

Published: September 13, 2006

Professional Development: How Do We Know If It Works?

By Pendred Noyce

Knowledgeable, skillful teachers form the bedrock of good schools. We all accept the notion that teachers who know and can do more in the classroom help students learn more. So it's no surprise that funders, educators, and researchers all tout professional development. Teaching teachers appeals to our intuition as a high-leverage strategy for boosting student achievement. But professional development is expensive to provide, hard to find time for, and difficult to do well. Worse, we have very little empirical evidence about how—or even whether—it works. As research, teaching, and funding communities, we need to start holding ourselves to a much higher standard of evidence about the effectiveness *for enhancing student learning* of professional-development interventions that we support.

Moving to this standard—judging professional development by its effect on student learning—does not mean that we should abruptly call a halt to existing professional development, label it ineffective, and instead embrace scripted curricula that expect little of teachers. Nor does it mean that every professional-development program, *forever*, will require an expensive research component. What it does mean is that for the next few years, funders and policymakers should make it a priority to build a research base about how and when our investments in professional development work.

School districts spend a lot of money on professional development, often much more than they realize. This spending is scattered among budget categories and funded from multiple sources, many of them outside the district. In a study of five urban districts, [Karen Hawley Miles](#) and her colleagues found that spending on professional development—including teacher time, coaching, materials, facilities, tuition, and fees—ranged from 2 percent to more than 5 percent of total district expenditures,

averaging more than \$4,000 per teacher. Extrapolated to the nation as a whole, these figures suggest that we spend \$5 billion to \$12 billion on professional development each year.

We must acknowledge first that the reason we do professional development is so that students will learn more. Every other outcome is important only insofar as it leads to that.

But there are additional costs. Teachers attending professional development miss days in front of students. Some of the strongest teachers are lifted out of the classroom to become coaches. And the dollar numbers cited don't begin to address the cost of steps on the salary scale that teachers earn for participating in graduate courses. So shouldn't we do everything we can to find out if this spending is making a difference?

Numerous groups of thoughtful experts have come up with checklists and guidelines for what constitutes high-quality teacher learning. The consensus documents stress a number of features: Professional development should build teacher content knowledge. It should align with the standards and curricula teachers are expected to teach. It should enhance teachers' knowledge of teaching practice, both theoretical and practical. It should be ongoing and job-embedded, not brief and external. It should include teacher collaboration, build leadership, and reflect teachers' input about their needs. It should help teachers use data, perhaps by examining student work.

There's nothing wrong with consensus documents, as far as they go. But they don't go far enough. In medicine, expert consensus is considered the lowest level of support for a particular practice, only to be used where no strong experimental evidence exists to help guide a decision.

Recently, the Noyce Foundation tried to do a study of effective professional development in mathematics and literacy. We canvassed practitioners and experts from around the country for nominations of programs that met the consensus criteria for quality and showed preliminary evidence of impact on student achievement. From multiple nominees, we identified as candidates eight strong professional-development programs, and asked their providers to select a best-case site for us to investigate.

Our finding? None of the eight sites, strong as they were, was engaged in the kind of data collection that would allow a convincing conclusion that professional development had made a broad, educationally significant difference to student learning. Teacher satisfaction and generally rising student scores are not enough—not when other schools also have rising scores, not when tests are changing, not when nobody is keeping track of which teachers received what professional development, not when nobody knows for certain which teachers are actually taking what they've learned back into the classroom.

Consumers and providers of professional development both know that feel-good statistics are inadequate. But they point to the many barriers and complications that make measuring the impact of professional development so daunting. There is such variation between teachers to begin with. There are so many school and district factors—changes in leadership, unstable curricula, mobility—that can interfere with good implementation of newly learned techniques. External accountability systems may not line up with what teachers are trying to do differently. And countless uncontrolled factors affect a student's learning—everything from family problems to museum visits to a toothache on test day.

There is no question that evaluating the impact of professional development is more complex than evaluating the effectiveness of a new drug or medical treatment. There are more intervening steps. Instead of being administered directly to the patient (student), the treatment is administered to an intermediary, the teacher. And broad educational outcomes are difficult to discern. It's easier to measure the number of new heart attacks than to measure gains in the understanding of rational numbers, or greater disposition to persist in solving mathematical problems. But hard is not impossible.

A new standard should be in place for funders, researchers, and district decisionmakers to use in deciding what professional development to support. First and most important is to acknowledge that *the reason we do professional development is so that students will learn more*. Every other outcome—building teacher leadership and satisfaction, increasing teacher content knowledge or confidence, decreasing teacher turnover, building a community of learners,

improving the quality of teacher questioning—is important only insofar as it leads to improved student learning. The standard should be that every professional-development intervention that does not already rest on a robust research base should include a strong, statistically defensible plan for letting all the actors know within a reasonable time (perhaps two or three years) whether, because of the intervention, students are learning more.

What might it mean to apply this standard in practice? How should we build the body of evidence? What should funders ask for, and researchers and educators expect to do? For starters, the following:

- *Demonstrate that the target intervention works in the best of circumstances.* Before launching a wide-scale professional-development effort, it's important to show that, if you get teachers to change their practice as envisioned, students will learn more. For example, if you propose teaching middle school teachers more about proportion or classroom discourse in the hope that they will teach more effectively, first demonstrate that the way you want them to teach works well for students. This could be accomplished with a small pilot intervention and assessment in advance of major, districtwide training.
- *Select a stable assessment regimen* for pre- and post-testing of the larger student group. This may include instruments specific to the intervention, such as portfolios, performance assessments, or specially designed tests. Try to include instruments that will be convincing to a general audience. Keep in mind that lack of continuity in testing is one of the biggest barriers to finding out whether students really are doing better. If your state test might change significantly during the period of your intervention, invest in another measure.
- *Assign treatment and control groups prospectively.* Randomizing students, teachers, classrooms, or schools to intervention and control groups makes for the strongest design, but is often not possible. Second best is matching intervention and control groups on a range of the most important variables known to affect student performance. Sometimes it is difficult to withhold a treatment from any group of teachers or students. In that case, a staggered implementation may suffice, for example by starting a third of the district schools in the program each year for three

years. Such a deliberate rollout may have other benefits, too, by allowing you to improve the offering each year.

If concurrent controls are unavailable, use historical controls, both longitudinally—by looking at achievement-growth trajectories of past and current students—and cross-sectionally, by looking at how teachers' past 6th grade students did compared to the 6th graders they teach after professional development.

- *Keep careful records of teacher participation.* Even when a professional-development program is aimed at all teachers, not all of them attend every session, receive the same amount of coaching, or join the same study groups. To measure the impact of the intervention, you will need to know the dose of intervention that different teachers receive.

- *Link student records to teacher records.* It does no good to track what dose and kind of professional development each teacher receives unless you know which students that teacher affects.

- *Think carefully about time frame.* If you are providing professional development to a teacher over the course of the year, when do you expect the teacher to begin doing things differently in the classroom? When do you expect to see student performance begin to change? Be sure you don't stop measuring just at the point you might reasonably begin to see student effects.

Evaluating the impact of professional development is more complex than evaluating the effectiveness of a new drug or medical treatment. There are more intervening steps.

- *Measure student gains* vs. control students, vs. historical controls, and vs. the general background—that is, what is happening to student performance in similar schools and districts in the state.

- *Monitor change in teacher knowledge and action.* Although teacher change logically precedes student change, from a practical standpoint, it is of secondary importance. Teacher change tells us how, not whether, professional development is having the impact we want.

To build our understanding of how professional development works, we need to find out what teachers are actually learning and what they are doing differently in the classroom. This may be difficult and expensive to measure. Asking them how they have changed their practice may not be enough. Teachers regularly overestimate the degree to which they are implementing a recommended or desired practice. Direct assessment of teacher understanding and observation of changed practice in a subset of classrooms will give a more complete picture.

This list may seem daunting. Carrying it out will require close collaboration between researchers or evaluators and district personnel. Planning for the impact study needs to begin at the same time as planning for the intervention itself. Studying impact will also require money. A reasonable rule of thumb might be that 10 percent of the initial cost (say, for the first three years) of a major intervention should go to seeing that we find out whether and how it works. Funders and practitioners should treat this expenditure as an opportunity and an insurance policy, to protect against a gnawing uncertainty, several years into a major investment, about whether the program they have promoted so assiduously is truly having the intended effect.

If we can hold ourselves to this standard—insisting that for a period of time each major new professional-development intervention include a careful study of its impact on student learning—we will finally be able to advance our common understanding of how to help teachers become more effective in the classroom. Professional development will begin to fulfill its promise. The payoff will come not only in increased teacher productivity and satisfaction, but also, and more importantly, in increased student learning and growth.

Pendred Noyce, a physician by training, has been active in school reform efforts nationally and in her home state of Massachusetts, where she has taken part in National Science Foundation-funded studies of mathematics, science, and technology education and been a principal investigator for the Massachusetts Parent Involvement Project. She serves on the boards of many education-related organizations, and has been a trustee of the Noyce Foundation, based in Palo Alto, Calif., since its inception in 1990.