kinds of PD programs can effectively increase facets of teachers' knowledge, including those that operate through less direct channels than an explicit knowledge development focus. Once more primary experimental studies on this topic are conducted, attempting to parse at a finer grain size how a PD focus on deepening specific facets of teachers' knowledge may foster changes in classrooms and student learning would be a useful next step for future syntheses.

Meanwhile, the variables that we examined relating to intervention type and dosage were not significantly related to effect size magnitudes for any of the teacher-level outcomes we examined. An absence of a statistically significant association between PD duration and teacher outcomes, which was retained after parsing the data for potential nonlinear relationships, hearkens to the findings of Kennedy (2016) and Lynch et al. (2019), which did not find a clear link between longer-duration PD experiences and student learning, as well as Garrett et al. (2019), who did not find a significant relationship between classroom observation indicators and the duration of teacher PD in studies pooled across different content areas. Consistent with conclusions drawn in Kennedy (2016) and Lynch et al. (2019), PD content and quality appeared to matter more than contact hours between teachers and PD leaders. It is also possible that, as Garrett et al. (2019) suggested, shorter-duration PD programs tend to focus on more discrete teacher skills, which are easier to improve than more complex pedagogies, but can nonetheless be valuable for incrementally strengthening teachers' instructional expertise (e.g., Cai et al., 2017). We also did not observe a significant relationship between PD programs' inclusion of a focus on new curriculum materials and teacher outcomes. This is perhaps surprising, given that curriculum materials are generally hypothesized to be educative for teachers (e.g., Davis et al., 2017), and Lynch et al. (2019) found a significant relationship between teacher PD's inclusion of

a focus on curriculum materials and student achievement outcomes. Lack of significant associations could be due to power constraints in the moderator analysis. Speculatively, it is also conceivable that students may be learning more in interventions that include new curriculum materials directly as a result of their own interactions with the new curricular content, in ways not captured by teacher outcome measures, or that curriculum-focused interventions are bolstering facets of teachers' instruction not typically captured in knowledge and practice assessments, like amount of classroom time dedicated to the content.

We note that although the current review quantitatively modeled a set of meta-analytic moderators derived from the literature, we could not test everything. Like most meta-analyses, we originally coded for and hoped to analyze additional moderator variables beyond those we were able to include in the analysis (e.g., whether the PD focused on integrating technology into the classroom; contextual factors such as the share of students receiving special educational services); however, not all variables could be analyzed because the included studies contained too little variation on these variables (i.e., coded feature present in fewer than 10% or more than 90% of studies or codes collinear with other codes). The research base also points toward the value of a range of further intervention features, such as an explicit PD focus on building on students' home and community funds of knowledge (e.g., McWayne et al., 2020). Meta-analytic testing of these additional features would be a generative step for future research once a larger pool of primary studies has been conducted.

We also found that program impacts did not differ significantly based on the contextual moderators we examined, including intervention scale, setting type, and characteristics of the student sample. Most of the PD studies identified were conducted in relatively high-poverty settings, as evidenced by the fact that roughly three-quarters of the studies with non-missing data

on this variable were conducted in settings in which a majority of students were low income or eligible for a free or reduced-price school lunch. The observation that PD interventions did not appear to be less effective in high-poverty settings is encouraging, given that high-poverty schools facing resource constraints are often those that PD interventions most aim to support (e.g., Ladson-Billings, 2005). We note that although the studies in our sample included students in grade levels ranging from prekindergarten to secondary school, we did not find evidence that intervention impacts on teacher outcomes differed by grade level. This is consistent with findings from prior meta-analyses examining the impacts of mathematics and science instructional interventions on student achievement, which generally yield similar patterns of positive impacts for students in elementary and secondary grades (e.g., Lynch et al. 2019; Slavin et al., 2009).

Links to Student Achievement

Bearing in mind that our meta-analytic data and design do not support causal inferences about the impacts of teachers' outcomes on student achievement, the data nonetheless allowed us to estimate the magnitude of relationships between interventions' impacts on proximal teacher outcomes and distal student learning. Pooling across outcome types, we found that PD interventions that had stronger impacts on teacher-level outcomes tended to have stronger impacts on student learning outcomes. Disaggregating by teacher outcome category, we found supportive evidence consistent with the notion that PD interventions that had stronger mean causal impacts on instruction tended to have significantly stronger impacts on students' mathematics and science achievement. On average, a 1 *SD* improvement in classroom instruction predicted a positive 0.24 *SD* difference in student test scores – the equivalent of an improvement in student achievement from the 50th to the 59th percentile. The association between improved teacher knowledge and improved student outcomes was also positive in sign, albeit not

statistically significant.

The particular pattern of findings we observed regarding the relationships between student impacts, impacts on instruction, and impacts on teacher knowledge may have occurred for one of several reasons. On the one hand, while both teacher knowledge improvements and classroom instruction improvements were associated with improvements in student achievement in the disaggregated models, the association was larger in magnitude for classroom instruction improvements. We note that perhaps interestingly, some of the included primary studies aimed at improving content knowledge posited that strengthening knowledge alone was not sufficient to improve student outcomes (e.g., Garet et al., 2010). For example, Garet et al. (2010, 2016) conducted two large-scale federally funded RCTs examining the impacts of PD interventions whose major aim was to improve teachers' mathematics content knowledge. Despite improving teacher content knowledge and some aspects of instructional quality, these studies returned null or negative results on student achievement (see also, e.g., Dash et al., 2012). However, this analysis is not causal. Observed differences in the relationships between knowledge and classroom instruction impacts and student achievement impacts could be due to other features of instructional programs that measured knowledge impacts that made them less strong programs, and hence less effective.

As a practical matter, we note that many contemporary mathematics and science PD programs reject an "either/or" approach to improving knowledge and classroom instruction, and instead emphasize both of these levers, providing teachers opportunities to deepen their learning of the content to be taught while also attending to pedagogical practices that teachers will use to portray the content. For instance, when teachers learn to use new curriculum materials, they may also be learning the specific subject-matter knowledge embedded in those materials (Remillard

& Kim, 2017). However, how, to what degree, and under what conditions teacher PD programs should focus teachers' attention on subject matter knowledge compared with centering instructional practices remains an important unsettled question in mathematics and science education research, which warrants further follow-up. Nevertheless, the observed pattern of findings provides support for the notion that among the kinds of PD interventions that have been rigorously evaluated using RCTs, there are positive links between improvements in teacher outcomes and improvements in student outcomes, as hypothesized in the PD logic model.

Strengths, Limitations, and Future Research Directions

Our findings move the field forward by providing empirical information about whether facets of the logic model implied in recent policy investments into PD interventions in mathematics and science are supported by the research available in the experimental evidence base. First, we show that PD interventions to improve mathematics and science instruction of the types evaluated in the contemporary experimental literature have, on average, been successful in improving teacher outcomes. In contrast to earlier reviews that examined only classroom instruction (e.g., Egert et al., 2018; Kraft et al., 2018), we also document that PD interventions of the types examined in RCTs also tend to improve the knowledge of in-service mathematics and science teachers. Third, in contrast to prior reviews that considered impacts on student outcomes (e.g., Lynch et al., 2019) or teacher outcomes (e.g., Garrett et al., 2019) alone, we provide empirical evidence that connects the dots between how instructional interventions influence teachers and how they influence students.

The current findings, and evidence gaps identified in the literature, point toward useful pathways for future research. First, as is usually true in meta-analyses, which rely on study report information, we faced the challenge of missing data. Empirical studies often reported impacts on

teachers' knowledge or classroom instruction, but not both. We urge future research studies to report both types of outcomes, to enable future synthesisists to empirically test hypothesized logic models of instructional interventions with larger data pools. Furthermore, access to the underlying individual-level data from the primary studies that permits testing of more of the additional paths represented in the Figure 1 logic model would be useful to the field, and we urge future primary study authors to make these data available. Additionally, information was often missing from primary studies on teachers' personal beliefs, attitudes, and values, precluding us from analyzing how these factors related to interventions' impacts despite their theoretical importance for shaping teachers' implementation of innovations (Pajares, 1992). We urge primary study authors to collect and report information on these factors as they are an important topic for future research syntheses.

Inconsistent reporting on study methodological details also suggested areas for future improvement. We urge future primary study authors to consult the most recent version of the WWC handbook, and report information on baseline equivalence of groups as needed, following its guidance. We further urge more detailed reporting on study attrition levels, in line with the reporting recommendations described in the WWC handbook, as this would permit future analysts to conduct finer-grained analyses of differences in study findings by attrition information. Another form of missing data pertained to the kinds of interventions that were studied. The kinds of PD interventions examined in randomized experimental studies are likely to be more time-intensive and resource-intensive compared with practices that are likely to be typically occurring in schools. As such, the current observed positive findings do not imply that schools' or districts' typical PD investments are paying off – rather, they indicate that PD interventions of the types examined in randomized studies and that met our relevance criteria

were effective at bolstering teacher-level outcomes, on average. We concur with arguments for more rigorous studies of interventions that are similar to common practices in school districts, as such studies would shed more light on the outcomes of investments of the kind districts typically make into teacher PD (e.g., Hill, 2004). Additionally, nearly all studies were conducted in the United States. Given the degree of variability in instructional systems and settings across country contexts, more future experimental research on the efficacy of instructional interventions in non-U.S. contexts would deepen the field's understanding of how interventions' efficacy may vary across settings.

Finally, other constructs of theoretical interest, such as the level of school leadership support allocated to the intervention, the degree of alignment between each intervention and the existing curriculum in study schools, and the financial and material supports expended (e.g., Penuel et al., 2010; Scher & O'Reilly, 2009; Wilson, 2013), could not be modeled due to minimal information reported in the primary studies. Reporting consistent data on students' baseline knowledge and attitudes towards mathematics and science could also aid future research in understanding potential variations in interventions' impacts, for example by identifying programmatic features that may be especially beneficial in classrooms serving students who began the intervention with less exposure to or knowledge of the content area (e.g., McCabe et al., 2020; Warne et al., 2019). Building a set of standardized reporting recommendations for primary experimental studies on professional development interventions that cover these substantive issues would be a productive step for the field, to enrich the ability of future syntheses to analyze how contextual factors relate to the efficacy of educational interventions (Hill et al., 2025; Lynch, 2024).

In sum, our findings indicate that mathematics and science teacher PD interventions of

the types evaluated in the experimental literature show promise for strengthening outcomes for teachers, and these improvements are in turn linked to stronger student achievement. Given the projected expansions in the PD market in the U.S. over the next five years (Technavio, 2022), incrementally advancing the field's understanding of how and under what conditions PD interventions in mathematics and science yield stronger teacher and student impacts can help districts and schools improve their odds of seeing meaningful changes in students' ultimate learning and attainment.

References

* Reference included in the meta-analytic sample

** Reference provided supplemental information for a reference in the meta-analytic sample

Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1105–1149). New Jersey: Laurence Erlbaum Associates.

Akçay, B. (2007). *The influence of the history of science course on pre-service science teachers' understanding of the nature of science concepts*. ProQuest Dissertations & Theses.

American Association for the Advancement of Science. (1990). *Science for all Americans*. New York, NY: Oxford University Press.

Atit, K., Power, J. R., Pigott, T., Lee, J., Geer, E. A., Uttal, D. H., ... & Sorby, S. A. (2022). Examining the relations between spatial skills and mathematical performance: A meta-analysis. *Psychonomic Bulletin & Review*, 1-22.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform?. *Educational Researcher*, 25(9), 6-14. <https://doi.org/10.3102/0013189X025009006>

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. *Teaching as the learning profession: Handbook of policy and practice*, 1, 3-22.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407.

Banilower, E. R., Smith, P. S., Weiss, I. R., & Pasley, J. D. (2006). The status of K-12 science teaching in the United States. *The impact of state and national standards on K-12 science*

teaching, 83-122.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... & Tsai, Y. M. (2010).

Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180.

<https://doi.org/10.3102/0002831209345157>

Bethell, G. (2016). *Mathematics education in sub-Saharan Africa: Status, challenges, and opportunities*. World Bank, Washington, DC.

<https://openknowledge.worldbank.org/handle/10986/25289> License: CC BY 3.0 IGO

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81-90.

Blank, R. K., & De las Alas, N. (2009). *The effects of teacher professional development on gains in student achievement: How meta analysis provides scientific evidence useful to education leaders*. Council of Chief State School Officers. One Massachusetts Avenue NW Suite 700, Washington, DC 20001.

*Borman, G. D., Gamoran, A., & Bowdon, J. (2008). A randomized trial of teacher development in elementary science: First-year achievement effects. *Journal of Research on Educational Effectiveness*, 1(4), 237–264. <https://doi.org/10.1080/19345740802328273>

*Bradshaw, T. J. (2012). *Impact of inquiry-based distance learning and availability of classroom materials on physical science content knowledge of teachers and students in central Appalachia*. University of Kentucky.

Bransford, J. D., A. L. Brown, & R. R. Cocking. *How people learn*. Vol. 11. Washington, DC:

National Academy Press, 2000.

Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed, pp. 328-375). New York: MacMillan.

Brown, H., Card, D., Dickersin, K., Greenhouse, J., Kling, J., & Littell, J. (2008). *Report of the What Works Clearinghouse expert panel*. Washington, DC: National Board for Education Sciences.

*Brown, J. A., Greenfelt, D. B., Bell, E., Juarez, C. L., Myers, T., & Nayfelt, I. (2013). Early Childhood Hands-on Science Efficacy Study. ECHOS: Early Childhood Hands on Science, Spring, 1–5.

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., & Hiebert, J. (2017). Clarifying the impact of educational research on learning opportunities. *Journal for Research in Mathematics Education*, 48, 230–236.

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., ... & Wilson, C. D. (2019). The refined consensus model of pedagogical content knowledge in science education. *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, 77-94.

*Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Lof, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26(4), 499–531.

<https://doi.org/10.3102/00028312026004499>

Charalambous, C. Y., Hill, H. C., Chin, M. J., & McGinn, D. (2019). Mathematical content knowledge and knowledge for teaching: exploring their distinguishability and contribution to student learning. *Journal of Mathematics Teacher Education*, 1–35.

- Cheung, A., Slavin, R. E., Kim, E., & Lake, C. (2017). Effective secondary science programs: A best evidence synthesis. *Journal of Research in Science Teaching*, 54, 58–81.
<https://doi.org/10.1002/tea.21338>
- Chmielewski, A. K., & Reardon, S. F. (2016). Patterns of cross-national variation in the association between income and academic achievement. *AERA Open*, 2(3), 2332858416649593.
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79–122. <https://doi.org/10.3102/0034654315582065>
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18(8), 947-967.
- *Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, 45(2), 443–494. <https://doi.org/10.3102/0002831207312908>
- *Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, 42(2), 127–166. <https://doi.org/10.5951/jresematheduc.42.2.0127>
- Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12, 311–330.
- Cohen, D. K., & Hill, H. C. (2008). *Learning policy: When state education reform works*. Yale University Press.
- Conference Board of the Mathematical Sciences. (2012). *The mathematical education of teachers*

II. Providence, RI, & Washington, DC: American Mathematical Society and
Mathematical Association of America.

Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science, 24*(8), 1408-1419.

*Dash, S., Magidin de Kramer, R., O'Dwyer, L. M., Masters, J., & Russell, M. (2012). Impact of online professional development on teacher quality and student achievement in fifth grade mathematics. *Journal of Research on Technology in Education, 45*(1), 1–26.

<https://doi.org/10.1080/15391523.2012.10782595>

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3–14.

<https://doi.org/10.3102/0013189X034003003>

Davis, E. A., Palincsar, A. S., Smith, P. S., Arias, A. M., & Kademian, S. M. (2017). Educative curriculum materials: Uptake, impact, and implications for research and design.

Educational Researcher, 46(6), 293-304.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics, 132*(4), 1593-1640.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development:

Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199.

*Doabler, C. T., Clarke, B., Kosty, D. B., Baker, S. K., Smolkowski, K., & Fien, H. (2016).

Effects of a core kindergarten mathematics curriculum on the mathematics achievement of Spanish-speaking English learners. *School Psychology Review, 45*(3), 343–361.

- Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of in-service professional development programs for early childhood teachers on quality ratings and child outcomes: A meta-analysis. *Review of Educational Research, 88*(3), 401–433.
<https://doi.org/10.3102/0034654317751918>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.
<https://doi.org/10.1136/bmj.315.7109.629>
- Farver, J. A. M., Lonigan, C. J., & Eppe, S. (2009). Effective early literacy skill development for young Spanish-speaking English language learners: An experimental study of two methods. *Child Development, 80*(3), 703–719.
- Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 147–164). New York, NY: Macmillan.
- Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., & Szejnberg, L. (2008). The impact of two professional development interventions on early reading instruction and achievement (NCEE 2008-4030). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- *Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., ... & Borman, G. D. (2016). Focusing on Mathematical Knowledge: The Impact of Content-Intensive Teacher Professional Development. NCEE 2016–4010. *National Center for Education Evaluation and Regional Assistance*.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes

- professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915-945.
- *Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Walters, K., Song, M., ... & Doolittle, F. (2010). Middle school mathematics professional development impact study: Findings after the first year of implementation (NCEE 2010–4009). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- **Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., ... & Doolittle, F. (2011). Middle school mathematics professional development impact study: Findings after the second year of implementation (NCEE 2011-4024). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Garrett, R., Citkowicz, M., & Williams, R. (2019). How responsive is a teacher's classroom practice to intervention? A meta-analysis of randomized field studies. *Review of Research in Education*, 43(1), 106–137. <https://doi.org/10.3102/0091732X19830634>
- Gersten, R., Taylor, M. J., Keys, T. D., Rolhus, E., & Newman-Gonchar, R. (2014). Summary of research on the effectiveness of math professional development approaches (REL 2014–010). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>
- Godbold, J. V. (1973). *Teacher Training for Effective Questioning*.
- Goldsmith, L. T., Doerr, H. M., & Lewis, C. C. (2014). Mathematics teachers' learning: A conceptual framework and synthesis of research. *Journal of Mathematics Teacher*

Education, 17(1), 5-36.

*Gropen, J., Kook, J. F., Hoisington, C., & Clark-Chiarelli, N. (2017). Foundations of science literacy: Efficacy of a preschool professional development program in science on classroom instruction, teachers' pedagogical content knowledge, and children's observations and predictions. *Early Education and Development*, 28(5), 607–631.

<https://doi.org/10.1080/10409289.2017.1279527>

Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis*, 25(1), 1-29.

Heaton, R. (1992). Who is minding the mathematics content? A case study of a fifth-grade teacher. *Elementary School Journal*, 93(2), 153–162.

*Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, 49(3), 333–362. <https://doi.org/10.1002/tea.21004>

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., ... & Gallimore, R. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis*, 27(2), 111-132.

Hill, H. C. (2004). Professional development standards and practices in elementary school mathematics. *The Elementary School Journal*, 104, 215–231.

<https://doi.org/10.1086/499750>

Hill, H. C. (2011). The nature and effects of middle school mathematics teacher learning

- experiences. *Teachers College Record*, 113(1), 205-234.
- Hill, H. C., Litke, E., & Lynch, K. (2018). Learning lessons from instruction: Descriptive results from an observational study of urban elementary classrooms. *Teachers College Record*, 120(12), 1-46.
- Hill, H. C., Lovison, V., & Kelley-Kemple, T. (2019). Mathematics teacher and curriculum quality, 2005 and 2016. *AERA Open*, 5(4).
- Hill, H., Lynch, K., Gonzalez, K., & Pollard, C. (2019). Professional development that improves STEM outcomes. *Phi Delta Kappan*, 101(5), 50-56.
- Hill, H. C., Lynch, K., & Polanin, J. (2025). Structured reporting guidelines for classroom intervention research. *Journal of Research on Educational Effectiveness*.
10.1080/19345747.2025.2470322
- Institute of Medicine. (2010). *Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5*. The National Academies Press.
<https://doi.org/10.17226/12999>.
- Jacob, B. A., Rockoff, J. E., Taylor, E. S., Lindy, B., & Rosen, R. (2018). Teacher applicant hiring and teacher performance: Evidence from DC public schools. *Journal of Public Economics*, 166, 81-97.
- *Jacob, R., Hill, H., & Corey, D. (2017). The impact of a professional development program on teachers' mathematical knowledge for teaching, instruction, and student achievement. *Journal of Research on Educational Effectiveness*, 10(2), 379–407.
<https://doi.org/10.1080/19345747.2016.1273411>
- *Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for*

- Research in Mathematics Education*, 258–288. <https://doi.org/10.2307/30034868>
- *Jayanthi, M., Gersten, R., Taylor, M. J., Smolkowski, K., & Dimino, J. (2017). Impact of the Developing Mathematical Ideas Professional Development Program on Grade 4 Students' and Teachers' Understanding of Fractions. REL 2017–256. Regional Educational Laboratory Southeast.
- Kärkkäinen, K., & Vincent-Lancrin, S. (2013). Sparking innovation in STEM education with technology and collaboration: A case study of the HP Catalyst Initiative. *OECD Education Working Papers*, 91, OECD Publishing.
- Kennedy, M. M. (1999). Form and substance in in-service teacher education (Research Monograph No.13). Arlington, VA: National Science Foundation.
- Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, 86(4), 945–980. <https://doi.org/10.3102/0034654315626800>
- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568–589. <https://doi.org/10.3102/0002831212437853>
- Kim, J. S., & Quinn, D. M. (2012, March 2012). A meta-analysis of K–8 summer reading interventions: The role of socioeconomic status in explaining variation in treatment effects. Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., ... & Baumert, J. (2015). Content knowledge and pedagogical content knowledge in Taiwanese and German mathematics teachers. *Teaching and Teacher Education*, 46, 115–126.

<https://doi.org/10.1016/j.tate.2014.11.004>

**Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, 1(3), 155–178.

<https://doi.org/10.1080/19345740802114533>

Knapp, M. S. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research*, 67(2), 227-266.

Kowalski, S. M., Taylor, J. A., Askinas, K. M., Wang, Q., Zhang, Q., Maddix, W. P., & Tipton, E. (2020). Examining factors contributing to variation in effect size estimates of teacher outcomes from studies of science teacher professional development. *Journal of Research on Educational Effectiveness*, 13(3), 430-458.

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>

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716. <https://doi.org/10.1037/0022-0663.100.3.716>

Ladson-Billings, G. (2005). No teacher left behind: Issues of equity and teacher quality. *Measurement and Research in the Accountability Era*, 141-162.

*Lanehart, R.E., Borman, K.M., Boydston, T. L., Cotner, B. A., & Lee, R. S. (2010, March).

Improving gender, racial, and social equity in elementary science instruction and student

- achievement: The impact of a professional development program. Paper presented Society for Research on Educational Effectiveness Spring Conference, Washington, DC.
- *Lang, L. B., Schoen, R. R., LaVenita, M., & Oberlin, M. (2014). Mathematics Formative Assessment System—Common Core State Standards: A Randomized Field Trial in Kindergarten and First Grade. Society for Research on Educational Effectiveness.
- Lemov, D. (2021). *Teach like a champion 3.0: 63 techniques that put students on the path to college*. John Wiley & Sons.
- *Lewis, C., & Perry, R. (2017). Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, 48(3), 261–299. <https://doi.org/10.5951/jresmetheduc.48.3.0261>
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., ... & Busick, M. D. (2012). Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. *National Center for Special Education Research*.
- *Llosa, L., Lee, O., Jiang, F., Haas, A., O'Connor, C., Van Booven, C. D., & Kieffer, M. J. (2016). Impact of a large-scale science intervention focused on English language learners. *American Educational Research Journal*, 53(2), 395–424. <https://doi.org/10.3102/0002831216637348>
- *Lockhart-Findling, M. J. (2016). *The relationship between two types of teacher professional development and Head Start student math outcomes*. Northcentral University.
- Lynch, K. (2024). The application of behavioral economics to teacher professional development. *Educational Researcher*, 53(9), 508-520.
- Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base

- that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, 41(3), 260–293.
<https://doi.org/10.3102/0162373719849044>
- Martin, D. B. (2009). Researching race in mathematics education. *Teachers College Record*, 111(2), 295–338.
- *Martin, T., Brasiel, S. J., Turner, H., & Wise, J. C. (2012). Effects of the Connected Mathematics Project 2 (CMP2) on the Mathematics Achievement of Grade 6 Students in the Mid-Atlantic Region. Final Report. NCEE 2012–4017. National Center for Education Evaluation and Regional Assistance.
- McCabe, K. O., Lubinski, D., & Benbow, C. P. (2020). Who shines most among the brightest? A 25-year longitudinal study of elite STEM graduate students. *Journal of Personality and Social Psychology*, 119(2), 390–416
- McWayne, C. M., Mistry, J., Brenneman, K., Zan, B., & Greenfield, D. B. (2020). A model of co-construction for curriculum and professional development in Head Start: The Readiness through Integrative Science and Engineering (RISE) Approach. *Teachers College Record*, 122(11), 1-46.
- Mevarech, Z. R., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal*, 34(2), 365-394.
- *Meyers, C. V., Molefe, A., Brandt, W. C., Zhu, B., & Dhillon, S. (2016). Impact results of the eMINTS professional development validation study. *Educational Evaluation and Policy Analysis*, 38(3), 455–476. <https://doi.org/10.3102/0162373716638446>
- Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional

development approaches for academically productive discussion. *Socializing intelligence through talk and dialogue*, 347-362.

*Miller, G., Jaciw, A., Ma, B., & Wei, X. (2007). *Comparative effectiveness of Scott Foresman Science: A report of randomized experiments in five school districts*. Empirical Education.

*Mutch-Jones, Puttick, & Demers (2014, April). *Differentiating science instruction: Teacher and student findings from the Accessing Science Ideas project*. Poster presented at the American Educational Research Association Annual Meeting, Philadelphia, PA.

National Academies of Sciences, Engineering, and Medicine. (2007). *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/11463>.

National Academies of Sciences, Engineering, and Medicine. (2020). *Changing Expectations for the K-12 Teacher Workforce: Policies, Preservice Education, Professional Development, and the Workplace*. The National Academies Press. <https://doi.org/10.17226/25603>.

National Governors Association. (2007). *Innovation America: A final report*. Retrieved from www.nga.org/files/live/sites/NGA/files/pdf/0707INNOvATIONFINAL.PDF

National Research Council. (2013). *Monitoring progress toward successful K-12 STEM education: A nation advancing?*. National Academies Press.

National Research Council Committee on Prospering in the Global Economy of the 21st Century. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. National Academies Press.

National Research Council, Division of Behavioral, Social Sciences, Board on Science Education, National Committee on Science Education Standards, & Assessment. (1996).

- National science education standards*. National Academies Press.
- *Newman, D., Finney, P., Bell, S. H., Turner, H., Jaciw, A., Zacamy, J., & Gould, L. (2012). Evaluation of the effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI). *Science, and Technology Initiative (AMSTI)*(February 7, 2012).
- No Child Left Behind Act of 2001. (2001). 20 U.S.C. § 6319.
- Oakes, J. (1990). Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science.
- Odden, A., Archibald, S., Fermanich, M., & Gallagher, H. A. (2002). A cost framework for professional development. *Journal of Education Finance*, 28(1), 51-74.
- OECD (2007), *PISA 2006: Science Competencies for Tomorrow's World: Volume 1: Analysis*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/9789264040014-en>.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332.
- Palm, T., Andersson, C., Boström, E., & Vingsle, C. (2017). A review of the impact of formative assessment on student achievement in mathematics. *Nordic Studies in Mathematics Education*, 22(3), 25-50.
- Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. E. (2020). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. *American Economic Journal: Economic Policy*, 12(1), 359-88.
- Pehmer, A. K., Gröschner, A., & Seidel, T. (2015). Fostering and scaffolding student engagement in productive classroom discourse: Teachers' practice changes and reflections in light of teacher professional development. *Learning, Culture and Social Interaction*, 7, 12-27.

- Pelligrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective programs in elementary mathematics: A meta-analysis. *AERA Open*, 7(1), 1-29.
- Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, 44(4), 921-958.
- *Penuel, W. R., Gallagher, L. P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in earth systems science: A comparison of three professional development programs. *American Educational Research Journal*, 48(4), 996–1025.
<https://doi.org/10.3102/0002831211410864>
- Penuel, W. R., Riel, M., Joshi, A., Pearlman, L., Kim, C. M., & Frank, K. A. (2010). The alignment of the informal and formal organizational supports for reform: Implications for improving teaching in schools. *Educational Administration Quarterly*, 46(1), 57–95.
<https://doi.org/10.1177/1094670509353180>
- Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. *Second Handbook of Research on Mathematics Teaching and Learning*, 1, 257-315.
- *Piasta, S. B., Logan, J. A., Pelatti, C. Y., Capps, J. L., & Petrill, S. A. (2015). Professional development for early childhood educators: Efforts to improve math and science learning opportunities in early childhood classrooms. *Journal of Educational Psychology*, 107(2), 407.
- Pustejovsky, J. (2018). clubSandwich: Cluster-robust (sandwich) variance estimators with small sample corrections. R package version 0.3.0. Retrieved from
<https://github.com/jepusto/clubSandwich>
- Putnam, R. T., Heaton, R., Prawat, R. S., & Remillard, J. (1992). Teaching mathematics for

- understanding: Discussing case studies of four fifth-grade teachers. *Elementary School Journal*, 93(2), 213–228.
- Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75(2), 211-246.
- Remillard, J., & Kim, O. K. (2017). Knowledge of curriculum embedded mathematics: Exploring a critical domain of teaching. *Educational Studies in Mathematics*, 96(1), 65–81.
- *Resendez, M., and Azin, M. (2008). *A study of the effects of Pearson's 2009 enVision Math Program*. Jackson, WY: PRES Associates, Inc.
- *Rimbey, K. A. (2013). *From the common core to the classroom: A professional development efficacy study for the common core state standards for mathematics*. Arizona State University.
- Robinson, W. H., McGee, E. O., Bentley, L. C., Houston, S. L., & Botchway, P. K. (2016). Addressing negative racial and gendered experiences that discourage academic careers in engineering. *Computing in Science & Engineering*, 18(2), 29-39.
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one?. *Education Finance and Policy*, 6(1), 43-74.
- Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, 52(3), 475-514.
- Roth, K. J., Wilson, C. D., Taylor, J. A., Stuhlsatz, M. A., & Hvidsten, C. (2019). Comparing the

effects of analysis-of-practice and content-based professional development on teacher and student outcomes in science. *American Educational Research Journal*, 56(4), 1217-1253.

Rust, A. H. (2011). *The impact of instruction incorporating content area reading strategies on student mathematical achievement in a community college developmental mathematics course*. University of Maryland, College Park.

*San Antonio, D. M., Morales, N. S., & Moral, L. S. (2011). Module-based professional development for teachers: a cost-effective Philippine experiment. *Teacher Development*, 15(2), 157–169. <https://doi.org/10.1080/13664530.2011.571496>

*Santagata, R., Kersting, N., Givvin, K. B., & Stigler, J. W. (2010). Problem implementation as a lever for change: An experimental study of the effects of a professional development program on students' mathematics learning. *Journal of Research on Educational Effectiveness*, 4(1), 1–24. <https://doi.org/10.1080/19345747.2010.498562>

*Sarama, J., Clements, D. H., Starkey, P., Klein, A., & Wakeley, A. (2008). Scaling up the implementation of a pre-kindergarten mathematics curriculum: Teaching for understanding with trajectories and technologies. *Journal of Research on Educational Effectiveness*, 1(2), 89–119. <https://doi.org/10.1080/19345740801941332>

Scher, L., & O'Reilly, F. (2009). Professional development for K–12 math and science teachers: What do we really know?. *Journal of Research on Educational Effectiveness*, 2(3), 209–249. <https://doi.org/10.1080/19345740802641527>

*Schneider, M. C., & Meyer, J. P. (2012). Investigating the efficacy of a professional development program in formative classroom assessment in middle school English language arts and mathematics. *Journal of Multidisciplinary Evaluation*, 8(17), 1–24.

- Sheridan, S. M., Smith, T. E., Moorman Kim, E., Beretvas, S. N., & Park, S. (2019). A meta analysis of family-school interventions and children's social-emotional functioning: Moderators and components of efficacy. *Review of Educational Research, 89*(2), 296-332.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.
- Siegle, D., & McCoach, D. B. (2007). Increasing student mathematics self-efficacy through teacher training. *Journal of Advanced Academics, 18*(2), 278-312.
- Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Goodrich, J., ... & Anders, J. (2023). Effective teacher professional development: New theory and a meta-analytic test. *Review of Educational Research, 00346543231217480*.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453.
- Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427-515.
<https://doi.org/10.3102/0034654308317473>
- Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research, 79*(2), 839-911. <https://doi.org/10.3102/0034654308330968>
- Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2014). Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching, 51*(7), 870-901. <https://doi.org/10.1002/tea.21139>
- Smith, M. S., & O'Day, J. (1990). Systemic school reform. *Journal of Education Policy, 5*, 233-

267. <https://doi.org/10.1080/02680939008549074>

- Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolk, S., Boys, C., & Teelucksingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39*(6), 521-542.
- *Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology, 40*, 41–54.
- Stein, M. K., Baxter, J. A., & Leinhardt, G. (1990). Subject-matter knowledge and elementary instruction: A case from functions and graphing. *American Educational Research Journal, 27*(4), 639–663.
- Stockard, J., Wood, T. W., Coughlin, C., & Rasplika Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research, 88*(4), 479-507.
- Sullins, J., Craig, S. D., & Graesser, A. C. (2010). The influence of modality on deep-reasoning questions. *International Journal of Learning Technology, 5*(4), 378-387.
- *Supovitz, J., & Sirinides, P. (2018). The linking study: An experiment to strengthen teachers' engagement with data on teaching and learning. *American Journal of Education, 124*(2), 161–189. <https://doi.org/10.1016/j.cedpsych.2014.05.005>
- Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods, 5*(1), 13–30. <https://doi.org/10.1002/jrsm.1091>
- Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data

structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, 2, 85-112.

*Taylor, J. A., Getty, S. R., Kowalski, S. M., Wilson, C. D., Carlson, J., & Van Scotter, P.

(2015). An efficacy trial of research-based curriculum materials with curriculum-based professional development. *American Educational Research Journal*, 52(5), 984–1017.

<https://doi.org/10.1177/2332858418791991>

Taylor, J. A., Kowalski, S. M., Polanin, J. R., Askinas, K., Stuhlsatz, M. A., Wilson, C. D., ... & Wilson, S. J. (2018). Investigating science education effect sizes: Implications for power analyses and programmatic decisions. *AERA Open*, 4(3), 2332858418791991.

Technavio (2022). *Professional Development Market in the US Growth, Size, Trends, Analysis Report by Type, Application, Region and Segment Forecast 2022-2026*.

https://www.technavio.com/report/professional-development-market-industry-in-the-us-analysis?utm_source=prnewswire&utm_medium=pressrelease+&utm_campaign=newn1_rep1_wk06_2023_007&utm_content=IRTNTR73438&nowebp

Vembye, M. H., Weiss, F., & Hamilton Bhat, B. (2024). The effects of co-teaching and related collaborative models of instruction on student achievement: A systematic review and meta-analysis. *Review of Educational Research*, 94(3), 376-422.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. doi:10.18637/jss.v036.i03

Waheed, H. (2023). Nudging smokers away from lighting up: A meta-analysis of framing effect in current smokers. *Journal of Behavioral and Experimental Economics*, 104, 101998.

Warne, R. T., Sonnert, G., & Sadler, P. M. (2019). The relationship between Advanced Placement mathematics courses and students' STEM career interest. *Educational*

- Researcher*, 48, 101–111.
- Wayne, A. J., Song, M., Bishop, A., Graczewski, C., Kitmitto, S., & Lally, H. (2023). Evaluation of MyTeachingPartner-Secondary Delivered Using Local Coaches during the COVID-19 Pandemic: Evidence from a Randomized Experiment. *American Institutes for Research*.
- Weller, N., & Barnes, J. (2014). *Finding pathways: Mixed-method research for studying causal mechanisms*. Cambridge University Press.
- What Works Clearinghouse. (n.d.). Questions and answers from demystifying the What Works Clearinghouse: A webinar for developers and researchers.
- What Works Clearinghouse. (2022). What Works Clearinghouse standards handbook, version 5.0. Institute of Education Sciences, U.S. Department of Education.
- Wijnia, L., Noordzij, G., Arends, L. R., Rikers, R. M., & Loyens, S. M. (2024). The effects of problem-based, project-based, and case-based learning on students' motivation: A meta-analysis. *Educational Psychology Review*, 36(1), 29.
- Wilson, S. M. (2008). *California dreaming: Reforming mathematics education*. Yale University Press.
- Wilson, S. M. (2013). Professional development for science teachers. *Science*, 340, 310–313. <https://doi.org/10.1126/science.1230725>
- Wu, H.-H. (2011). *Understanding numbers in elementary school mathematics*. Providence, RI: American Mathematics Society.
- *Yang, R., Porter, A. C., Massey, C., & Merlino, F. J. (2016). *Applying cognitive science principles to a traditional middle school science curriculum: comparing two approaches to increasing student achievement*. Working Paper. Philadelphia, PA: University of Pennsylvania.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues & Answers. REL 2007-No. 033. Regional Educational Laboratory Southwest (NJ1).

Tables and Figures

Table 1

Sample Sizes and Study Characteristics

<i>Sample sizes</i>		
Total number of studies (treatment-control contrasts)	46 (57)	
Total number of teacher outcomes effect sizes	200	
<i>Study characteristics</i>	<i>Number of studies</i>	<i>Percent of studies or mean (SD)</i>
Intervention type ^a		
Professional development only	21	45.7%
Professional development + New curriculum materials	25	54.3%
Subject matter focus		
Mathematics only	28	60.9%
Science only	16	34.8%
Mathematics + Science	2	4.3%
Grade level ^b		
Preschool	7	15.2%
Kindergarten	2	4.3%
Early Elementary	8	17.4%
Upper Elementary	22	47.8%
Middle School	15	32.6%
High School	3	6.5%
Professional development hours ^c		56.5 (48.2)
Professional development timespan ^d		
One month or less	5	11.6%
One semester	6	14.0%
Two semesters	20	46.5%
More than one year	12	27.9%
Professional development focus: Teacher knowledge		
Improve content knowledge/pedagogical content knowledge or knowledge of how students learn	32	69.6%
Professional development focus: Instructional practice		
Content-specific instructional strategies	45	97.8%
Generic instructional strategies	7	15.2%
Content-specific formative assessment	8	17.4%
Professional development focus: Teacher knowledge and instructional practice together	32	69.6%
Contextual factors		
Average percent of low income or free or reduced-price lunch-eligible students		61.8 (23.6) ^e
Majority of students are low income or free or reduced-price lunch-eligible	27 ^f	73.0%
Average percent of emergent bilingual students/students		15.9 (13.8) ^g

learning English as a new language		
Average percent of students of color		59.6 (26.6) ^h
Study conducted in the US	45	97.8%
<hr/>		
	<i>Number of</i>	<i>Percent of effect</i>
<i>Effect size characteristics</i>	<i>effect sizes</i>	<i>sizes</i>
<hr/>		
Outcome type		
Teacher content knowledge or pedagogical content knowledge	55	27.5%
Classroom instruction	145	72.5%
Observational	58	29.0%
Self-reported	87	43.5%
Effect size adjusted for covariates	137	68.5%
Outcome measure type		
Standardized	41	20.5%
Intervenor-developed	150	75.0%
Other	9	4.5%

^a Studies with at least one treatment arm that provided new curriculum materials and professional development were included in “Professional development + New curriculum materials.”

^b Studies may have included multiple grade levels.

^c If professional development hours varied across treatment arms, we calculated the study average.

^d Out of 43 studies with non-missing information on this variable. If professional development timespan varied across treatment arms, we used the maximum.

^e Out of 29 studies with non-missing information on this variable.

^f Out of 37 studies with non-missing information on this variable. Studies conducted in Head Start programs were included in the “Majority of students are low income or free or reduced-price lunch-eligible” category.

^g Out of 24 studies with non-missing information on this variable.

^h Includes the average percent of students who are not white. Out of 32 studies with non-missing information on this variable.

Table 2

Results of Estimating CHE Models Examining the Impacts of PD Interventions on Pooled Teacher Outcomes, Teacher Knowledge, and Classroom Instruction

	(1) All teacher outcomes	(2) Teacher knowledge outcomes	(3) Classroom instruction outcomes
Intercept	0.521*** (0.054)	0.518*** (0.081)	0.489*** (0.061)
<i>N</i> effect sizes	200	55	145
<i>N</i> studies	46	22	36
τ^2 ^a	0.031	0.025	0.029
.95 prediction interval ^b	[0.176, 0.866]	[0.209, 0.828]	[0.155, 0.823]

Notes: Standard errors in parentheses. All models were estimated using the *metafor* (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky, 2018) packages in R.

+ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$.

^b The prediction interval was calculated as: $\hat{\mu} \pm 1.96 * \hat{\tau}$, where $\hat{\mu}$ is the estimated average effect size and $\hat{\tau}$ is the square root of the estimate of the between-study variance component.

Table 3

Results of Estimating CHE Models Examining the Impacts of PD Interventions on Pooled Teacher Outcomes, Teacher Knowledge, and Classroom Instruction, including Intervention Type and Dosage as Moderators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All teacher outcomes			Teacher knowledge outcomes			Classroom instruction outcomes		
Intervention type:									
Both PD + curriculum	0.008 (0.114)			-0.345 (0.234)			0.142 (0.098)		
PD contact hours ^a		0.004 (0.011)			-0.016 (0.021)			0.005 (0.012)	
PD contact hours: Above sample median (>49 hours)			-0.004 (0.114)			-0.102 (0.155)			-0.007 (0.132)
<i>N</i> effect sizes	200	195	195	55	55	55	145	140	140
<i>N</i> studies	46	44	44	22	22	22	36	34	34
Controls for study covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. All models were estimated using the *metafor* (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky, 2018) packages in R.

+ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$.

^a PD contact hours measured as PD contact hours/10.

Table 4

Results of Estimating CHE Models Examining the Impacts of PD Interventions on Pooled Teacher Outcomes, Teacher Knowledge, and Classroom Instruction, including Professional Development Foci as Moderators

	(1) All teacher outcomes	(2) Teacher knowledge outcomes	(3) Classroom instruction outcomes
PD focus: Teacher knowledge			
Improve content knowledge/PCK and/or knowledge of how students learn	0.182+ (0.093)	-0.072 (0.182)	0.223+ (0.114)
PD focus: Instructional practice			
Generic instructional strategies	-0.023 (0.147)	-0.160 (0.275)	0.126 (0.137)
Content-specific formative assessment	0.190+ (0.100)	0.120 (0.207)	0.272+ (0.111)
<i>N</i> effect sizes	200	55	145
<i>N</i> studies	46	22	36
Controls for study covariates	Yes	Yes	Yes

Notes: Standard errors in parentheses. All models were estimated using the *metafor* (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky, 2018) packages in R.

Study covariates include whether intervention focused on math or math/science (vs. science only) and whether effect sizes were adjusted for covariates. Study-level mean values of all covariates were included in the model.

+ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$.

Table 5

Results of Estimating Weighted Regression Models Examining the Associations Between Treatment Impacts on Teacher Outcomes and Student Outcomes

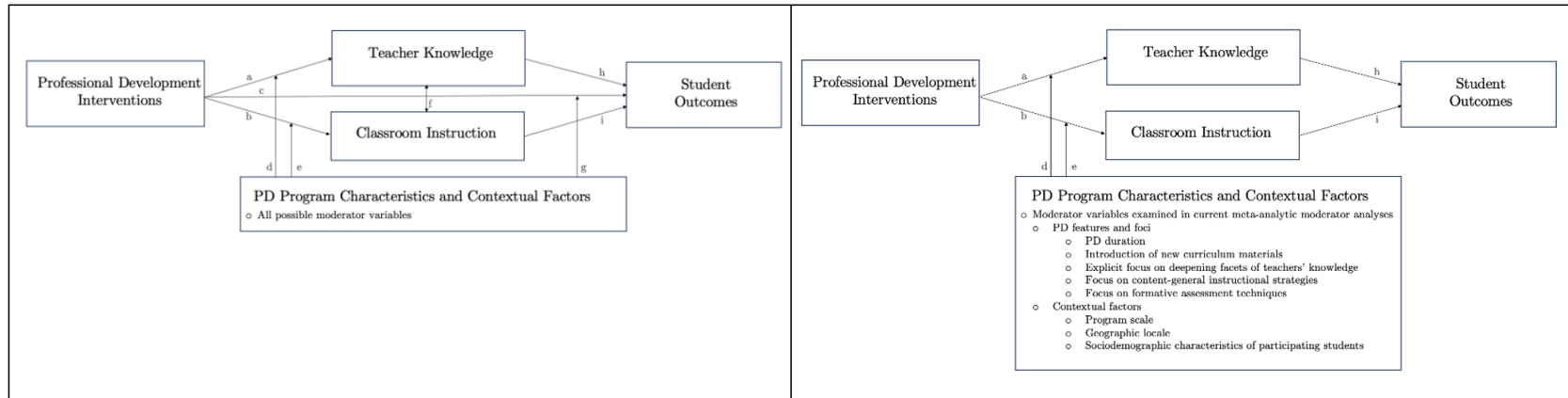
	(1)	(2)	(3)
	Student achievement		
All teacher outcomes	0.179* (0.089)		
Teacher knowledge		0.084 (0.134)	
Classroom instruction			0.236** (0.083)
Controls for study covariates	Yes	Yes	Yes
N treatment arms	57	27	44
N studies	46	22	36

Notes: Standard errors in parentheses. Table presents results of estimating regression models predicting intervention (treatment arm) mean impacts on student outcomes as a function of intervention (treatment arm) mean impacts on teacher outcomes, weighted by the average inverse effect size effect for teacher outcomes. Study covariates include whether intervention focused on math or math/science (vs. science only) and whether effect sizes were adjusted for covariates. Study-level mean values of all covariates were included in the model.

+ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$.

Figure 1

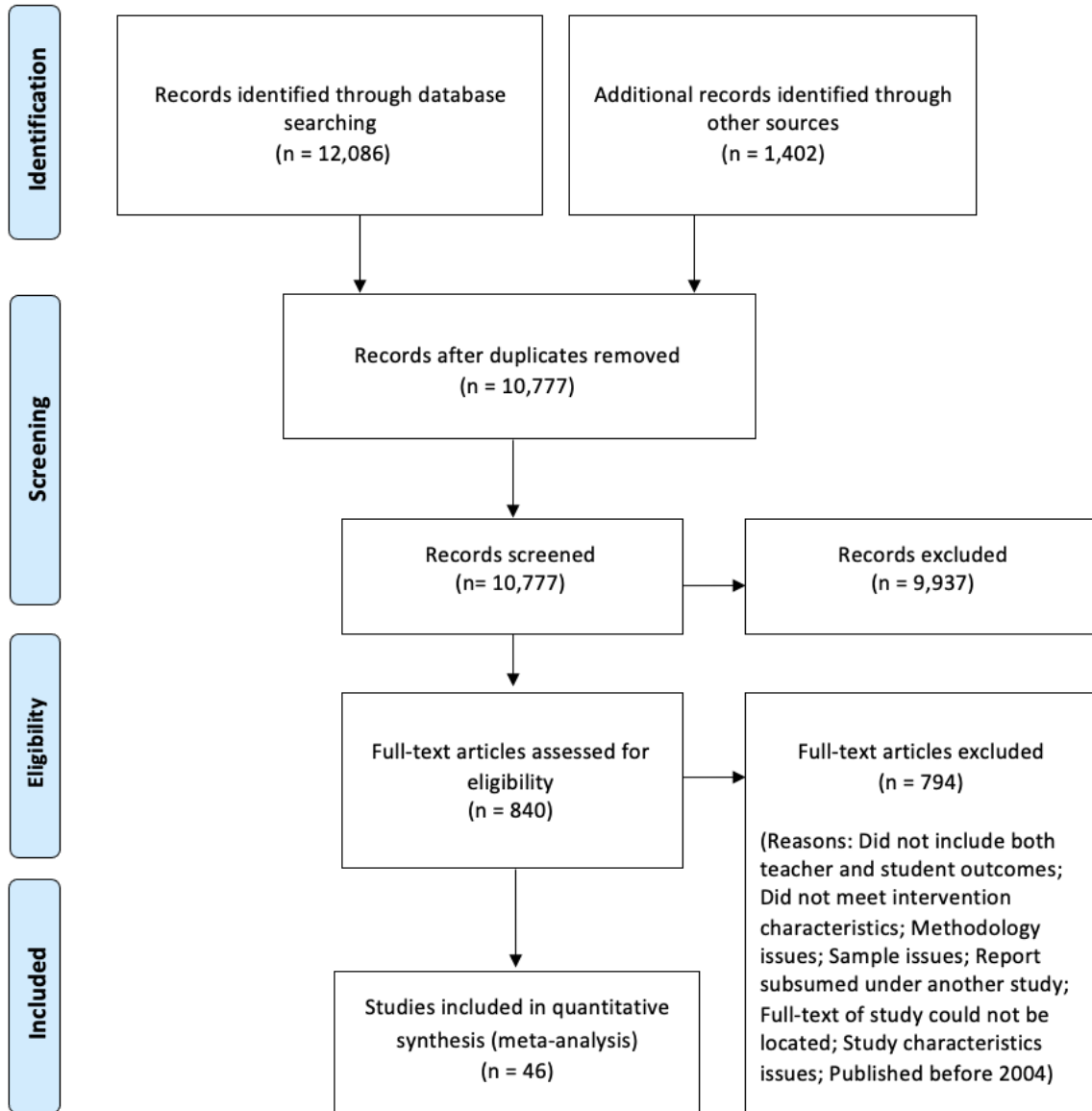
Hypothesized Logic Model for Links Between Professional Development Interventions, Teacher Outcomes, and Student Outcomes



Note: The left panel of the figure shows an illustration of potential mechanisms by which PD interventions are hypothesized to lead to changes in student learning. The right panel illustrates the subset of mechanisms examined in the current analyses. (Adapted from Garrett et al., 2019)

Figure 2

PRISMA Study Screening Flowchart



Source: Moher, Liberati, Tetzlaff, Altman, and The PRISMA Group (2009)

Note: PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Flowchart indicates whether studies were identified and screened as part of the meta-analysis conducted by Lynch et al. (2019), or as part of the supplemental search for studies published between 2016 and 2024 conducted by the study authors for the present analysis. Portions of this flowchart adapted from Figure 1 of Lynch et al. (2019).