



**FOCUS ON KEY POINTS TO DEVELOP THE BEST STRATEGY TO EVALUATE PROFESSIONAL LEARNING**

**By Thomas R. Guskey**

A few years ago, I learned an important lesson about gathering evidence on the outcomes of professional learning endeavors. Several colleagues and I were asked to evaluate the effects of a new program that had been implemented in elementary schools throughout an entire state. The program involved extensive professional learning for the educators responsible for implementation and was quite costly. We collected information on student achievement, affect, and behavior from four years before

implementing the program and three years following. We also gathered data on teachers' recommendations of students for special services and student disciplinary actions during that time.

Comparing year-to-year results based on hundreds of students, we found no significant improvement on any measure of student learning. Some measures actually showed declines. According to the evidence gathered, the new program had produced no overall benefits in student achievement, affect, or behavior, despite its high cost.

We presented the results of our evaluation to a legislative committee, using colorful charts and graphs to show that the program had not yielded the promised improve-

ments. When the committee chair asked for questions or comments following our presentation, one committee member responded, “All this may be true, but ... ” and went on to relate a touching story about her niece, a shy and quiet young girl who had struggled terribly in school. She was embarrassed by her learning difficulties and felt belittled by her classmates. Following implementation of this program, according to the committee member, her niece gained new confidence, interest, and enthusiasm for learning. She was now doing better than ever, seemed excited about learning, and loved going to school each day, much to the surprise and delight of her parents.

Later that day, the committee voted to continue funding the program for two more years.

The lesson I learned that day is this: When gathering evidence on outcomes, one must always consider the perspective of stakeholders. Both the committee member and we presented detailed evidence on program outcomes. But in the end, an impassioned story about one particular child carried more weight than did impersonal charts and graphs based on data from hundreds of children. The committee member’s story was evidence that the other members of the committee trusted and believed.

Gathering evidence on the outcomes of any professional learning experience can be a challenging and complicated task. It involves consideration of a wide variety of perceptual and contextual issues, some obvious to education leaders and others not. Those who want to succeed in this process may find the following points helpful.

#### **ALWAYS BEGIN WITH THE OUTCOMES.**



Many educators consider outcomes to be synonymous with results or consequences. But if viewed in a broader context, outcomes also may be seen as goals or aspirations. They describe what we hope to accomplish and set forth the criteria by which success will be judged. In this sense, outcomes are not something to consider only at the end when activities are completed. Instead, they must be where we begin planning all professional learning endeavors (Guskey, 2005, 2007a).

As Covey (2004) reminded us, we must always “begin with the end in mind.” Before thinking about the content or format of any professional learning, planners must first

consider the outcomes they hope to accomplish. This requires addressing two essential questions: What outcomes do we want to achieve, especially with regard to student learning, and what evidence best reflects the achievement of those outcomes? These two questions should mark the starting point in all planning discussions.

Deciding what outcomes we want to influence typically involves careful analysis of current data on student learning. Results from large-scale assessments, common formative assessments (Ainsworth & Viegut, 2006), and individual classroom assessments can show areas where students might be struggling or not performing as well as hoped. School records can identify behavior problems related to attendance or discipline. Classroom observations and discussions with students often help pinpoint areas of concern. Interviews with teachers, focus groups, or discussions in professional learning communities (DuFour, 2004) are especially valuable when trying to identify persistent trouble spots in efforts to help all students succeed in mastering complex concepts and skills.

Analyzing the performance of subgroups of students can bring additional insights to these discussions. Considering the learning progress of students of different backgrounds and ability levels, language experiences, ethnicity, race, and gender can be particularly informative. Looking at differences between classrooms and between schools often yields new understandings of problem areas as well.

The key point in these discussions is to ensure that the focus remains on student learning outcomes. Because of concerns about professional learning processes, conversations often skip to the content and activities in which participating educators will be involved. We begin debating new ideas, strategies, innovations, programs, and instructional technologies. While these are important issues, remember that they are means to an important end that must be determined first. After deciding the specific desired outcomes with regard to student learning, decisions about the most appropriate means will be much easier to make.

**Gathering evidence on the outcomes of any professional learning experience can be a challenging and complicated task. It involves consideration of a wide variety of perceptual and contextual issues.**



**DIFFERENT STAKEHOLDERS TRUST DIFFERENT EVIDENCE.**

Addressing the second essential question — what evidence best reflects achievement of those outcomes? — can prove just as thorny. One might think after reaching consensus about outcomes that it would be easy to decide what evidence to gather on those outcomes. But the validity and believability of different sources of evidence varies among stakeholders.

A few years ago, I conducted a study asking groups of educators in three states to judge the validity of various sources of evidence on student achievement (Guskey, 2007b). All three states had implemented comprehensive, statewide assessment programs with high-stakes consequences for educators and students. The educators ranked 15 indicators of student learning based on which they believed “provides the most trustworthy information about students’ academic performance” (p. 19). Evidence included teacher observations, portfolios of student work, grades, and scores on state assessments and nationally normed standardized assessments. I then compared the rankings of school administrators to those of teachers.

Results showed that, while educators generally held similar views, the perspectives of administrators and teachers differed significantly. Administrators perceived nationally normed standardized assessments, state assessments, and district assessments to be more valid indicators of student achievement than did teachers. In contrast, teachers granted more validity to classroom assessment results, classroom observations, homework completion and quality, and students’ class participation and behavior than did administrators.

Even when planners agree on the student learning goals or outcomes of professional learning endeavors, different stakeholders may not agree on what evidence best reflects improvement in those outcomes. And just as was true in the story of our presentation to the legislative committee, the trust stakeholders place in that evidence can impact their interpretations of effectiveness and their subsequent decisions based on those interpretations.



**USE MULTIPLE SOURCES OF EVIDENCE.**

As Learning Forward’s Data standard states, “Professional learning that increases educator effectiveness and results for all students uses a variety of sources and types of student, educator, and system data to plan, assess, and evaluate professional learning” (Learning Forward, 2011). Since stakeholders vary in their trust of different sources of evidence, it is unlikely that

any single indicator of success will prove adequate or sufficient to all. Providing acceptable evidence for judging the effects of professional learning will therefore require multiple sources of evidence. In addition, these sources of evidence must be carefully matched to the needs and perceptions of different stakeholder groups.

Take the testimonial offered by the committee member at the start of this article. From a technical perspective, testimonials are a very poor source of evidence. They are highly subjective, often unreliable, and based on a limited sample of observations, but they can be emotionally stirring, personally compelling, and extremely influential. The story told by the committee member turned out to be the kind of evidence that the other committee members understood, trusted, and believed.

It would be inappropriate to use testimonials as the only source of evidence in evaluating the effectiveness of professional learning endeavors. But as one of several sources of evidence, testimonials offer a rich, powerful, and personalized account that should never be ignored.

Results from large-scale state assessments and nationally normed standardized exams may be important for accountability purposes and need to be included. School administrators generally consider these to be valid indicators of success. Teachers, however, see limitations in large-scale assessment results. These assessments are generally administered only once a year, and results may not be available until several months later. By that time, the school year may have ended and students promoted to another teacher’s class. So, while these assessments are important, many teachers do not find such results particularly useful.

Teachers put more trust in results from their own assessments of student learning: classroom assessments, common formative assessments, and portfolios of student work. They turn to these sources of evidence for feedback to determine if the new strategies or practices they are implementing really make a difference. Classroom assessments provide timely, targeted, and instructionally relevant information that also can be used to plan revisions when needed. Since teachers comprise a major stakeholder group in any professional learning, sources of evidence that they trust and believe will be particularly important to include.

Finally, while evidence on student academic achievement will always be essential, affective and behavioral indicators of student performance can be relevant as well. These include student surveys designed to measure how much students like school, their perceptions of teachers, fellow students, and themselves, their sense of self-efficacy, and their confidence in new learning situations. Evidence on school attendance, enrollment patterns, dropout rates, class disruptions, and disciplinary actions are also important. In some areas, parents’ or families’ perceptions may be a vital consideration. This is especially true in initiatives that involve changes in grading practices, report

**Teachers put more trust in results from their own assessments of student learning: classroom assessments, common formative assessments, and portfolios of student work.**

cards, or other aspects of school-to-home and home-to-school communication (Epstein & Associates, 2009; Guskey, 2002).



#### HOW EVIDENCE IS GATHERED IS JUST AS IMPORTANT AS THE EVIDENCE ITSELF.

Gathering evidence needs to be an explicit and transparent process. Just as students should never be surprised by the evidence used to evaluate their performance, educators should not be surprised by the evidence selected to measure the outcomes of their professional learning. Not only should they know what those sources of evidence will be, they should have a voice in choosing them.

The best way to ensure transparency is to address questions about what evidence to gather during the initial planning process. Deciding at the start what evidence best reflects a particular outcome brings purpose and direction to professional learning. Involving different stakeholders in deciding what evidence to use and in gathering that evidence further guarantees results will be seen as credible and trustworthy. It also reinforces the idea that improvement is an ongoing process that requires input and collaboration among all stakeholders.



#### PLAN FOR COMPARISONS.

In many cases, evidence on outcomes is gathered from a single school or school district in a single setting for a restricted time period. Unfortunately, from a design perspective, such evidence lacks reliability and validity. Whether results are positive or negative, so many alternative explanations may account for the results that most authorities would consider such outcomes dubious at best and meaningless at worst (Guskey & Yoon, 2009).

It may be, for example, that the planned professional learning did lead to noted improvements. But maybe the improvements were the result of a change in leadership or personnel instead. Maybe the community or student population changed. Maybe changes in state policies or assessments made a difference. Maybe other simultaneously implemented interventions were responsible. The possibility that these or other extraneous factors influenced results makes it impossible to draw definitive conclusions.

The best way to counter these threats to the validity of results is to include a comparison group — a similar group of educators or schools not involved in the current activity or perhaps engaged in a different activity. Ideal comparisons involve the random assignment of students, teachers, or schools to different groups. Because that is rarely possible in most education settings, finding similar classrooms, schools, or school districts is the next best option. In some cases, involvement in professional learning can be staggered so that half of the teachers or schools that volunteer can be randomly selected to take part initially, while the others delay involvement and serve as the comparison

group. In other cases, comparisons can be made to “matched” classrooms, schools, or school districts that share similar characteristics related to motivation, size, and demographics.

Using comparison groups does not eliminate the effects of extraneous factors that might influence results. It simply allows planners greater confidence in attributing the results attained to the particular program or activity being considered.

#### FOCUS ON OUTCOMES AND EVIDENCE

Just as we urge teachers to become more purposeful in planning instructional activities, we need to become more purposeful in planning professional learning. We must determine up front what improvements we seek in terms of student learning and what evidence best reflects that improvement to the satisfaction of all stakeholders involved. That will improve the likelihood of our success and yield more valid evidence on the effectiveness of current activities while informing future professional learning.

#### REFERENCES

- Ainsworth, L. & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction and assessment*. Thousand Oaks, CA: Corwin Press.
- Covey, S.R. (2004). *The 7 habits of highly effective people*. New York, NY: Free Press.
- DuFour, R. (2004, May). What is a “professional learning community”? *Educational Leadership*, 61(8), 6-11.
- Epstein, J.L. & Associates. (2009). *School, family, and community partnerships: Your handbook for action* (3rd ed.). Thousand Oaks, CA: Corwin Press.
- Guskey, T.R. (2002). *How's my kid doing? A parent's guide to grades, marks, and report cards*. San Francisco, CA: Jossey-Bass.
- Guskey, T.R. (2005, Winter). Taking a second look at accountability. *JSD*, 26(1), 10-18.
- Guskey, T.R. (2007a). Leadership in the age of accountability. *Educational Horizons*, 86(1), 29-34.
- Guskey, T.R. (2007b). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice*, 26(1), 19-27.
- Guskey, T.R. & Yoon, K.S. (2009, March). What works in professional development? *Phi Delta Kappan*, 90(7), 495-500.
- Learning Forward. (2011). *Standards for Professional Learning*. Oxford, OH: Author.

•  
**Thomas R. Guskey (guskey@uky.edu) is professor of educational psychology in the College of Education, University of Kentucky. ■**