

***The effects of teachers' professional development on student achievement:
Findings from a systematic review of evidence***

Kwang Suk Yoon (American Institutes for Research)
Teresa Duncan (American Institutes for Research)
Sylvia Lee (Taiwan National University)
Kathy Shapley (Edvance Research)

Paper presented at the Annual Meeting of the American Educational Research Association,
March 24-28, 2008, New York

We would like to thank those who participated in the review study as coder or assistant: They are Beth Scarloss, James Taylor, Stacey Merola, Kaz Uekawa, Andrea Martin, Andrei Streke, Adam Segal, and Sandra Tang.

Overview

In the past 20 years, teacher professional development (PD) has been used as one of key policy mechanisms for educational reform initiatives (Cohen & Hill, 2000; Corcoran et al., 1998). The passage and implementation of the No Child Left Behind Act (*NCLB*) heightened the attention given to teacher PD as a potential strategy to boost students' academic gains and close the achievement gaps. *NCLB* mandates all teachers to receive high quality PD. In addition, *NCLB* requires that activities supported with Title II funds must be based on a review of scientifically based research that shows how such interventions are expected to improve student achievement. Furthermore, the law stipulates that PD activities are regularly evaluated for their impact on teacher effectiveness and improved student academic achievement. In light of *NCLB*'s emphasis on assessing the impact of PD on improved student learning, it is imperative to examine the available research-based evidence on the effectiveness of PD as it relates to student academic outcomes. It is also important to identify effective professional development programs and models that exist.

Accordingly, we embarked on a systematic review of evidence on the effects of PD on growth in student learning. Our major objective is to use the rigorous standards of the What Works Clearinghouse (WWC) to assess “what works” in PD and to make this information accessible to researchers and practitioners so that they can make informed decisions about PD in terms of planning, designing, implementing, and allocating resources. The main focus of this review was on PD activities that are designed to enhance K-12 teachers' knowledge and skills, to transform their classroom practices, and to improve their students' achievement in three core academic subjects: English/language arts/reading, mathematics, and science. The project was sponsored by the Regional Education Laboratory – Southwest (REL SW), and funded by the Institute of Education Sciences of the U.S. Department of Education.¹

Theoretical Framework

Showing that professional development translates into gains in student achievement poses tremendous challenges, despite an intuitive and logical connection between them (Borko, 2004; Loucks-Horsley & Matsumoto, 1999; Supovitz, 2001). To substantiate the empirical link between professional development and student achievement, studies should *ideally* establish two points. One is that there are links among professional development, teacher learning and practice, and student learning. The other is that the empirical evidence is of high quality—that the study proves what it claims to prove. This paper focuses on the second point, treating the first only briefly as a heuristic or logic model.

The links among PD, teacher learning and practice, and student learning

¹ This AERA paper is based on our final report that was submitted to the U.S. Department of Education Institute of Education Science. <http://ies.ed.gov/ncee/edlabs/projects/project.asp?id=70>.

Consistent with several existing models of effective PD (Cohen & Hill, 2000; Fishman, Marx, Best, & Tal, 2003; Garet et al., 2001; Guskey & Sparks, 2004; Kennedy, 1998; Loucks-Horsley & Matsumoto, 1999), we assume that the effects of PD on student achievement are mediated by teacher knowledge and teaching in the classroom. As our logic model (shown in Figure 1) indicates, we also assume that teacher's PD and instruction as well as student learning take place in the context of high standards, challenging curriculum, system-wide accountability, and high-stakes assessment.

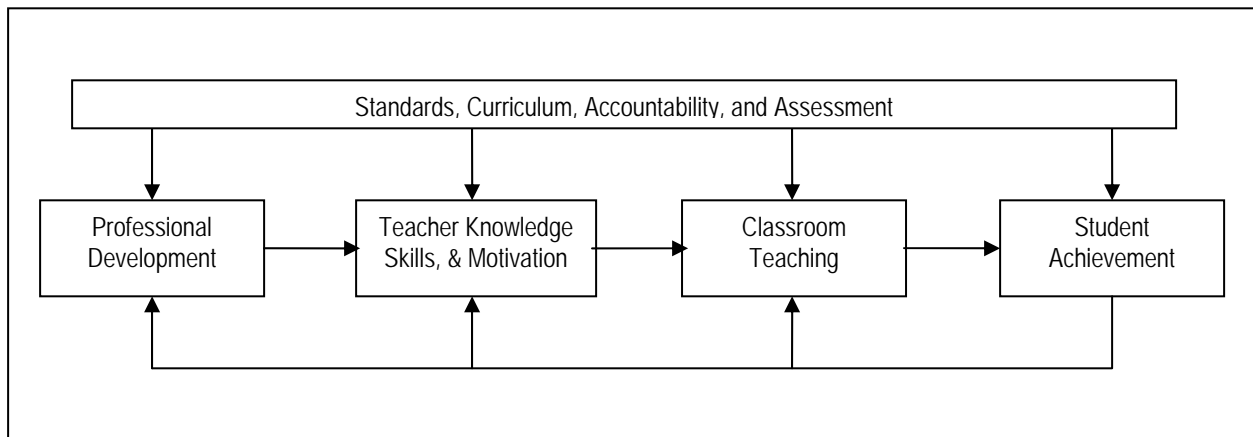


Figure 1. Logic model of the impact of professional development on student achievement

Professional development affects student achievement through three steps. First, professional development enhances teacher knowledge, skills, and motivation. Second, better knowledge, skills, and motivation improve classroom teaching. Third, improved teaching raises student achievement. If one link is weak or missing, better student learning cannot be expected. If a teacher fails to apply new ideas from professional development to classroom instruction, for example, students will not benefit from the teacher's professional development. In other words, the effect of PD on student learning is possible through two mediating outcomes: teachers' learning, and instruction in the classroom.

In the first step, for professional development to improve teachers' knowledge, skills, and motivation, it must be of high quality in its theory of action, planning, design, and implementation. More specifically:

- It should be based on a carefully constructed and empirically validated theory of teacher learning and change (Ball & Cohen, 1999; Richardson & Placier, 2001; Sprinthall, Reiman, & Thies-Sprinthall, 1996).
- It should promote and extend effective curricula and instructional models—or materials based on a well defined and valid theory of action (Cohen, Raudenbush, & Ball, 2002; Hiebert & Grouws, 2007; Rossi, Lipsey, & Freeman, 2004).
- It should be intensive, sustained, content-focused, coherent, well defined, and strongly implemented (Garet et al., 2001; Guskey, 2003; Loucks-Horsley, Hewson, Love, & Stiles, 1998; Supovitz, 2001; Wilson & Berne, 1999).

In the second step, teachers apply their enhanced knowledge, skills, and motivation to classroom teaching (Borko, 2004; Showers, Joyce, & Bennett, 1987), supported by ongoing school collaboration and follow-up consultations with experts. Doing so could require overcoming such barriers to new practices as lack of time for preparation and instruction, limited materials and human resources, and lack of follow-up support from professional development providers.

In the third step, teaching—improved by professional development—raises student achievement. The challenge is measuring and validating the gains.

The quality of empirical evidence

Establishing the second point—that the empirical evidence is of high quality—is the primary focus of this paper, which examines the rigor of empirical studies conducted to validate the effects of professional development (National Research Council, 2004). Even if professional development enhances teacher knowledge and skills and improves classroom instruction, a poorly designed evaluation or inadequate implementation would make it difficult to detect any effects from the professional development.

What is required for establishing the empirical link between professional development and student achievement? That empirical link is based on at least four elements:

- A rigorous research design must ensure the internal validity of causal inferences about the effectiveness of professional development. Using a study design with strong internal validity (a randomized controlled trial, for example) can rule out competing explanations for gains in student academic achievement. The research design should be able to measure the value that professional development adds to student learning separately from the value added by innovative curricula, instruction, or materials. A rigorous research design must also have externally valid findings, adequate statistical power to detect true effects, and sufficient time between the professional development and the measurement of teacher and student outcomes.
- The study design must be executed with high fidelity and sufficient implementation of professional development
- Psychometric properties of measures must be adequate (measures of classroom teaching practices, of student achievement, and of teacher knowledge, beliefs, and behaviors). Measures should be valid, reliable, age-appropriate, and sensitive to and aligned with the intervention.
- Analytic models must be well-specified and statistical methods must be appropriate

Given these requirements, it is unsurprising that few rigorous studies address the effect of professional development on student achievement (Borko, 2004; Clewell, Campbell, & Perlman, 2004; Kennedy, 1998; Killion, 1999; Loucks-Horsley & Matsumoto, 1999; Supovitz, 2001). There is more literature on the effects of professional development on teacher learning and teaching practice, falling short of demonstrating effects on student achievement (e.g., Garet et al., 2001). In addition, even more literature addresses curricular or instructional effectiveness (National Research Council, 2004; various What Works Clearinghouse intervention reports).

One systematic review of the effects of professional development on student achievement is Kennedy (1998). That review analyzes the relative effects on student outcomes from professional development programs for math and science, examining the professional development's subject, content focus, skill level, form, and other features (e.g., intensity and concentration).² She concluded that:

“Programs whose content focused mainly on teachers’ behaviors demonstrated smaller influences on student learning than did programs whose content focused on teachers’ knowledge of the subject, on the curriculum, or on how students learn the subject” (p. 18).

Kennedy’s seminal review indicates the importance of content focus in high quality professional development (see also Desimone, Porter, Garet, Yoon, & Birman, 2002; Garet et al., 2001; Yoon, Garet, Birman, & Jacobson, 2007). There are three reasons, however, for a new systematic review to supplement those of Kennedy and of Clewell, Campbell, and Perlman (2004). First, the volume of literature has grown, especially after standards-based reform prompted a wave of professional development–related studies. Second, most of the literature reviews and research syntheses are limited in scope, source, and subject. Few literature reviews encompass the three core academic subjects under No Child Left Behind accountability requirements (reading/English/language arts, mathematics, and science). A more comprehensive and systematic review of evidence that professional development works in these critical subject areas is needed. Third, the growing emphasis on effective professional development practices supported by scientifically based research makes it imperative to apply rigorous evidence standards—such as those of the What Works Clearinghouse—in new literature reviews and syntheses.

Methods

This systematic review work involved a series of carefully planned steps. It began with keyword searches of seven electronic databases: ERIC, PsycINFO, ProQuest, EBSCO’s Professional Development Collection, Dissertation Abstracts, Sociological Collection, and Campbell Collaboration.³ A deliberately wide net of keywords was used to capture literature – published and unpublished – on professional development and student learning in three core content areas: language arts, mathematics, and science. The search identified 1,343 citations as potentially addressing the impact of professional development on student learning outcomes.

Next, prescreening was performed by scanning the abstracts or full texts of the 1,343 studies to determine if they met broad relevance and methodology criteria (e.g., an empirical study involving professional development and some measure of student achievement). The prescreening process reduced the list to 132 studies that were considered relevant for systematic

² For Kennedy’s (1998) classification of PD content, see Appendix B.

³ For the complete list of keywords, refer to Appendix D in the final report.

review. These studies were then subjected to three stages of coding.⁴

Stage 1 Coding examined the relevance of the studies using the following criteria:

- **Topic.** The study had to deal with the effects of professional development on student learning in at least one of three core content areas (language arts, mathematics, and science).
- **Population.** The sample had to include teachers of language arts, mathematics, or science and their students in grades K-12.
- **Outcome.** The study had to measure student learning outcomes.
- **Study design.** The study had to be empirically based and use randomized controlled trials or some form of quasi-experimental design.
- **Time.** The study had to be published between 1986 and 2006.
- **Country.** The study had to take place in Australia, Canada, the United Kingdom, or the United States, due to concerns about the external validity of the findings.

It is important to note that about two thirds of studies failed to meet the study design criterion at Stage 1 Coding (see Table 1). It turned out that a great majority of empirical investigations examining the effects of PD on student achievement were not rigorous enough to meet the WWC’s standards of eligible studies.

Table 1. Studies failing and passing stage 1 criteria

Stage 1–full screening criterion	Failing		Passing	
	Number	Percentage	Number	Percentage
Focus on in-service professional development	30	22.7	102	77.3
Focus on K–12 teachers and their students	2	1.5	130	98.5
Country	8	6.1	124	93.9
Time of study	13	9.9	119	90.1
Study design	84	63.6	48	36.4
Focus on the specified subjects	24	18.2	108	81.8
Focus on the effects of in-service professional development on student achievement outcomes	38	28.8	94	71.2
Overall stage 1 screening decision	105	79.5	27	20.5

Note: Each row contains 132 studies. Questions about adequate psychometric properties were asked only if all seven preceding criteria were met. Because not all 132 studies were subject to that question, it is excluded from this table.

The results of Stage 1 Coding yielded 27 relevant studies that were eligible for review in terms of study quality ratings.

Stage 2 Coding focused on quality rating of the 27 eligible studies using the U.S. Department of Education’s What Works Clearinghouse (WWC) Evidence Standards.⁵ At this stage, each study was given one of three possible ratings in accordance with the WWC technical guidelines: “Meets Evidence Standards” (e.g., randomized controlled trials that provided the strongest

⁴ See Appendix A for a flowchart of the review process.

⁵ See <http://ies.ed.gov/ncee/wwc/overview/review.asp?ag=pi> for the WWC Evidence Standards.

evidence of causal validity), “Meets Evidence Standards with Reservations” (e.g., quasi-experimental studies or randomized controlled trials that had problems with randomization, attrition, teacher-intervention confound, or disruption), and “Does Not Meet Evidence Standards” (e.g., studies that did not provide strong evidence of causal validity).⁶ Only nine studies were found to meet the WWC Evidence Standards.

Stage 3 Coding involved the computation of effect sizes and thorough descriptions of the characteristics of professional development activities (e.g., form, content, duration) employed in the nine studies.

Results

We reviewed more than 1,300 studies identified as potentially addressing the impact of teacher professional development on student achievement. Only nine meet What Works Clearinghouse evidence standards—attesting to the paucity of rigorous studies that directly examine the effect of in-service teacher professional development on student achievement in the three core academic subjects. For studies not meeting evidence standards despite focusing on teacher professional development and including a student achievement measure, a frequent problem was study design, particularly for quasi-experimental designs with problems in baseline equivalence between treatment and comparison groups. The nine studies are:

- Carpenter, Fennema, Peterson, Chiang, & Loef (1989).
- Cole (1992).
- Duffy et al. (1986).
- Marek & Methven (1991).
- McCutchen et al. (2002).
- McGill-Franzen, Allington, Yokoi, & Brooks (1999).
- Saxe, Gearhart, & Nasir (2001).
- Sloan (1993).
- Tienken (2003).

These nine studies can be characterized as follows:⁷

- All focused on elementary school teachers and their students. About half focused on lower elementary grades (kindergarten and first grade), and about half on upper elementary grades (fourth and fifth grades).
- Four of the investigations included measures of student learning in reading and language arts. Two studies focused on mathematics; one on science; and two on language arts, mathematics, and science.

⁶ The factors that affected the quality rating of the studies included the following:

- Problems with randomization in case of RCTs
- Problems with baseline equivalence in case of QED studies
- Differential attrition
- Statistical properties important for computing accurate effect sizes.

⁷ For detailed descriptions of the nine studies and their effects, refer to our final report.

- Six studies were published in peer-reviewed journals; three were unpublished doctoral dissertations.
- The studies were not particularly recent, ranging from 1986 to 2003.
- Five studies were RCTs that met evidence standards without reservations.
- The remaining four studies met evidence standards with reservations (one RCT with group equivalence problems and three QEDs).
- Seven studies used standardized measures of achievement. One used researcher-developed measures of students' knowledge of fractions, and one used Piagetian conservation tasks as the outcome.
- The number of teachers involved in these studies ranged from five to 44; the number of students from 98 to 779.
- All nine studies employed workshops or summer institutes.
- In all but one study, the main PD event was supported with a number of follow-up sessions.
- In all nine studies, PD was provided directly to teachers (rather than using a train-the-trainer approach), and delivered by the author(s) or their affiliated university researchers.

Twenty effect sizes and improvement indices were computed across the nine studies (see Table 2).

- The average effect size across the nine studies was 0.54, ranging from -0.53 to 2.39. The average improvement index was 21 (which can be translated to boosting students' achievement by 21 percentile points), ranging from -20 to 49.
- Only one effect was negative (in Saxe et al., 2001), and only one effect was zero (in Duffy et al., 1986). The other 18 effects were positive, with effect sizes ranging from 0.12 to 2.39 (with improvement indices from 5 to 49).
- Of the 20 effects, 12 were not statistically significant after applying necessary corrections for unaddressed clustering and multiple outcomes. Nine of those twelve, however, are substantively important according to What Works Clearinghouse conventions.

[Table 2 about here]

Disaggregating the studies by their content-area outcomes allowed computing averages and ranges for science, mathematics, and reading and English/language arts. Science had only 2 effects, mathematics had 6, and reading and English/language arts had 12. The average effect was remarkably consistent across the three content areas. The average effect size in science was 0.51; in mathematics, 0.57; and in reading and English/language arts, 0.53.

Effects by form, contact hours, intensity, and duration of professional development

All nine studies employed workshops or summer institutes. In all but one follow-up sessions supported the main professional development event. Marek and Methven (1991) was the exception; that study provided an intensive four-week summer workshop without follow-up support. In all nine studies professional development went directly to teachers rather than through a train-the-trainer approach and was delivered by the authors or their affiliated researchers.

The professional development in these studies varied in total contact hours, duration, and intensity. The total contact hours ranged from 5 hours to 100, with an average of 49 hours. Marek and Methven (1991) provided 100 hours of professional development over four weeks, while McCutchen et al. (2002) provided about the same number of contact hours but over 10 months, offering more sustained, if less intensive, development. Studies that had greater than 30 hours of professional development showed a positive and significant effect on student achievement from professional development. The three studies that involved the least amount of professional development (5–14 hours of contact time) showed no statistically significant effects on student achievement.

Because of the lack of variability in form and the great variability in duration and intensity in this small number of studies, discerning any pattern between these characteristics and their effects on student achievement is difficult. A larger number of rigorous studies on the link between professional development and student achievement might have made it possible to determine whether intensive, sustained, and content-focused professional development is more effective (Ball & Cohen, 1999; Garet et al., 2001; Joyce & Showers, 1995; Loucks-Horsley, Stiles, & Hewson, 1996; Wilson & Berne, 1999; Yoon et al., 2007).

Effects by the content of professional development

The fourfold content-group classification scheme for professional development in Kennedy (1998) helps characterize the professional development models and theories of actions in the nine studies (see Appendix B for the classification). The professional development in the nine studies varied much more in content and substance than in form—as predicted in Kennedy (see Table 3). Likewise, Spillane (2000, p. 23) notes that “structural similarities in district professional development approaches (e.g., classroom demonstrations, peer coaching) camouflaged substantial differences in the underlying theories of teacher learning and change.” The limited number of studies and the variability in their professional development models precludes drawing any definitive links between content-group classification and effects on student achievement.

[Table 3 about here]

Conclusions and Discussion

Few studies meet evidence standards. But the average effect size of 0.54 in mathematics, science, and reading and English/language arts—and the consistency of that effect size—indicates that providing professional development to teachers has a moderate effect on student achievement across the nine studies. Average control group students would have increased their achievement by 21 percentile points if their teacher had received professional development.

Results in mathematics are of particular note, given the recent study on professional development in mathematics that was carried out to evaluate the state and local implementation of NCLB (Birman et al., 2007). Birman et al. (2007) showed that few teachers receive intensive, sustained, and content-focused professional development in mathematics. They found that teachers

averaged 8.3 hours of professional development on how to teach mathematics and 5.2 hours on the “in-depth study” of topics in mathematics during the 12 months spanning the 2003/04 school year and the summer of 2004. Of elementary teachers, 71 percent participated in professional development focused on instructional strategies for teaching mathematics. But only 9 percent participated for more than 24 hours during the one-year period. Even fewer elementary school teachers (49 percent) reported that they participated in professional development focused on the in-depth study of mathematics during the same time period, and only 6 percent participated for more than 24 hours. Of secondary mathematics teachers, 51 percent attended professional development focused on the in-depth study of mathematics, but only 10 percent spent more than 24 hours on that content during the year.

Four studies in mathematics reviewed here generated six effects, averaging 0.57, with an improvement index of 22 percentile points. The PD contact hours in the four studies averaged just over 53 hours, ranging from 30 hours to 83 hours, over a period of four months to one year. This professional development is longer than that of the typical elementary school teacher—only 9 percent of elementary school teachers participated in mathematics professional development for more than 24 hours over a year in Birman et al. (2007).

This report cannot determine definitively whether the professional development in the four studies meets other criteria for high quality professional development in the literature (using active learning and collective participation, for example) or in No Child Left Behind (consistent with state academic content standards and involving strategies from scientifically based research). Even so, the gap between the amount of professional development found effective in the four studies and the average received by elementary school teachers is worth considering.

The limited number of studies and the variability in their professional development approaches preclude any conclusions about the effectiveness of specific professional development programs or about the effectiveness of professional development by form, content, or intensity. Recently publicized National Mathematics Advisory Panel’s final report (2008) reached a similar conclusion as we did:

“Although the Panel did find some positive effects of professional development on students’ achievement gains, research does not yield sufficient evidence on the features of any particular approach to permit detailed conclusions about the forms of or approaches to effective professional development” (p. 40).

Some might argue that the What Works Clearinghouse’s evidence standards used to select the studies included in this paper are unduly rigorous and that their use eliminated many good studies that other adequate but less restrictive criteria would not. Including these other studies might substantially change the complexion of the analysis and yield quite different results. Mary Kennedy’s review, for example, included a different set of empirical studies, mostly due to different selection criteria. Using less stringent criteria, her review – if not a systematic review as employed in our study – yielded a broader range of PD programs varying in forms and content. Possibly due to this difference in the selection criteria and review approach, our study was not able to replicate her finding that PD content matter; that is, PD activities that provide teachers with pedagogical content knowledge are more effective than those based on content-free

pedagogical practice. On the other hand, as the National Mathematics Advisory Panel's review demonstrated, when a similar systematic review approach is employed, a convergent conclusion is drawn as noted above.

These findings are important, but note a number of limitation and caveats of this study:

First, none of the nine studies focused on professional development's effects on middle or high school students. We simply lack evidence of effective PD program at the secondary level in reading, math, and science.

Second, even the studies meeting evidence standards were generally underpowered. In addition, some of the studies failed to address statistical problems with clustering or multiple comparisons. In this study, we corrected for such problems. As a result of lack of power and statistical corrections, 12 effects of 20 were not statistically significant.

Third, greater resources and time would allow a more comprehensive literature search for comparison. Using different keywords for search might generate a larger pool, for example. And more studies might meet evidence standards if authors could be contacted for additional information.

Fourth, each of the 9 studies and the 20 effects are treated equally, regardless of differences in type of professional development, sample sizes, or quality of research design. Because some studies included several outcome measures, those studies are overrepresented in the average overall effect. For example, McGill-Franzen et al. (1999) accounts for six effect sizes in the overall average.

Fifth, this research synthesis conducts none of the additional data manipulations of traditional meta-analysis, such as differential weighting. The intent was to adhere as closely as possible to What Works Clearinghouse procedures.⁸ Although the What Works Clearinghouse computes an average effect size for a study and uses an average of study averages to report an overall average effect size, the studies in a What Works Clearinghouse intervention report address one intervention. This report, however, addresses several interventions, and the studies were few enough to merit limiting any additional aggregation, given the diversity among the nine studies in content areas and professional development approaches. So, the individual effects and the overall average are the only ones included. Interpreting the overall average effect size of 0.54 also requires caution.⁹ This effect is only a preliminary marker on the sparsely populated terrain of professional development research, still at its developmental stage (Borko, 2004).

⁸ Traditional meta-analysis would weight the studies to account for differences in numbers of effects in each study and the variability in sample sizes across studies. The argument for doing so is that differential weighting affords greater power and precision. The What Works Clearinghouse, however, has not adopted a traditional meta-analysis approach.

⁹ Following What Works Clearinghouse procedures, this report does not conduct a test of statistical significance on the average effect size, as would have been done in a traditional meta-analysis.

This report is a first step. There remain a lot of questions to be addressed. For example, how effective is coaching as an approach to PD compared to other traditional approach such as workshop? What's the optimal level of sustained and intensive PD? Unfortunately, we do not yet have sufficient credible evidence on relative efficacy or effectiveness of various PD approaches, content, forms, and durations. At this early stage of PD impact research, we have demonstrated that there is only a limited amount of what Borko (2004) called “existence proofs of effective professional development” based on several small-scale, underpowered, efficacy trials. We need to expand the research base by doing new good efficacy trials, replicating some of the existing studies that demonstrated some efficacy or potential, and subsequently conducting a number of effectiveness trials. As professional development research matures, individual empirical studies of multiple professional development programs will eventually make it possible to judge the effectiveness of individual programs, taking into account such factors as the quality of the study design, statistical significance of the findings, and direction and magnitude of the findings—as does the What Works Clearinghouse classification. Fortunately, several large-scale, randomized studies of the impact of professional development on student learning funded by the Institute of Education Sciences are now underway to answer questions that could not be answered in this analysis.

Implications for Researchers and Practitioners

Our research synthesis highlighted the problems and pitfalls of many studies of PD effectiveness.¹⁰ By documenting how studies dropped off as we proceeded through the coding process, we hope this helps to provide researchers direction for designing methodologically rigorous studies that produce valid evaluation of PD effects. In this regard, efforts are being made to improve the rigor of studies specifically designed to examine the effectiveness of PD on student achievement, especially in the context of randomized controlled trials (e.g., Wayne, Yoon, Cronen, Garet, & Zu, 2007). For example, PD impact studies are to address a number of methodological issues such as appropriate counterfactuals of PD, alignment between PD and outcome measures, attrition of samples (teachers and students), and potentially confounding effects of PD and other instructionally relevant variables such as curriculum and instructional materials. Even if RCTs are not feasible, other methods such as QED or regression discontinuity design are also available to substantiate potential effects of PD. We urge researchers who are designing QEDs to ensure that they provide information about and evidence for the baseline equivalence of the treatment and comparison groups.

In addition, to facilitate future research synthesis, we suggest researchers provide structural abstract in their report (Kelly & Yin, 2007; Mosteller et al, 2004).

Future studies of the effect of professional development on *both* teachers *and* students would be particularly useful—fully addressing PD's *mediated* effects on students' achievement outcomes

¹⁰ The final report describes in details why some studies failed to pass prescreen or meet evidence standards at different stages of review.

through teachers' enhanced knowledge and skill as well as their improved instruction in the classrooms.¹¹

Finally, our findings suggest that practitioners need to be vigilant of PD programs that do not have research evidence to base their claims of effectiveness (e.g., "anecdotal evidence"). Practitioners must demand evidence-based strategies and practices from their PD providers. As of yet, though, such valuable evidence is in a short supply. Standards-based school reform efforts including the implementation of NCLB may not reap students' gains without teachers' enhanced knowledge, skills, and instruction in the classrooms. Researchers and practitioners alike should strive to strengthen the evidence base of the very core of their school-reform efforts by doing more of rigorous studies on PD, teachers' learning and practice, and students' achievement.

¹¹ Our study was one of fast-turnaround projects awarded by the Institute of Education Science. Due to such limited scope, timeline, and resources of our work, we could not afford to examine the full mediated model of the effects of PD on students, as shown in our logic model (see Figure 1). Instead, our study was focused solely on the overall effect of PD on student achievement, without regard to any mediating effects of teachers' knowledge and instruction.

References

- Ball, D. L., & Cohen, D. K. (1999). Developing practices, developing practitioners: Toward a practice-based theory of professional development. In G. Sykes & L. Darling-Hammonds (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 30–32). San Francisco, CA: Jossey-Bass.
- Birman, B., LeFloch, K. C., Klekotka, A., Ludwig, M., Taylor, J., Walters, K., Wayne, A. and Yoon, K. S. (2007). *State and Local Implementation of the No Child Left Behind Act, Volume II—Teacher Quality Under NCLB: Interim Report*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15.
- Carpenter, T. P., Fennema, E., Peterson, P.L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children’s mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26(4), 499–531.
- Clewell, B. C., Campbell, P. B., & Perlman, L. (2004). *Review of evaluation studies of mathematics and science curricula and professional development models*. Submitted to the GE Foundation. Washington, DC: Urban Institute.
- Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, 102(2), 294–343.
- Cohen, D. K., Raudenbush, S., & Ball, D. L. (2002). Resources, instruction, and research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research*, (pp. 80–119). Washington, DC: Brookings Institution Press.
- Cole, D. C. (1992). The effects of a one-year staff development program on the achievement of test scores of fourth-grade students. *Dissertation Abstracts International*, 53(06), 1792A. (UMI No. 9232258)
- Corcoran, T. B., Shields, P. M., & Zucker, A. A. (1998). *The SSIs and professional development for teachers*. Menlo Park, CA: SRI International.
- Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan*, 76(8), 597–604.
- Desimone, L., Porter, A. C., Garet, M., Yoon, K. S., & Birman, B. (2002). Does professional development change teachers’ instruction? Results from a three-year study. *Educational Evaluation and Policy Analysis*, 24(2), 81–112.
- Duffy, G. G., Roehler, L. R., Meloth, M. S., Vavrus, L. G., Book, C., Putnam, J., & Wesselman, R. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: A study of reading teacher effects. *Reading Research Quarterly*, 21(3), 237–252.

- Elmore, R. F. (1997). *Investing in teacher learning: Staff development and instructional improvement in Community School District #2, New York City*. New York, NY: National Commission on Teaching & America's Future.
- Fishman, B. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*, 643–658.
- Garet, M., Birman, B. F., Porter, A. C., Desimone, L., Herman, R., & Yoon, K. S. (1999). *Designing effective professional development: Lessons from the Eisenhower program*. Washington, DC: American Institutes for Research.
- Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., and Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945.
- Guskey, T., & Sparks, D. (2004). Linking professional development to improvements in student learning. In E. M. Guyton & J. R. Dangel (Eds.), *Research linking teacher preparation and student performance: Teacher education yearbook XII* (pp. 233–247). Dubuque, IA: Kendall/Hunt.
- Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *The second handbook of research in mathematics education*. Reston, VA: New Age and National Council of Teachers of Mathematics.
- Hunter, M. (1984). Knowing, teaching, and supervising. In P.L. Hosford (Ed.), *Using what we know about teaching* (pp. 169-192). Alexandria, VA: Association for Supervision and Curriculum Development.
- Joyce, B. & Showers, B. (1995). *Student achievement through staff development: Fundamentals of school renewal* (2nd ed.). New York: Longman.
- Kennedy, M. (1998). *Form and substance of inservice teacher education* (Research Monograph No. 13). Madison, WI: National Institute for Science Education, University of Wisconsin–Madison.
- Killion, J. (1999). *What works in the middle: Results-based staff development*. Oxford, OH: National Staff Development Council.
- Little, J. W. Teachers' professional development in a climate of educational reform. (1993). *Educational Evaluation & Policy Analysis, 15*(2), 129–151.
- Loucks-Horsley, S., Stiles, K. & Hewson, P. (1996). *Principles of effective professional development for mathematics and science education: a synthesis of standards*. Madison, WI: University of Wisconsin at Madison, National Institute for Science Education.
- Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.
- Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and Mathematics, 99*(5), 258–271.

- Marek, E. A., & Methven, S. B. (1991). Effects of the learning cycle upon student and classroom teacher performance. *Journal of Research on Science in Teaching*, 28(1), 41–53.
- McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., Quiroga, T., & Gray, A. L. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities*, 35(1), 69–86.
- McGill-Franzen, A., Allington, R. L., Yokoi, L., & Brooks, G. (1999). Putting books in the classroom seems necessary but not sufficient. *Journal of Reading Research*, 93(2), 67–74.
- National Commission on Teaching and America’s Future (1996). *What matters most: Teaching for America’s future*. New York, NY: Author.
- National Mathematics Advisory Panel (2008). *Foundations for Success*. Washington, DC: Authors
- National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations*. Washington, DC: National Academies Press.
- Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.). *Handbook of Research on Teaching* (4th Ed., pp. 905–947) Washington, DC: American Education Research Association.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th Ed.) Thousand Oaks, CA: Sage Publications.
- Saxe, G. B., Gearhart, M., & Nasir, N. S. (2001). Enhancing students’ understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education*, 4, 55–79.
- Showers, B., Joyce, B., & Bennett, B. (1987). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. *Educational Leadership*, 45(3), 77–87.
- Sloan, H. A. (1993). Direct instruction in fourth and fifth grade classrooms. *Dissertation Abstracts International*, 54(08), 2837A. (UMI No. 9334424)
- Spillane, J. P. (2000). *District leaders’ perceptions of teacher learning* (No. CPRE Occasional Paper Series OP-05). Philadelphia: Consortium for Policy Research in Education.
- Sprinthall, N. A., Reiman, A. J., & Thies-Sprinthall, L. (1996). Teacher professional development. In J. Sikula (Ed.), *Handbook of research on teacher education* (2nd Ed., pp. 666–703). New York, NY: Macmillan.
- Supovitz, J. A. (2001). Translating teaching practice into improved student achievement. In S. Fuhrman (Ed.), *National Society for the Study of Education Yearbook*. Chicago, IL: University of Chicago Press.
- Tienken, C. H. (2003). The effect of staff development in the use of scoring rubrics and reflective questioning strategies on fourth-grade students’ narrative writing performance. *Dissertation Abstracts International*, 64(02), 388A. (UMI No. 3081032)
- U.S. Department of Education (2001). *Teacher preparation and professional development: 2000* (National Center for Education Statistics Report No. 2001-088). Washington, DC: Author.

- Wayne, A., Yoon, K. S., Cronen, S., Garet, M. S., & Zhu, P. (November 2007). *Issues in designing randomized field trails of teacher professional development*. Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, Washington, DC.
- Wilson, S. M. & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, 24, 173–209.
- Yoon, K. S., Garet, M., Birman, B. & Jacobson, R. (2007). Examining the effects of mathematics and science professional development on teachers' instructional practice: Using professional development activity log. Washington, DC: Council of Chief State School Officers.

Table 2. Effects of professional development on student achievement, by study

Study (study design)	Outcome measure	Effect size	Applied correction for clustering or multiple comparisons?	Recomputed statistical significance	Improvement index
Carpenter et al., 1989 (RCT)	Iowa Test of Basic Skills Level 7, computation	0.41	None applied if author did not report significant results	Not significant, but substantively important	16
	Iowa Test of Basic Skills Level 7, problem solving	0.41	None applied if author did not report significant results	Not significant, but substantively important	16
Cole, 1992 (RCT)	Average for math	0.50	Yes	Statistically significant	19
	Average for reading	0.82	Yes	Statistically significant	29
	Average for language	0.24	None applied if author did not report significant results	Not significant	9
Duffy et al., 1986 (RCT)	Gates-MacGinitie Reading Test	0.00	None applied if author did not report significant results	Not significant	0
Marek & Methven, 1991 (QED)	Average for conservation test	0.39	Yes	Statistically significant	15
McCutchen et al., 2002 (QED)	Gates-MacGinitie Word Reading Subtest	0.39	No	Statistically significant	15
McGill-Franzen et al., 1999 (RCT)	Concepts about print	1.11	Yes	Statistically significant	37
	Letter identification	0.69	Yes	Statistically significant	25
	Writing vocabulary	0.32	Yes	Not significant, but substantively important	13
	Ohio Word Test	0.66	Yes	Not significant, but substantively important	24
	Hearing sounds in words	0.97	Yes	Statistically significant	33
Saxe et al., 2001 (QED)	Peabody Picture Vocabulary Test	0.12	None applied if author did not report significant results	Not significant	5
	Fraction concepts	2.39	No	Statistically significant	49
Sloan, 1993 (RCT)	Fraction computation	-0.53	No	Not significant, but substantively important	-20
	Comprehensive Test of Basic Skills, reading	0.68	Yes	Not significant, but substantively important	25
	Comprehensive Test of Basic Skills, mathematics	0.26	None applied if author did not report significant results	Not significant, but substantively important	10
Tienken, 2003 (RCT with group equivalence problems)	Comprehensive Test of Basic Skills, science	0.63	Yes	Not significant, but substantively important	23
	Content/organization score on narrative writing test	0.41	Yes	Not significant, but substantively important	16
<i>Average effect size across all studies:</i>		<i>0.54</i>	<i>Average improvement index across all studies:</i>		<i>27</i>
<i>Minimum effect size across all studies:</i>		<i>-0.53</i>	<i>Minimum improvement index across all studies:</i>		<i>-20</i>
<i>Maximum effect size across all studies:</i>		<i>2.39</i>	<i>Maximum improvement index across all studies:</i>		<i>49</i>

RCT is a randomized controlled trial; QED is a quasi-experimental design.

Source: Authors' calculations based on data described in text.

Table 3: Features of the professional development in the nine studies that meet evidence standards

Study (study design)	Name or type of professional development	Philosophy	Content	Provider and delivery	Contact hours and duration	Kennedy content group
Carpenter et al., 1989 (randomized controlled trial)	Cognitively guided instruction	Giving teachers knowledge from research on students' thinking and learning about mathematics changes teachers' teaching. It also improves how teachers assess student knowledge, which in turn changes teachers' instruction.	How students learn math, relationships between math problems and how students process to solve them, research on math acquisition, examination of curricula, how materials affect teaching, planning instruction	Authors Four-week workshop and one follow-up meeting	83 hours over four months	Group 4: Focused on how students learn and how to assess student learning
Cole, 1992 (randomized controlled trial)	Mississippi Teacher Assessment Instrument staff development	Teachers who use state-defined competencies will teach better, and therefore their students will learn more.	Modeling of the 14 Mississippi Teacher Assessment Instrument teacher (pedagogical) behavior competencies (for example, planning instruction to achieve selected objectives, organizing instruction to take into account individual differences among learners, and obtaining and using information about the needs and progress of individual learners)	Mississippi State Department of Education Eight three-hour sessions over a two-month period with follow-up observational visits throughout the year, plus two half day follow-up conferences	40+ hours over a year	Group 1: Focused on teaching behaviors applying generically to all subjects
Duffy et al., 1986 (randomized controlled trial)	Incorporating explicit verbal explanations during reading instruction	Training teachers in the use of explicit verbal explanations during reading instruction to poor readers will increase student awareness of what was taught, which in turn will enhance students' strategic reading skills.	How to recast teacher skill at prescriptive basal text techniques into strategies for helping students be better readers when removing blockages to meanings; how to make explicit statements about the reading skills being taught; how to organize these statements for presentation to students	Authors using Houghton-Mifflin basal text Five two-hour sessions	10 hours over four months	Group 2: Focused on teaching behaviors applying to a particular subject

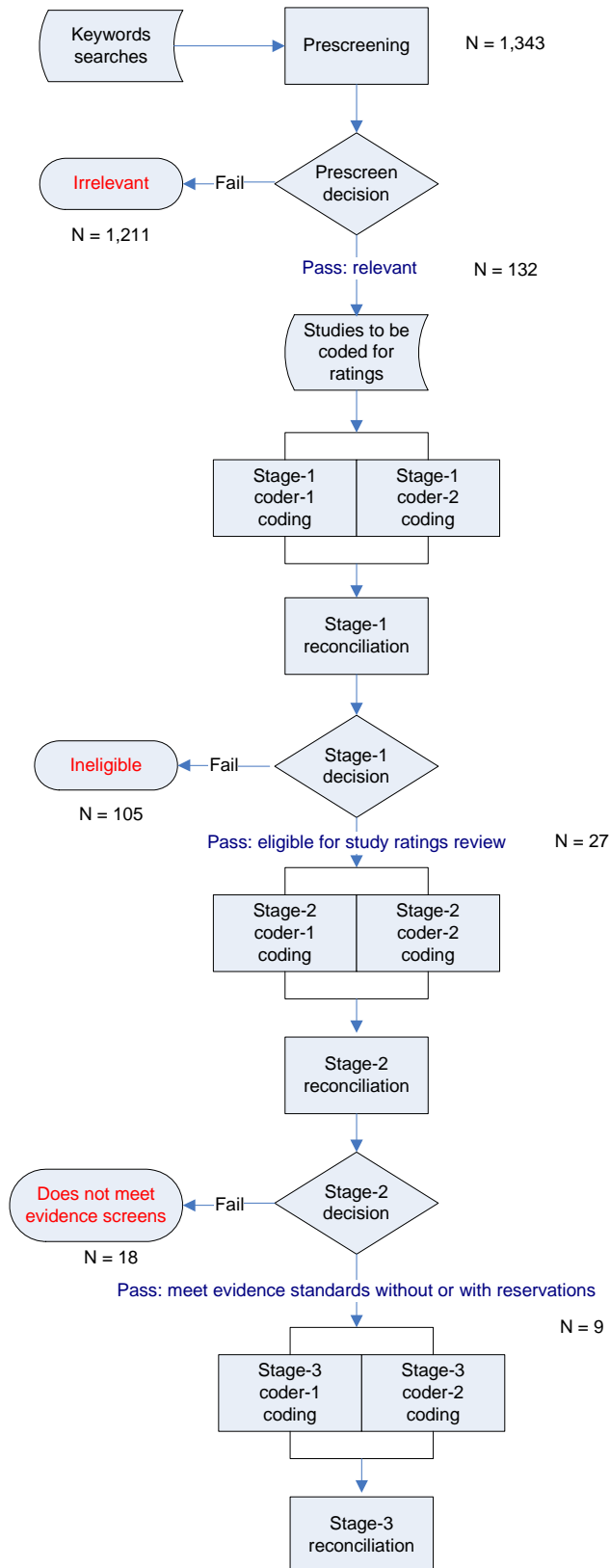
Systematic review of evidence

Study (study design)	Name or type of professional development	Philosophy	Content	Provider and delivery	Contact hours and duration	Kennedy content group
Marek & Methven, 1991 (quasi-experimental design)	Utilizing the Learning Cycle in Elementary School Science	Teaching science as a search for knowledge will lead students to construct their own knowledge about the world around them.	How to develop a curriculum (learning cycles) that represents science, allows students to experience science as a search for knowledge, and is compatible with their students' learning abilities.	National Science Foundation-funded workshop delivered by the Science Education Center at the University of Oklahoma Four-week summer workshop	100 hours over four weeks	Group 2: Focused on teaching behaviors applying to a particular subject
McCutchen et al., 2002 (quasi-experimental design)	n/a	Teachers should incorporate phonological awareness instruction into classroom practices. A deep understanding of phonology, pronunciation, reading skill development, and the links among them must be used in the classroom.	Deepening teachers' understanding of phonology, phonological awareness, analysis of sounds, development of phonological awareness in children, children's mistakes revealing underlying conception of phonemics.	University research team Two-week summer institute plus three follow-up meetings; informal interactions and classroom visits with support	About 100 hours over 10 months	Group 4: Focused on how students learn and how to assess student learning
McGill-Franzen et al., 1999 (randomized controlled trial)	n/a	Improving children's access to books in their classrooms is not enough to develop literacy among kindergartners. It must be supplemented by enhancing their teachers' instructional routines involving the book collection.	Physical design of the classroom; effective book displays; importance of reading aloud to children; environmental print; author, genre, and content themes created with the book collection; small-group lessons using teacher-made materials based on books read	Authors Three whole-day sessions and seven two-hour follow-up sessions	About 30 hours over six months	Group 3: Focused on curriculum and pedagogy, justified by how students learn

Systematic review of evidence

Study (study design)	Name or type of professional development	Philosophy	Content	Provider and delivery	Contact hours and duration	Kennedy content group
Saxe et al., 2001 (quasi-experimental design)	n/a	Although good curriculum materials can provide rich tasks and activities that support students' mathematical investigations, such materials may not be sufficient to enable deep changes in instructional practice. With professional development, teachers must transform the ways they use curriculum materials with students.	Teacher knowledge of mathematics (particularly fractions), teacher knowledge of how students learn mathematics and fractions, and teacher understanding of student motivation in math	Authors/ university-based developer A weeklong summer workshop with 13 follow-up meetings	About 60 hours over six and a half months	Group 4: Focused on how students learn and how to assess student learning
Sloan, 1993 (randomized controlled trial)	n/a	Training teachers to exhibit behaviors related to Direct Instruction using Hunter's (1984) Seven Steps of the Teaching Act will lead to global changes in teachers' instructional and questioning behaviors, which in turn will improve student learning in various subjects.	Use of instructional and questioning strategies associated with Direct Instruction and Hunter's (1984) Seven Steps of the Teaching Act (for example, anticipatory set, objective and purpose, instructional input, modeling, checking for guidance)	Author with district support Summer sessions and seven follow-up meetings	About five hours over two months	Group 1: Focused on teaching behaviors applying generically to all subjects
Tienken, 2003 (randomized controlled trial with group equivalence problems)	n/a	There is a need for focused and sustained professional development in writing instruction. Job-embedded professional development or the "environmental" model of instruction will be effective in training teachers in using rubrics to enhance student self-monitoring and thinking about writing and the writing process.	How to provide instruction to students in the use of the criteria in the New Jersey Registered Holistic Scoring Rubric and a set of high-order reflective questions as self-assessment and reflection devices when composing, revising, and editing narrative essays.	Author Eight one-hour sessions with six follow-up conferences	14 hours over three and a half months	Group 3: Focused on curriculum and pedagogy, justified by how students learn

Appendix A: Flowchart of the systematic review process



Appendix B: Kennedy's professional development content groups

Kennedy's (1998) classification scheme for professional development differentiates between four types.

Group 1—focused on teaching behaviors applying generically to all subjects. These behaviors might result from process–product research or might include strategies like cooperative grouping. The methods are expected to be equally effective across school subjects.

Group 2—focused on teaching behaviors applying to a particular subject. Although presented for a particular subject, the behaviors have a generic quality and are expected to be generally applicable in that subject.

Group 3—focused on curriculum and pedagogy, justified by how students learn. Such professional development provides general guidance on curriculum and pedagogy for teaching a subject and justifies its recommendations using knowledge about how students learn the subject.

Group 4—focused on how students learn and how to assess student learning. Such professional development provides knowledge about how students learn particular subjects but does not provide specific guidance on practices for teaching the subject

