ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-142

Strengthening STEM Instruction in Schools: Learning from Research

Kathleen Lynch
Brown University

Heather C. Hill
Harvard University

Kathryn E. Gonzalez
Harvard University

Cynthia Pollard
Harvard University

More than half of U.S. children fail to meet proficiency standards in mathematics and science in fourth grade. Teacher professional development and curriculum improvement are two of the primary levers that school leaders and policymakers use to improve children's science, technology, engineering and mathematics (STEM) learning, yet until recently, the evidence base for understanding their effectiveness was relatively thin. In recent years, a wealth of rigorous new studies using experimental designs have investigated whether and how STEM instructional improvement programs work. This article highlights contemporary research on how to improve classroom instruction and subsequent student learning in STEM. Instructional improvement programs that feature curriculum integration, teacher collaboration, content knowledge, pedagogical content knowledge, and how students learn all link to stronger student achievement outcomes. We discuss implications for policy and practice.

VERSION: October 2019

Suggested citation: Lynch, Kathleen, Heather C. Hill, Kathryn Gonzalez, and Cynthia Pollard. (2019). Strengthening STEM Instruction in Schools: Learning from Research. (EdWorkingPaper: 19-142). Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-142

Strengthening STEM Instruction in Schools:
Learning from Research

Kathleen Lynch
Brown University

Heather C. Hill, Kathryn Gonzalez, & Cynthia Pollard
Harvard Graduate School of Education

October 2019

Suggested Citation:
Lynch, K., Hill, H.C., Gonzalez, K, & Pollard, C. (2019). Strengthening STEM instruction in schools: Learning from research. *Policy Insights from the Behavioral and Brain Sciences*, *6*(2), 236–242. https://doi.org/10.1177/2372732219864385

Correspondence concerning this article can be addressed to Kathleen Lynch at kathleen.lynch@brown.edu.

This material is based upon work supported by the National Science Foundation under Grant #1348669. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Abstract

More than half of U.S. children fail to meet proficiency standards in mathematics and science in fourth grade. Teacher professional development and curriculum improvement are two of the primary levers that school leaders and policymakers use to improve children's STEM learning, yet until recently, the evidence base for understanding their effectiveness was relatively thin. In recent years, a wealth of rigorous new studies using experimental designs have investigated whether and how STEM instructional improvement programs work. This article highlights contemporary research on how to improve classroom instruction and subsequent student learning in STEM. Instructional improvement programs that feature curriculum integration, teacher collaboration, content knowledge, pedagogical content knowledge, and how students learn all link to stronger student achievement outcomes. We discuss implications for policy and practice.

Keywords: Teachers, Professional development, Curriculum, Mathematics, Science, STEM

Bulleted Highlights

1. Teacher professional development and curriculum improvement programs have, on average, positive impacts on student outcomes.
2. Instructional improvement programs that include both professional development and curriculum materials are more effective than those that include PD or curriculum alone.
3. Affording teachers opportunities to study the curriculum they will teach and to improve their own STEM content and pedagogical content knowledge benefits student learning.
4. Carving out time for teachers to collaborate with their same-school colleagues and to meet regularly to discuss program implementation is linked to better program outcomes.

Teachers learn from one another, and social motivation and the opportunity to share expertise encourage teachers to persist in implementing ambitious new practices.

5. Schools and districts should invest strategically in STEM teacher professional development that is curriculum-focused and embeds teacher collaboration. These investments tend to strengthen the skills of the STEM teacher workforce, leading to improved student learning in mathematics and science.

Strengthening STEM Instruction in Schools:

Learning from Research

U.S. students' poor performance in STEM is a widely documented problem. Sixty percent of children in the U.S. fail to meet standards for mathematical proficiency in fourth grade (NAEP, 2017), while 62 percent fail to meet benchmarks in science (NAEP, 2015). These problems are particularly acute for students from less socioeconomically advantaged backgrounds, who disproportionately attend under-resourced schools with less experienced teachers. Poor performance in mathematics and science limits children's access to advanced STEM coursework in the later grades, and subsequently may block students' access to future STEM careers.

Teacher professional development and curriculum improvement are two critical policy levers to improve children's STEM learning. As such, teacher professional development and new curriculum materials constitute major investments for districts. Districts spend between 1 and 6% of their budgets on teacher PD (see, e.g., Corcoran, 1995; Miles, Odden, Fermanich & Archibald, 2004; Miller, Lord & Dorney, 1994), and most U.S. teachers participate in professional development at least once annually (NCES, 2001). Furthermore, the market for instructional materials is estimated to be in the neighborhood of $12 billion per year (Cavanagh, 2015). The size of these investments dictates understanding when professional development and curriculum interventions are most effective. From a policy standpoint, distilling the empirical research evidence is particularly important because the *Every Student Succeeds Act* requires that districts receiving Title I funds must adopt "evidence-based interventions," including programs and strategies proven to be effective in raising student achievement.

The current article highlight the contemporary research base that bears on how to improve classroom instruction and subsequent student learning in STEM. Specifically, we survey the empirical evidence on two categories of STEM programs that intervene *in the classroom*: (1) *teacher professional development*, which aims to improve teachers' knowledge and skills, and (2) *new curriculum,* which aims to improve the richness of the content available to students. Here, without changes to classroom processes -- what is happening in the classroom each day between teachers, students, and content -- student learning is unlikely to improve.

This article is organized as follows. First, we review prior efforts to understand the impacts of teacher professional development and curriculum improvement interventions on student learning in STEM. Second, we highlight recent work examining the characteristics of STEM PD and curriculum programs that are associated with stronger student achievement outcomes. In particular, our large-scale statistical meta-analysis of the recent experimental and quasi-experimental evaluations shows these programs influence student achievement. Third, we provide recommendations for policy, research, and practice stemming from the recent research on STEM instructional improvement initiatives, with the goal of strengthening teacher practice and bolstering students' opportunities to advance in STEM.

How Teacher PD and Curriculum Programs Influence Student Learning: Theory of Action

What is the process by which teacher professional development and curriculum programs influence student learning? Informative prior work (Scher &O'Reilly, 2009) posited a framework for understanding how elements of teacher professional development influence teacher learning, teacher practices, and student outcomes. Developing logic model entailed reviewing both existing research on teacher professional development, as well as expert opinion on factors theorized to contribute to PD effectiveness as discussed in the literature and in policy documents (Scher & O'Reilly, 2009). Specific features of professional development content, context, and implementation are likely to improve student learning outcomes. First, the *resources* available for the professional development, such as duration, format (e.g. workshop, study group, workshop plus coaching), and structural characteristics, may contribute to PD outcomes. Second, the district, school, and political climate, or *contexts of environmental support*, may help or hinder the success of the PD. Third, PD activities may be more or less *coherent* in their alignment with school curricula, standards, and teachers' beliefs and practices. PD *content* may vary in terms of mathematics versus science focus, instructional strategies taught, and focus on how students learn, as well as fidelity of implementation.

So far, the Scher-O'Reily model focused on professional development alone. Because we seek to understand the role of curriculum materials, we amend their model to include characteristics of curriculum materials interventions that are hypothesized to contribute to student learning. A similar theoretical framework, developed by the National Research Council for curriculum materials, suggests that curriculum-materials evaluations should take into consideration how each component influences student outcomes (Confrey, 2006): program components (such as curriculum content and design elements), implementation components (such as teacher support and pedagogical elements), and secondary components (such as contextual and systemic factors). These curriculum elements may include the program's focus on active student learning (such as in labs), alignment with NCTM/NSTA standards, and amount of teacher professional development associated with the materials.

Professional development (and by extension, curriculum materials) is hypothesized to influence student outcomes via a three-phase process (Scher & O'Reilly, 2009). The first phase, *immediate outcomes,* includes changes in teacher content knowledge, pedagogical knowledge, and attitudes or beliefs. The second phase, *intermediate outcomes,* includes changes in teachers' instructional practices. Finally, the third phase, *long-term outcomes,* includes changes in students' attitudes and achievement. Here, we focus primarily on the research evidence on the third phase, student outcomes. We return to the issue of proximal outcomes on teachers in the discussion.

Impacts on Distal Outcomes: Student Achievement

In recent decades, increasingly sophisticated methods have honed the field's understanding of the impacts of teacher PD and curriculum programs on student learning. Prior to 2002, studies often relied on expert views and teacher self-report data to distinguish best practices in teacher PD and curriculum development (e.g., Borasi & Fonzi, 2002; Elmore, 2002; Porter, Garet, Desimone, Yoon, & Birman, 2000). Although these studies illuminate teachers' and experts' opinions, they lack causal evidence about the impacts of the recommended features on student outcomes, limiting their conclusions.

Several reviews during this period ventured to synthesize the existing empirical evidence on teacher PD and curriculum improvement programs; however, these early efforts often found that the evidence base for making recommendations was thin. Searching from 1990 through 2004, yielded only 18 studies for a meta-analysis of mathematics and science professional development programs (Scher & O'Reilly, 2009), despite relaxing methodological criteria from an initial focus on randomized experiments to the eventual inclusion of any relevant study with a comparison group. The small sample size enabled examining only a limited number of variables as potential moderators of program impact. Multiyear professional development interventions were more effective than those that transpired over a single year in mathematics, but not in science. Also, programs that focused on both content and pedagogy together were more effective on average than programs that focused on content or pedagogy alone. Another survey of the research encountered a similarly small pool of studies (Yoon et al., 2007). In a pool of studies produced through 2003, only nine studies met the What Works Clearinghouse's stringent criteria for methodological rigor, despite searching in math, science, and English Language Arts. Programs with teachers participating in 14 or more hours of professional development yielded larger average impacts on student achievement than programs offering shorter-duration PD. Yet again, the small pool of studies allowed few conclusions about the characteristics associated with effective teacher PD. Synthesists from this period frequently advocated that more research reviews should be conducted at a future time when more rigorous studies had been conducted.

However, with changes to the Institute of Education Sciences (IES) funding guidelines in 2002, researchers seeking federal IES grants were newly required both to use research methods that support causal inference and to measure the impacts of such interventions on student outcomes. In the same period, the National Science Foundation's (NSF) grantmaking also began to reflect similar interests, particularly around interventions focused on teacher learning and the use of novel curriculum materials in classrooms. Several influential reports in the early and mid-2000s also called for increased use of rigorous study designs in education research and specifically in research on teaching (Confrey & Stohl, 2004; Raudenbush, 2008; Shavelson & Towne, 2001).

Partially in response, many new studies of instructional programs are more methodologically rigorous, including many classroom- and school-level cluster randomized experiments. At the outset, our own work hypothesized that these new studies should provide a thicker evidence base than had been available in the past for understanding the effectiveness of teacher PD and curriculum programs in STEM. These studies also examined interventions using a range of new formats, such as online teacher professional development and peer coaching.

Prompted by this larger, more rigorous, and more thematically varied pool of studies, a new meta-analytic review synthesized updated research evidence on instructional improvement programs in STEM (Lynch, Hill, Gonzalez, & Pollard, 2019). First, we sought to understand how effective STEM PD and curriculum improvement programs are on average at improving student outcomes. Second, we asked whether specific program or study characteristics were associated with differential effects of STEM PD and curriculum improvement programs on student outcomes. Third, given the persistent and severe challenges of improving STEM achievement for children from less socioeconomically advantaged backgrounds, we explored whether PD and curriculum interventions are more or less effective in high-poverty settings.

The goal was to identify all relevant studies published in 1989 or later and focused on classroom-level STEM instructional improvement through professional development, curriculum materials, or both. Studies for review came from library electronic reference databases, research organizations' websites, lists of IES and NSF STEM grant awardees, and prior reviews. Including only randomized or strong quasi-experimental design yielded 95 studies that met these criteria, notably more than reviewers found two decades ago. Newly developed codes captured features of the instructional improvement programs evaluated in the studies. Statistical meta-analytic techniques estimated the average impact of STEM instructional improvement programs on student achievement in math and science. Subsequent analyses then examined whether instructional improvement programs with specific features posted larger effects on student achievement. (See the full report for details; Lynch, Hill, Gonzalez, & Pollard, 2019.)

Typical STEM professional development and curriculum interventions improved student achievement by about 8 percentile points (+0.21 standard deviations). Researchers have estimated that a typical teacher who increases student achievement by +0.14 SD creates marginal gains of roughly $7,000 per child in present value future earnings (Chetty, Friedman, & Rockoff, 2014). Extrapolating from this, the estimated average test score impact of STEM PD and curriculum interventions of +0.21 SD would be expected to yield approximately $10,500 in present value future earnings per student.

The next analyses investigated relationships between professional development characteristics and student impact estimates. Programs that included both teacher PD and curriculum materials together were more effective on average than programs that included only one. For programs that incorporated only professional development or only new curriculum materials, a typical student in the treatment group could be expected to rank about 6 percentile points higher than a typical student in the control group. However, for programs that included *both* PD and curriculum materials together, a typical treatment group student could be expected to score about 10 percentile points higher than a typical control group student.

Several specific program *foci* and *formats* were associated with stronger than typical program impacts. Student outcomes were significantly larger among programs that focused on *how to use curriculum materials*, and among programs that focused on *improving teachers' content and pedagogical content knowledge and/or how students learned the content*, relative to programs that did not have these focus areas. Regarding formats, on average, PD programs that had teachers participate alongside other teachers in their school, *same-school collaboration*, as well as programs that included PD with *implementation meetings*, yielded outcomes that were larger than programs that lacked these components. Implementation meetings typically allowed teachers to convene with other teachers— after initially trying curriculum materials or ideas from professional development— to troubleshoot and discuss ways to improve use of the intervention. Programs that included a summer workshop component were also more effective, on average, than programs that did not meet during the summer; perhaps the summer workshop format provided a prospective, concentrated dose of training at a time when demands on teachers' time and attention were lower, thus bolstering teachers' take-up of the program during the school year.

On the other hand, compared with programs where the professional development was conducted entirely in person, programs where the professional development included an *online component* had smaller (but still positive) impacts, on average.

No significant relationship emerged between the *length* of teacher professional development (measured either as contact hours or as the timespan over which the PD was spread) and impacts on student achievement outcomes at posttest. Likewise, the specific *activities* that teachers engaged in during the professional development (e.g., developing their own curriculum materials or lessons, observing teaching demonstrations) did not reveal any significant relationships between the coded activities and student impact estimates. Nor did significant associations emerge between specific features of the *curriculum materials* (e.g., whether they provided implementation guidance, included lab work and hands-on activities; or the intended length of time for students' use of the materials) and the magnitude of student achievement outcomes.

Impacts of Instructional Improvement Interventions in High-Poverty Settings

The need for instructional improvement in STEM is particularly acute in high-poverty settings. Children from low socioeconomic backgrounds are significantly less likely than their more socioeconomically advantaged counterparts to persist in STEM fields (National Science Foundation, 2017), due in part to insufficient K-12 school quality and STEM coursework opportunities (Tyson et al., 2007).

The current review leveraged data from the just-described meta-analysis (Lynch et al., 2019) to examine the impacts of instructional improvement interventions in high- versus mixed-poverty settings. Controlling for study methods, the analysis suggests a marginal trend toward smaller impacts of instructional improvement interventions in high-poverty settings. This pattern may fit low-SES schools having teachers with weaker baseline preparation (Lankford, Loeb, & Wyckoff, 2002) and fewer resources to support program implementation (Cohen, Raudenbush, & Ball, 2003), thus allowing students in more advantaged settings to garner larger gains from new programs (Ceci & Papierno, 2005). These results are exploratory, as many studies did not include student demographic data. Future empirical reports should include detailed data on student and school socioeconomic characteristics for future research on this issue.

Impacts on Proximal Outcomes: Teacher Practices

As the core purpose of schooling is student learning, analyses have so far focused on instructional improvement programs' impact on student learning in STEM. However, in order for professional development programs to change student learning, they must first alter teachers' classroom practices (e.g., Hamre et al., 2012). Most curriculum programs also have the goal of altering teachers' actions. A handful of meta-analyses pull together the evidence on whether PD and curriculum interventions result in these teacher-level changes. Synthesizing the research base on how teacher professional development in mathematics and science impacts teachers' attitudes and practice (Scher and O'Reilly, 2009) found average positive impacts, but only seven studies met the inclusion criteria. Examining the impact of teacher coaching interventions on instructional practices—operationalized using classroom observation scores from measures of teachers' pedagogical practices, teacher–student interactions, student engagement, and/or classroom climate—found a moderate pooled effect size (Kraft, Blazar, & Hogan, 2018); however, the analysis could not evaluate the measurement properties of the teacher practice measures, and did not differentiate between outcomes on standardized versus researcher-developed teacher practice measures. Finally, a meta-analysis of professional development interventions (Garrett, Citkowicz, & Williams, 2019) reviewed the literature on the impacts of programs aimed at improving classroom practice—defined broadly to include domains such as classroom management, classroom environment, and instructional practices. Although generally positive, impacts on classroom observation outcomes were variable.

Teachers' Reasons for Changing (or Maintaining) Their Practice

Conspicuously absent from most studies of the impacts of teacher professional development and curriculum improvement programs is an underlying theory of adult learning and adult behavior change, specifying why and how teachers abandon their existing practices in favor of new ones. Teaching is a complex behavior (e.g., Clark & Lampert, 1986; Lampert, 2003), and changing teachers' practice often involves challenging and replacing long-held beliefs, attitudes, and habits learned over decades-long "apprenticeships of observation" -- the teachers' experiences as students themselves (Fang, 1996; Philipp, 2007).

One underexplored area of research is whether and how employing principles of *adult behavior change* from adult learning theory or behavioral economics may enhance the effectiveness of teacher professional development and curriculum interventions. Adult learning theory points toward the importance of self-directed learning, adults' intrinsic motivation, problem-centered approaches with practical applications, and reference to adults' prior life experience and knowledge for bolstering adults' learning of complex new topics (e.g., Knowles, Holton, & Swanson, 2012; Merriam, 2001). Yet with some exceptions, relatively few PD and curriculum programs explicitly incorporate concepts from adult learning theory in their design. One exception is a case study of a school psychology intervention (Sanetti, Kratochwill, & Long, 2013), which used adult behavior change principles with the goal of enhancing teachers' take-up and maintenance of a student behavior support intervention. The researchers began the PD by asking the teacher to identify possible barriers to her implementation, then devised specific solutions. The researchers also repeatedly measured the participating teacher's self-efficacy for maintaining the intervention at multiple timepoints during the study, and had a plan in place for mid-stream interventions if her motivation level dipped. The participating teacher implemented the program with high fidelity and quality.

Insights from Behavioral Economics and Social Psychology

Meanwhile, behavioral economists and social psychologists have pointed to the potential importance of behavioral barriers for influencing educational attainment. Recently, the educational literature features nudging—policies that re-frame the choice, aimed at changing behaviors in a predictable fashion without closing off alternatives (Thaler & Sunstein, 2008); factors such as social identity, concern for social belonging, and limited attention span seem influential for educational attainment (Damgaard & Nielsen, 2018). In a handful of studies, several disparate social belonging and identity mechanisms affect teachers' attitudes and behavior: Teachers learn from their interactions with colleagues (Spillane et al., 2012; Sun, Penuel, Frank, Gallagher, & Youngs, 2013), and benefit from opportunities to learn from highly-skilled peers (Jackson & Bruegmann, 2009; Sun, Loeb, & Grissom, 2017). With regard to social identity, teachers' perceptions of shared identity--that they hold interests and values similar to their students--has a positive influence on teacher-student relationships and student outcomes (Gehlbach et al., 2016). Encouraging teachers to find commonalities with their students is thus a promising approach.

Other nudges involve reminders, but the data are sparse. Webinars and an online discussion board for teachers did not significantly increase the effectiveness of a math curriculum intervention (Jackson & Makarin, in press). On the other hand, in the arena of parenting, another highly complex practice, simple text message reminders to engage in literacy activities, such as letter identification, can effectively alter parenting practices and improve student achievement (York, Loeb, & Doss, 2018). To date, we are unaware of analogous studies that examine the efficacy of these reminder interventions in the teacher PD context, pointing toward the possibility of productive future research in this area.

The nudging literature is in its infancy. Yet behavioral economics and adult learning theory could guide the design of PD and curriculum materials in several ways. For example, text messages reminding teachers to use new curriculum materials or to try out a practice discussed in PD could resurface these tasks to the tops of teachers' minds, when they might otherwise be forgotten amidst competing demands. Building ongoing, structured team meetings into a plan for a new curriculum implementation may be beneficial, both by providing social accountability for trying out the new materials, and by meeting teachers' needs for collegial support and advice when challenges to implementation arise.

Policy Insights: Strengthening STEM Instruction in Schools

Teacher professional development and new curriculum can both change teachers' practices and improve students' STEM achievement. The estimates of overall impacts of STEM instructional improvement programs on student achievement are based on experimental and strong quasi-experimental evidence, lending confidence to the overall results. The analyses examining relationships between specific program features and the strength of program impacts are correlational, rather than experimental in nature; as a result, these conclusions about moderators are less definitive but rather suggestive of promising practices. Nonetheless, the findings from dozens of recent evaluations of STEM instructional improvement programs align with prior work (e.g., Boyd, Grossman, Lankford, Loeb, and Wyckoff, 2009; Cohen & Hill, 1998) and are consistent with several recommendations for practice. Specifically, the findings point toward three recommendations for practices that should be implemented in schools and districts:

1. *Focus Professional Development on Curriculum Materials*

Although researchers have sometimes conceptualized teacher professional development and new curriculum materials in isolation, we studied them together. This allowed us to observe that on average, instructional improvement programs that included both components were more effective than those that included PD or curriculum alone. When teachers attend professional development that is not grounded in the specific content they teach, they may find it more difficult to map the PD's techniques onto their own lessons. On the other hand, curriculum materials alone, even those designed to educate teachers, may not be enough to improve teachers' instruction. Programs such as Building Blocks provide an example of how PD that is explicitly tied to the curriculum can be particularly effective. Building Blocks is an effective preschool mathematics intervention that involves both new curriculum materials and ongoing teacher professional development focused on learning about the materials, receiving targeted coaching, and practicing implementation (Clements et al., 2011).

2. Focus on Improving Teachers' Content Knowledge and Understanding of How Students Learn

Affording teachers opportunities to study the curriculum they will teach and to improve their own STEM content and pedagogical content knowledge is beneficial for student learning. What is more, these opportunities should be intellectually engaging to teachers and built on teachers' motivation to learn, rather than being didactic or overly prescriptive. The type of content knowledge needed for teaching is specialized. Given this, programs that offer teachers opportunities to increase their content knowledge specifically in the service of broader pedagogical goals, such as exposing student thinking, may be more effective than programs that simply deliver content (Kennedy, 2016).

3. *Provide Teachers Opportunities to Collaborate and Discuss Implementation Regularly with Teachers in their School*

Carving out time for teachers to collaborate with their same-school colleagues and to meet regularly to discuss program implementation is linked to better program outcomes. Teachers learn from one another, and social motivation and the opportunity to share expertise may also encourage teachers to persist in implementing ambitious new practices. Yet the content of these meetings matters: not all group learning experiences are created equal. For example, simply providing groups of teachers with data on their students' test scores or classroom practices, without guidance on what the group should do with this information, does not appear promising. On the other hand, focusing group work specifically on curricular content, employing discussion leaders to steer conversations on track, and ensuring that group work is intellectually rich are promising avenues to support effective group participation (Kennedy, 2016).

Improving student achievement in STEM is a critical concern for policy. Students' early success in mathematics and science coursework is clearly related to their long-term educational attainment and career prospects, and foundational math and science skills also enhance civic competence and general quality of life. Empirical research evidence can and should inform curriculum selection and teacher workforce development decisions. Strategic investments in STEM instructional improvement programs work, and they can improve schools' productivity and students' long-run opportunities.

References

Borasi, R., & Fonzi, J. (2002). Foundations (vol. 3): Professional development that supports school mathematics reform. *National Science Foundation*.

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, *31*, 416–440.

Carter, D. F. (2006). Key issues in the persistence of underrepresented minority students. New Directions for Institutional Research, 130, 33–46.

Cavanagh, S. (2015, August 17). Spending on instructional materials, construction climbing in K-12. EdWeek. Retrieved from https://marketbrief.edweek.org/marketplace-k-12/instructional_materials_construction_spending_climbing_in_k-12/

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: when the "have-nots" gain but the "haves" gain even more. *American Psychologist*, *60*(2), 149.

Chetty, R., Friedman, J., & Rockoff, J. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9): 2633–2679.

Clark, C., & Lampert, M. (1986). The study of teacher thinking: Implications for teacher education. *Journal of teacher education*, *37*(5), 27-31.

Clements, D. H., Sarama, J., Farran, D., Lipsey, M., Hofer, K. G., & Bilbrey, C. (2011). An Examination of the Building Blocks Math Curriculum: Results of a Longitudinal Scale-Up Study. *Society for Research on Educational Effectiveness*.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 119-142.

Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. New

Haven, CT: Yale University Press.

Confrey, J., & Stohl, V. (Eds.). (2004). On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations. Washington, DC: National Academies Press.

Corcoran, T. B. (1995). Helping teachers teach well: Transforming professional development. Philadelphia, PA: CPRE Policy Briefs. Retrieved from https://files.eric.ed.gov/fulltext/ED388619.Pdf

Damgaard, M. T., & Nielsen, H. S. (2018). Nudging in education. *Economics of Education Review*, *64*, 313-342.

Elmore, R. F. (2002). Bridging the gap between standards and achievement: The imperative for professional development in education. *Secondary lenses on learning participant book: Team leadership for mathematics in middle and high schools*, 313-344.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational research*, *38*(1), 47-65.

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.

Gehlbach, H., Brinkworth, M. E., King, A. M., Hsu, L. M., McIntyre, J., & Rogers, T. (2016). Creating birds of similar feathers: Leveraging similarity to improve teacher–student relationships and academic achievement. *Journal of Educational Psychology*, *108*(3), 342.

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., ... & Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, *49*(1), 88-123.

Jackson, C. K., & Makarin, A. (Forthcoming). Can online off-the-shelf lessons improve student outcomes? Evidence from a field experiment. American Economic Journal: Economic Policy.

Jackson, CK, & Bruegmann, E. (2009). Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers. *American Economic Journal: Applied Economics*, *1*(4): 85-108.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, 86, 945–980.

Knowles, M. S., Holton III, E. F., & Swanson, R. A. (2012). *The adult learner*. Routledge.

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.

Lampert, M. (2003). *Teaching problems and the problems of teaching*. Yale University Press.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational evaluation and policy analysis*, *24*(1), 37-62.

Lavecchia, A. M., Lui, H., & Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. In E. A. Hanushek, S. Machin, & L. Woessmann (Vol. Eds.), Handbook of the economics of education: 5, (pp. 1–74). Elsevier.

Lynch, K., Hill, H.C., Gonzalez, K, & Pollard, C. (in press). Strengthening the Research Base that Informs STEM Instructional Improvement Efforts: A Meta-Analysis. *Educational Evaluation & Policy Analysis*. doi: 10.3102/0162373719849044.

Merriam, S. B. (2001). Andragogy and self-directed learning: Pillars of adult learning theory. *New directions for adult and continuing education*, *2001*(89), 3-14.

Miles, K. H., Odden, A., Fermanich, M., & Archibald, S. (2004). Inside the black box of school district spending on professional development: Lessons from five urban districts. Journal of Education Finance, 30, 1–26.

Miller, B., Lord, B., & Dorney, J. A. (1994). Staff development for teachers: A study of configurations and costs in four districts. Boston, MA: Education Development Center.

NAEP (2017). 2017 NAEP Mathematics and Reading Assessments: Highlighted Results at Grades 4 and 8 for the Nation, States, and Districts. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018037

NAEP (2015). The Nation's Report Card: 2015 Science at Grades 4, 8 and 12. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2016162

National Center for Education Statistics (NCES). Teacher Preparation and Professional Development: 2000. NCES2001-088. Washington D.C.: U.S. Department of Education, 2001.

National Science Foundation. (2017). Women, minorities, and persons with disabilities in science and engineering (WMPD) report. Retrieved from https://nsf.gov/statistics/2017/nsf17310/

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. *Second handbook of research on mathematics teaching and learning*, *1*, 257-315.

Porter, A. C., Garet, M. S., Desimone, L., Yoon, K. S., & Birman, B. F. (2000). Does professional development change teaching practice? Results from a three-year study.

Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. American Educational Research Journal, 45, 206–230.

Russell, M. L., & Atwater, M. M. (2005). Traveling the road to success: A discourse on persistence throughout the science pipeline with African American students at a predominantly white institution. *Journal of Research in Science Teaching*, *42*(6), 691-715.

Sanetti, L. M., Kratochwill, T. R., & Long, A. C. (2013). Applying adult behavior change theory to support mediator-based intervention implementation. *School Psychology Quarterly*, *28*(1), 47.

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,* 209–249.

Shavelson, R. J., & Towne, L. (Eds.). (2001). Scientific research in education. Washington, DC: The National Academies Press.

Spillane, J. P. (2012). *Distributed leadership* (Vol. 4). John Wiley & Sons.

Sun, M., Loeb, S., & Grissom, J. A. (2017). Building teacher teams: Evidence of positive spillovers from more effective colleagues. *Educational Evaluation and Policy Analysis*, *39*(1), 104-125.

Sun, M., Penuel, W. R., Frank, K. A., Gallagher, H. A., & Youngs, P. (2013). Shaping professional development to promote the diffusion of instructional expertise among teachers. *Educational Evaluation and Policy Analysis*, 35(3), 344-369.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk*, *12*(3), 243–270.

Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal*, *50*(5), 1081-1121.

Yoon, K. S., Duncan, T., Lee, W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs

York, B. N., Loeb, S., & Doss, C. (2018). One Step at a Time: The Effects of an Early Literacy Text Messaging Program for Parents of Preschoolers. *Journal of Human Resources*, 0517-8756R.