What Works in Professional Development?

A research synthesis confirms the difficulty of translating professional development into student achievement gains despite the intuitive and logical connection. Those responsible for planning and implementing professional development must learn how to critically assess and evaluate the effectiveness of what they do.

THOMAS R. GUSKEY AND KWANG SUK YOON

■ *THOMAS R. GUSKEY is distinguished service professor at Georgetown College, Georgetown, Kentucky. KWANG SUK YOON is a research scientist for the American Institutes for Research, Washington, D.C.*

Image: Liquid Library

What do we really know about the relationship between professional development and improvements in student learning? What evidence validates that relationship, and how trustworthy is that evidence? What does that evidence tell us about the characteristics of truly effective professional development activities?

These questions guided one of the largest and most inclusive syntheses of research on effective professional development conducted to date. Scholars from the American Institutes for Research analyzed findings from over 1,300 studies that potentially address the effect of professional development on student learning outcomes. 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.

The findings from this comprehensive analysis, titled *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement* (Yoon et al. 2007), shed new light on the complex relationship between professional development and improvements in student learning. It is hoped that they also will lead to new and better research on this vitally important dimension of the educational improvement

process. (Details of the research are available in the online version of this article.)

One of the most discouraging findings in the project was the discovery that only nine of the original list of 1,343 studies met the standards of credible evidence set by the What Works Clearinghouse, the arm of the U.S. Department of Education that is charged with providing educators, policy makers, researchers, and the public with scientific evidence about "what works" in education. All nine studies focused on elementary schools and were conducted between 1986 and 2003. No studies of professional development at the middle school or high school levels met the WWC standards, nor did any of the studies published between 2004 and 2006.

Workshops are not the poster child of ineffective practice that they are often made out to be.

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. Among the achievement measures considered, seven studies used standardized assessments of achievement, one involved researcher-developed measures of students' knowledge of fractions, and one used Piagetian conservation tasks. The number of teachers involved in these studies ranged from five to 44; the number of students from 98 to 779.

What Was Learned

Researchers reviewed these nine well-designed investigations to determine whether the professional development efforts on which they focused shared common elements or characteristics. They noted that information about the professional development described in the studies was far from perfect and varied in its quality and effect. In addition, given an initial pool of more than 1,300 citations, the nine studies represent a relatively modest research base. Nevertheless, several common elements emerged. These shared characteristics were not what many would have guessed, and several differ from those factors frequently noted as contributing to the effectiveness of professional development endeavors (Guskey 2003).

Workshops. Of all professional development activities, none has been more disparaged in recent years than workshops, particularly those of short duration. Criticized as the epitome of ineffective practice, many education leaders regard workshops as a waste of both time and money. And, indeed, a lot of workshops are wasteful, especially the one-shot variety that offers no genuine follow-up or sustained support. But ironically, *all* of the studies that showed a positive relationship between professional development and improvements in student learning involved workshops or summer institutes. These workshops focused on the implementation of research-based instructional practices, involved active-learning experiences for participants, and provided teachers with opportunities to adapt the practices to their unique classroom situations. So while undoubtedly many workshops are poorly organized and focus on unproven ideas and strategies, as a form of professional development, they are not the poster child of ineffective practice that they are often made out to be.

Outside Experts. Many writers in education today stress that professional development should be strictly site-based and should build on the combined expertise of in-house staff members. They believe that the most effective way to bring improvement is to have educators in each school meet regularly to explore common problems and seek solutions based on shared experiences and collective wisdom. But while this may be an appropriate starting point, it is seldom, if ever, sufficient (Holloway 2000; Latham 1998). A review by Thomas Corcoran and his colleagues found, for example, that when decisions about professional development were primarily school-based, "school staff members paid lip service to the use of research" and "were more interested in designs that drew on research about practices that they already felt were 'good' than in designs that were producing results." According to these researchers, "the decentralization of decision making appear[s] to be undermining the use of knowledge rather than promoting it" (2001, p. 81).

In the current analysis, the professional development efforts that brought improvements in student learning focused principally on ideas gained through the involvement of outside experts. These individuals were either program authors or researchers who presented ideas directly to teachers and then helped facilitate implementation. None of the successful efforts used a train-the-trainer approach, peer coaching, collaborative problem solving, or other forms of school-based professional learning. This does not imply that these practices are ineffective. Rather, it simply points out that at the present time, we have no strong, valid, and scientifically defensible evidence

demonstrating that they are effective.

Time. Professional development advocates have long lamented the lack of sufficient time for staff members to engage in high-quality professional learning. Obviously, educators need time to deepen their understanding, analyze students' work, and develop new approaches to instruction. But simply providing more time for professional development yields no benefit if that time is not used wisely. An analysis by Mary Kennedy (1998) showed, in fact, that differences in the time spent in professional development activities were unrelated to improvements in student outcomes. Why? Presumably because doing ineffective things longer does not make them any better.

No improvement effort has ever succeeded in the absence of thoughtfully planned and well-implemented professional development.

In this analysis, time was found to be a crucial factor to success. While the number of contact hours ranged widely, from five to over 100 hours depending on the study, those initiatives that showed positive effects included 30 or more contact hours. It thus seems clear that effective professional development requires considerable time, and that time must be well organized, carefully structured, purposefully directed, and focused on content or pedagogy or both (Birman et al. 2000; Garet et al. 2001; Guskey 1999).

Follow-up. For decades professional development experts have stressed the importance of follow-up activities. Educators at all levels need just-in-time, job-embedded assistance as they struggle to adapt new curricula and new instructional practices to their unique classroom contexts. This analysis confirmed the vital importance of follow-up. Virtually all of the studies that showed positive improvements in student learning included significant amounts of structured and sustained follow-up after the main professional development activities.

Activities. Discussions about "best practices" have dominated professional development circles in recent years. Debates frequently arise from these discussions about what particular professional development activities or designs are most effective and work best (Easton 2004). Yet this analysis of well-designed studies identified no set of common activities or designs linked to effect on student learning outcomes. In each case, the structural features of the professional development activity were determined by the specific content involved, the nature of the work, and the context in which that work took place. This corroborates the position taken by the National Staff Development Council (2001), which argues that the most effective professional development comes not from the implementation of a particular set of "best practices," but from the careful adaptation of varied practices to specific content, process, and context elements.

Content. Equally debated in recent years is what professional development content is most likely to lead to improvements in student learning. The analysis noted considerable consistency regarding this aspect. The nine studies focused on specific subject-related content or pedagogic practices. In addition, most also emphasized teacher discretion in implementing that content and those pedagogic practices, justified by how students learn. In other words, the professional development efforts in every one of these investigations centered directly on enhancing teachers' content knowledge and their pedagogic content knowledge (Shulman 1986). The activities were designed to help teachers better understand both what they teach and how students acquire specific content knowledge and skill.

Interpreting the Findings

Many professional developers are likely to be surprised by these results, and some may be disappointed. Many will be stunned, just as we were, to learn that only nine investigations from a pool of over 1,300 potentially useful citations met the WWC standards for inclusion in the analysis. Obviously, these findings paint a dismal picture of our knowledge about the relationship between professional development and improvements in student learning. Such a paucity of rigorous studies of the impact of professional development on student learning outcomes was corroborated by the recent National Mathematics Advisory Panel's report (2008), which concluded that most studies of professional development in mathematics were descriptive in nature and lacking in the methodological rigor needed to warrant sound causal inferences (e.g., "one-group pretest/posttest designs" without a comparison group).

Nevertheless, these results should not be taken as an indictment of professional development advocates or their work. In the history of education, no improvement effort has every succeeded in the absence of thoughtfully planned and well-implemented pro-

fessional development. This analysis shows simply that sound, trustworthy, and scientifically valid evidence on the specific aspects of professional development that contribute to such improvement is in dreadfully short supply and that dedicated efforts to enhance that body of evidence are sorely needed. Furthermore, this research synthesis confirms the difficulty of linking professional development to specific student achievement gains despite the intuitive and logical connection. It is hoped that a better understanding of what the current evidence reveals will help guide those efforts.

Educators at all levels need just-in-time, job-embedded assistance as they struggle to adapt new curricula and new instructional practices to their unique classroom contexts.

We also want to emphasize that the results from this analysis should *not* be taken to mean that alternative professional development activities and designs — such as coaching, the use of collective internal expertise, different allocations of time, or other types of professional development content — do not work. Rather, the results illustrate that at this time, we simply have no reliable, valid, and scientifically defensible data to show that these strategies do work. The best that can be said is that their true value has yet to be determined.

Some might argue that the "What Works Clearinghouse Evidence Standards" used to select the studies included in this analysis 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 (1998), for example, included a different set of investigations, mostly due to different selection criteria. Using less stringent criteria could have yielded a broader range of effective professional development models, activities, and designs.

In defense of these criteria, however, we would counter that when educators ask what professional development approaches are most likely to lead to improvements in student learning, answers should be based on the most valid and scientifically defensible evidence available. The results from carefully designed, experimental or quasi-experimental studies provide such evidence. Furthermore, if the advocates of alternative professional development models, practices, and designs want their approaches to gain professional credibility and acceptance, then they should take responsibility for demonstrating effectiveness through rigorous and scientifically valid means. In other words, rather than simply appealing to practitioners' intuition and making claims of common sense, take the time to conduct thorough and systematic investigations of the true effects. Doing so will not only establish credibility, it will go far in enhancing the professionalism of our field.

Implications

The implications of this analysis for professional developers are fourfold. First, at all levels of education, those responsible for planning and implementing professional development must learn how to critically assess and evaluate the effectiveness of what they do. This means that discussions about the specific goals of professional development, what evidence best reflects the achievement of those goals, and how that evidence can be gathered in meaningful and scientifically defensible ways must become the starting point for all planning activities (Guskey 2000; Guskey 2001). Only when gathering data on the effectiveness of professional development becomes a central focus in the planning process will the pool of valid and trustworthy evidence expand.

Second, practitioners at all levels must demand better evidence from consultants and purveyors of new strategies and practices. Stories about what happened at one time in a single school or district may be interesting, but they do not justify broader implementation. What we need is trustworthy, verifiable, replicable, and comparative data. In addition, those promoting particular ideas or techniques often preface their comments with the phrase, "Research says . . ." in order to enhance presumed credibility. School-based educators must be prepared to dispute such claims, asking such questions as: "What research?" "When was it conducted?" "Was it done in contexts similar to ours?" "Are the results applicable to our setting?" and "How trustworthy are those results?" Consultants have the responsibility to know that research in sufficient depth to answer these questions. And if they do not, then at least they should have the courage and integrity to say, "I don't know."

Third, implementation of any new professional development strategy should always begin with small-scale, carefully controlled, pilot studies designed to

test its effectiveness. Before embracing any new strategy or committing large amounts of time, money, and other resources to any new approach, that new strategy should be carefully examined in that context to determine if the promised effects in terms of student learning gains can be realized. Comparing the progress of one group of educators engaged in the new approach with that of another, matched group of educators in similar teaching situations can yield important evidence on the likelihood of success. Positive results will enhance the credibility of the new approach and will provide a foundation on which larger scale implementation and evaluation can build. In the absence of positive results, either needed adaptations can be considered or resources can be redirected to other, more promising approaches.

Effective professional development requires considerable time, and that time must be well organized, carefully structured, purposefully directed, and focused on content or pedagogy or both.

Finally, researchers as well as practitioners must pursue greater rigor in the study of professional development. If public schools are spending about $20 billion annually on professional development activities (NCES 2008), then it merits serious study. The research community must dramatically improve the precision of studies of the relationship between professional development, changes in teaching practices, and improvements in student learning. Practitioners likewise should insist on better evidence when making decisions about how to spend their limited professional development resources.

Rigor, however, does not imply that only one method of inquiry is required to produce credible evidence. Although randomized designs (i.e., true experimental studies) represent the gold standard in scientific research, especially in studies of causal effects, a wide range of quasi-experimental designs can produce valid results. When such studies are replicated with similar findings, that validity is further enhanced. Comparing the progress of one group to a similar group that has been "matched" on relevant measures, for example, can be especially useful if data are available on pertinent background characteristics of the participating teachers and their students. Randomly selecting half of those who volunteer to take part in a new approach and then comparing their results with those from the other half who were not included but will be next year also can offer valuable information. In addition, other investigative methods may be used to formulate important research questions and develop new measures relating to professional growth (Raudenbush 2005).

The amount of valid and scientifically defensible evidence we currently have on the relationship between professional development and improvements in student learning is exceptionally modest. Nine studies from an initial group of 1,343 potentially relevant citations represent a very small percentage. Given this limited number of studies, we also have to be cautious about making a definitive conclusion about the effectiveness of specific elements of professional development. This conservative stance is echoed by the National Mathematics Advisory Panel. It concluded that "Although the Panel did find some positive effects of PD 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 PD" (2008, p. 40).

Still, we are now in a better position than ever before to organize and conduct professional development so that valid evidence can be gathered, both to determine the effectiveness of

current practice and to inform future endeavors. In addition, several large-scale, randomized studies of the impact of professional development on student learning funded by the Institute of Education Sciences are now under way to answer questions that could not be answered in this analysis. Efforts are also being made to improve the rigor of studies specifically designed to examine this important relationship (Wayne et al. 2008). Moving in this direction will improve the likelihood of success and also elevate professional development to an inquiry-based profession, rather than a haphazard set of activities based on intuition, hearsay, tradition, and folklore. ■

Those responsible for planning and implementing professional development must learn how to critically assess and evaluate the effectiveness of what they do.

REFERENCES

Birman, Beatrice F., Laura Desimone, Andrew C. Porter, and Michael S. Garet. "Designing Professional Development That Works." *Educational Leadership* 57 (May 2000): 28-33.

Corcoran, Thomas B., Susan H. Fuhrman, and C.L. Belcher. "The District Role in Instructional Improvement." *Phi Delta Kappan* 83 (September 2001): 78-84.

Easton, Lois B., ed. *Powerful Designs for Professional Learning.* Oxford, Ohio: National Staff Development Council, 2004.

Garet, Michael S., Andrew C. Porter, Laura Desimone, Beatrice F. Birman, and Kwang Suk Yoon. "What Makes Professional Development Effective? Results from a National Sample of Teachers." *American Educational Research Journal* 38 (Winter 2001): 915-945.

Guskey, Thomas R. "Apply Time with Wisdom." *Journal of Staff Development* 20 (Spring 1999): 10-15.

Guskey, Thomas R. *Evaluating Professional Development.* Thousand Oaks, Calif.: Corwin Press, 2000.

Guskey, Thomas R. "The Backward Approach." *JSD* 22 (Summer 2001): 60.

Guskey, Thomas R. "What Makes Professional Development Effective?" *Phi Delta Kappan* 84 (June 2003): 748-750.

Holloway, John H. "The Promise and Pitfalls of Site-Based Management." *Educational Leadership* 57 (October 2000): 81-82.

Kennedy, Mary M. *Form and Substance in Inservice Teachers Education*. Research Monograph No. 13. Madison, Wis.: National Institute for Science Education, University of Wisconsin-Madison, 1998.

Latham, Andrew S. "Site-Based Management: Is It Working?" *Educational Leadership* 55 (April 1998): 85-86.

National Center for Educational Statistics. *Revenues and Expenditures for Public Elementary and Secondary Education: School Year 2005-06 (Fiscal Year 2006)*. Washington, D.C.: Institute for Education Sciences, U.S. Department of Education, 2008. http://nces.ed.gov/pubs2008/expenditures.

National Mathematics Advisory Panel. *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*. Washington, D.C.: U.S. Department of Education, 2008.

National Staff Development Council (NSDC). *NSDC's Standards for Staff Development*. Oxford, Ohio: National Staff Development Council, 2001.

Raudenbush, Stephen W. "Learning from Attempts to Improve Schooling: The Contribution of Methodological Diversity." *Educational Researcher* 34 (February 2005): 25-31.

Shulman, Lee S. "Those Who Understand: Knowledge Growth in Teaching." *Educational Researcher* 15 (October 1986): 4-14.

Wayne, Andrew, Kwang Suk Yoon, Stephanie Cronen, Pei Zhu, and Michael S. Garet. "Experimenting with Teacher Professional Development: Motives and Methods." *Educational Researcher* 37 (May 2008): 469-479.

Yoon, Kwang Suk, Teresa Duncan, Silvia Wen-Yu Lee, Beth Scarloss, and Kathy L. Shapley. *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement*. Issues and Answers Report, REL 2007 – No. 033. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest, 2007. http://ies.ed.gov/ncee/edlabs.

DESIGN OF THE SYNTHESIS

This broad research synthesis 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. A deliberately wide net of keywords was used to capture literature 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 review. These studies were then subjected to three stages of coding.

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.

The results of this stage of coding yielded 27 relevant studies that were eligible for review in terms of study quality ratings.

Stage 2 coding focused on quality ratings of the 27 eligible studies using the U.S. Department of Education's What Works Clearinghouse (WWC) Evidence Standards (see http://ies.ed.gov/ncee/wwc/overview/review.asp?ag=pi). 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 evidence of causal validity).
- "Meets Evidence Standards with Reservations" (e.g., quasi-experimental studies or controlled trials that had problems with randomization, attrition, teacher-intervention confound, or disruption).
- "Does Not Meet Evidence Standards" (e.g., studies that did not provide strong evidence of causal validity).

Only nine of the 27 studies were rated at the first or second level as having met the WWC Evidence Standards. The other 18 studies were rated at the third level: "Does Not Meet Evidence Standards."

Descriptive Results

The next step in the analysis was to review the selected studies for shared descriptive characteristics. Among the nine studies that met the What Works Clearinghouse Evidence Standards for causal validity, six were published in peer-reviewed journals, while three were unpublished doctoral dissertations. All of the studies focused on elementary schools and were conducted between 1986 and 2003. No studies of professional development conducted at the middle school or high school levels met the standards, nor did any of the studies published more recently, between 2004 and 2006.

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. Among the achievement measures considered, seven studies used standardized assessments of achievement, one involved researcher-developed measures of students' knowledge of fractions, and one used Piagetian conservation tasks. The number of teachers involved in these studies ranged from five to 44, the number of students from 98 to 779.

Twenty different effect sizes were computed across the nine studies, ranging from -.53 to +2.4. Eighteen of these effect sizes were positive, one was zero, and another was negative but not statistically significant. Eight of the 20 effect sizes proved statistically significant, and 12 were not. But among those 12, nine would be considered substantively important accord-

ing to What Works Clearinghouse conventions.

Analytic Results

Following the descriptive analysis, the researchers reviewed these well-designed investigations to determine whether or not the professional development efforts on which they focused shared common elements or characteristics. They noted that information about the professional development activities described in the studies was far from perfect and varied in its quality and effect. In addition, given an initial pool of more than 1,300 citations that were found in electronic literature searches to be linked to the keywords of professional development and improvements in student learning, the nine studies that met the guidelines of causal validity represent a relatively modest research base. Nevertheless, several common elements emerged from the research synthesis. Surprisingly, these shared characteristics were not what many would have guessed, and several differ from the factors frequently noted as contributing to the effectiveness of professional development endeavors.