BY THOMAS R. GUSKEY

My article, "The Age of Our Accountability", was first published in the Fall 1998 Journal of Staff Development. We were just entering a new era of accountability. Today and for the foreseeable future, accountability is our way of life. The No Child Left Behind Act (U.S. Congress, 2001) demands that only strategies and methods "proven effective by the standard of scientifically based research should be included in school reform programs" and that programs be evaluated based on specific measures of student learning (U.S. Department of Education, 2002, p. 2).

Surprisingly, some professional development leaders still seem oblivious to the demands of accountability. In their innocence, they persuade us that their professional development programs are state-of-the-art efforts that turn teachers and school administrators into reflective, brain-based, consensus-seeking, multiple-intelligence capacity builders. They remain largely confident that what they do brings priceless benefits to students, teachers, parents, board members, and
the community at large.

Slowly but surely, however, most are recognizing that without strong evidence reflecting those benefits, their very existence may be in jeopardy. A new president, governor, state legislator, superintendent, or board member might come along who wants to know the payoff from the district’s or school’s investment in professional development. If the evidence isn’t there, heads may roll and programs may get axed. To provide evidence in this age of accountability, we must get serious about evaluation (Guskey, 2002a).

Let’s revisit professional development evaluation in the context of accountability by considering three basic questions:
- What is evaluation?
- What are the purposes of evaluation?
- What are the critical levels of professional development evaluation?

Finally, we examine the implications of these questions.

WHAT IS EVALUATION?

Just as there are many forms of professional development, there also are many forms of evaluation. While not everyone agrees on the best definition of evaluation, a useful operational definition is: Evaluation is the systematic investigation of merit or worth (adapted from the Joint Committee on Standards for Educational Evaluation, 1994).

The word “systematic” in this definition distinguishes the evaluation process from the multitude of informal evaluation acts in which we consciously engage every day. “Systematic” implies that evaluation in this context is thoughtful, intentional, and purposeful. It’s done for clear reasons and with explicit intent. Although its specific purpose may vary from one setting to another, all good evaluations are organized and deliberate.

Because it’s systematic, some educators mistakenly believe that professional development evaluation is...
They demand an appraisal of quality and judgments of value, based on the best evidence available.

WHAT ARE THE PURPOSES OF EVALUATION?

The purposes of evaluation generally are classified into three broad categories: planning, formative, and summative. Most evaluations fulfill all three purposes, although the emphasis on each changes in various stages of the evaluation process. Although this blending of purposes blurs their distinction, differentiation their intent helps clarify our understanding of evaluation procedures (Stevens, Lawrence, & Schtyaj, 1999).

Planning
Planning evaluation occurs before a program or activity begins, although certain aspects may be continual and ongoing. It involves those involved in program development and implementation to develop an understanding of what the program is supposed to accomplish, what procedures will be used, and how success will be determined. Planning evaluation lays the groundwork for all other evaluation activities. While some advocate an "evaluation" assessment prior to planning to determine if a program or activity is "evaluable" (Wholey, Harty, & Newcomer, 2004), most experts contend that planning evaluation alone will not make such assessment unnecessary.

Planning evaluation involves appraising a program or activity's critical attributes, usually on the basis of previously established standards. These include the specified goals, the proposal or plan to achieve those goals, the concept or theory underlying the proposal (a critical attribute in meeting NCLB requirements), the overall evaluation plan, and the likelihood that the plan can be carried out with the time and resources available. In addition, planning evaluation typically includes determining needs, assessing the characteristics of participants, carefully analyzing the context, and gathering relevant baseline information.

Evaluation for planning purposes is sometimes referred to as "formative evaluation" (Stevens, 1993) and may be thought of as prescriptive evaluation. It helps decision makers know if efforts are headed in the right direction and likely to produce the desired results. Planning evaluation also helps identify and remedy difficulties that might plague later evaluation efforts. Furthermore, evaluation during the planning stage helps ensure that other evaluation purposes can be met in an efficient and timely manner.

Formative
Formative evaluation occurs during the operations of a program or activity. It gives those responsible for the program ongoing information about whether the work is going as planned and whether expected progress is being made. If progress is not as expected, formative evaluation information helps guide necessary improvements (Stevens, 1967). The most useful formative evaluations focus on the conditions for success. They address issues such as:

1. What conditions are necessary for success?
2. Have the conditions for success been met?
3. Can conditions be improved?

In many cases, formative evaluation is a recurring process that takes place many times throughout the program. Many program developers, in fact, constantly engage in formative evaluation and use the evidence they gather at each step of development and implementation to adjust, modify, or revise the program (Freidreich, Sanders, & Wren, 2003).

To keep formative evaluations effective, Stevens (1993) recommends using them as "early warning" evaluative...
Summative evaluation takes place at the completion of a program or activity. It provides program developers and decision makers information with which to make judgments about the program's overall merit or worth. Summative evaluations describe what was accomplished, what the consequences were (positive and negative), what the final results were (intended and unintended), and, in some cases, whether the benefits justify the costs (Phillips, 2002).

Unlike formative evaluations that are used to guide improvements, summative evaluations present decision makers with information to make crucial decisions about a program. Should it be continued? Continued with modifications? Expanded? Discontinued? The focus of summative evaluation is the bottom line.

An excellent description of the distinction between formative and summative evaluations is offered by Robert Stake: "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative." (quoted in Seriven, 1991, p. 169).

Unfortunately, many educators associate evaluation only with its summative purposes. Important information that could help guide planning, development, and implementation is often neglected, even though such information can be key in helping a program succeed.

Summative evaluation, although necessary, comes too late to be of much help. Thus, while the relative emphasis on planning, formative, and summative evaluation changes through the life of a program or activity, all three are essential to a meaningful evaluation.

WHAT ARE THE LEVELS OF PROFESSIONAL DEVELOPMENT EVALUATION?

Planning, formative, and summative evaluation all involve collecting and analyzing information. Effective professional development evaluation requires combining the five critical stages or levels of information shown on page 14 (Guise, 2000a, 2002a).

These five levels are adapted from an evaluation model developed by Kirkpatrick (1959, 1998) for judging the value of supervisory training programs in business and industry. Kirkpatrick's model, although widely applied, has been limited use in education because of inadequate explanatory power. While helpful in addressing a broad range of "what" questions, such as, "What took place?" and "What were the results?", many find it lacking for answering "why" questions, such as, "Why were practices not implemented?" or "Why were improvements not evidenced?" (Alliger & Janak, 1989; Holton, 1990).

The five levels in this model are arranged hierarchically, from simple to more complex. With each succeeding level, the process of gathering evaluation information requires more time and resources. And because each level builds on the levels before, success at one level is necessary to succeed at higher levels. (See chart on Page 14.)

Level 1: Participants' reactions

The first level of evaluation considers participants' reactions to the professional development experience. Assessing reactions is the most common form of professional development evaluation and the easiest type of information to gather and analyze. At Level 1, the questions focus on "whether". These questions ask, "Did they feel their time was well spent? Did the material make sense to them? Were the activities well-planned and meaningful? Was the leader knowledgeable and helpful? Did they find the information useful?"

Also important for some professional development experiences are questions related to the context, such as, "Was the room the right temperature? Were the chairs comfortable? Was the coffee hot and ready on time? Were the refreshments fresh and tasty? To some, questions such as these may seem silly and inconsequential. But experienced professional developers know the importance of attending to these basic human needs.

Information on participants' reactions usually is gathered through questionnaires or surveys handed out at the end of a program or activity. These questionnaires typically include a combination of rating-scale items and open-ended response questions that allow participants to provide comments. Because of the general nature of this information, many organizations use the same questionnaire for professional development activities, regardless of the format.

Some educators refer to these measures of participants' reactions as "happiness questionnaires," inquiring that they reveal only the entertainment value of an activity, not its quality or worth. But measuring participants' initial satisfaction provides information that can help improve the design and delivery of programs or activities in valid ways. In addition, positive reactions from participants usually are a necessary prerequisite to higher level evaluation results.
## Five levels of professional development evaluation

<table>
<thead>
<tr>
<th>EVALUATION LEVEL</th>
<th>WHAT QUESTIONS ARE ADDRESSED?</th>
<th>HOW WILL INFORMATION BE GATHERED?</th>
<th>WHAT IS MEASURED OR ASSESSED?</th>
<th>HOW WILL INFORMATION BE USED?</th>
</tr>
</thead>
</table>
| **1**  
**PARTICIPANTS’ REACTIONS** | • Did they like it?  
• Was their time well spent?  
• Did the material make sense?  
• Will it be useful?  
• Was the leader knowledgeable and helpful?  
• Were the refreshments both and tasty?  
• Was the room the right temperature?  
• Were the chairs comfortable? | • Questionnaires or surveys administered at the end of the session | • Perceptual satisfaction with the experience | • To improve program design and delivery |
| **2**  
**PARTICIPANTS’ LEARNING** | • Did participants acquire the intended knowledge and skills? | • Paper and pencil instruments  
• Simulations  
• Demonstrations  
• Participant reflections (oral and/or written)  
• Participant portfolios | • New knowledge and skills of participants | • To improve program content, format, and organization |
| **3**  
**ORGANIZATIONAL SUPPORT AND CHANGE** | • Were sufficient resources made available?  
• Were problems addressed quickly and efficiently?  
• Was implementation advocated, facilitated, and supported?  
• Were successes recognized and shared?  
• Was the support public and overt?  
• What was the impact on the organization?  
• Did it affect organizational climate and procedures? | • Minutes from follow-up meetings  
• Questionnaires  
• Structured interviews with participants and district or school administrators  
• District and school records  
• Participant portfolios | • The organization’s advocacy, support, accommodation, facilitation, and recognition | • To document and improve organizational support;  
• To inform future change efforts |
| **4**  
**PARTICIPANTS’ USE OF NEW KNOWLEDGE AND SKILLS** | • Did participants effectively apply the new knowledge and skills? | • Questionnaires  
• Structured interviews with participants and their supervisors  
• Participant reflections (oral and/or written)  
• Participant portfolios  
• Direct observances  
• Video or audiotapes | • Degree and quality of implementation | • To document and improve the implementation of program content |
| **5**  
**STUDENT LEARNING OUTCOMES** | • What was the impact on student learning?  
• Did student performance improve?  
• Did student attendance improve?  
• Are dropout rates decreasing?  
• Did it influence students’ physical or emotional well-being?  
• Are students more confident as learners? | • Student records  
• School records  
• Questionnaires  
• Structured interviews with students, parents, teachers, and/or administrators  
• Participant portfolios | • Student learning outcomes  
• Cognitive (performances and achievement),  
• Affective (attitudes and dispositions)  
• Psychomotor (skills and behaviors) | • To focus and improve all aspects of program design, implementation, and follow-up  
• To document the overall impact of professional development |

In addition to liking their professional development experience, participants also should learn something from it. Level 2 focuses on measuring the new knowledge, skills, and perhaps attitudes that participants gained (Guskey, 2002b). Depending on the goals of the program or activity, an evaluation of learning can involve anything from a pencil-and-paper assessment to a simulation or full-scale skill demonstration. Presented with a variety of classroom conflicts, can participants diagnose each situation and then prescribe and carry out a fair and workable solution? Participants also might be asked to document their learning through oral or written personal reflections or through portfolio entries.

Evaluators sometimes can gather Level 2 information after the session is completed, but evaluations typically require more than a standardized form. And because these evaluations typically focus on specific learning goals, evaluators must work before the sessions begin to identify what indicators provide the best evidence that participants have successfully learned the material. If program developers have concerns that participants may already have the requisite knowledge and skills, some form of pre- and post-assessment may be necessary.

Careful evaluators also consider the possibility of unintended learning outcomes, both positive and negative. On the positive side, for example, professional development activities that engage educators in collaboration also can foster a positive sense of community and shared purpose among participants (Supsists, 2002). But in some instances, individual collaboration to block change or inhibit advancement (Little, 1996). Recent investigations further show that collaborative efforts sometimes run headlong into enormous conflicts over professional beliefs and practices, conflicts that can impede progress (Ashbychenko, 2002). Even the best planned professional development occasionally yields unintended negative consequences. Without information about such unintended outcomes, program developers and decision makers may make inaccurate judgments of a program’s merit or worth. Analyzing Level 2 information provides a basis for improving the content, format, and organization of professional development programs and activities.

LEVEL 3: Organizational support and change

At Level 3, the focus of the evaluation shifts from participants to organizational variables that are central to the success of the professional development effort. Did the professional activities promote changes that were aligned with the mission of the school and district? Were changes as the individual level encouraged and supported at all levels? (Conceron, Furtman, & Belcher, 2001)? Were sufficient resources made available, including time for sharing and reflection? (Langer & Colvin, 1994)? Were successes recognized and shared? Issues such as these play a large part in the success of any professional development effort.

Organizational factors also can hinder or prevent success, even when individual aspects of professional development are done right (Sparks, 1996). Support, for example, that a group of high school educators participate in a professional development program on cooperative learning. Group members gain a thorough understanding of the theory and organize a variety of classroom activities based on cooperative learning principles. After their training, they try implementing these activities in schools where students are graded “on the curve,” according to their relative standing among classmates, and where great importance is attached to selecting the class valedictorian.

Organizational policies and practices such as these make learning highly competitive and thwart the most valuable efforts to have students cooperate and help each other learn (Guskey, 2006b).

The lack of positive results in this case does not reflect poor training or inadequate learning, but rather organizational policies that are incompatible with implementation efforts. Problems at Level 3 essentially can cancel the gains made at Levels 1 and 2 (Sparks & Hibsh, 1997). Addressing such conflicts is a true test of professional development; evaluations must include information on organizational support and change.

Procedures for gathering information at Level 3 depend on the goals of the program or activity. They may involve analyzing districts or school records, examining the minutes from follow-up meetings, administering questionnaires that tap issues related to the organization’s advocacy, support, accommodation, facilitation, and recognition change efforts. Structured interviews with participants and district or school administrators also can be helpful. This information is used not only to document and improve organizational support, but also to inform future change measures.

LEVEL 4: Participants’ use of new knowledge and skills

Level 4 evaluations include questions such as: Did the new knowledge
and skills that participants learned make a difference in their professional practice? The key to gathering relevant information at this level lies in specifying clear indicators of both the degree and quality of implementation. Unlike at Levels 1 and 2, this information cannot be gathered at the end of a professional development session. Enough time must pass to allow participants to adapt to new ideas and practices to their settings. Because implementation is often a gradual and uneven process, measures of progress may need to be gathered at several time intervals.

Depending on the goals of the program or activity, this information may involve questionnaires or structured interviews with participants and their supervisors. Evaluators might consider participants’ oral or written personal reflections and also might examine participants’ journals and supervisors’ observations. The most accurate information typically comes from direct observations, either by trained observers or using video and audio recordings. These observations, however, should be kept as unobtrusive as possible (for example, see Hall & Ford, 1987).

Analyzing this information gives evaluators evidence about the extent to which participants are using the new knowledge and helps professional development leaders structure follow-up programs and activities to facilitate improved and more consistent implementation.

LEVEL 5: Student learning outcomes

Level 5 addresses the bottom line in education: What impact did the professional development have on students? Did educators’ professional learning benefit students in any way? The student learning outcomes of interest will depend, of course, on the goals of that specific professional development program or activity. In addition to the stated goals, the program or activity may result in important unintended outcomes. Suppose, for example, that students’ scores on statewide assessments went up, but so did the dropout rate. Such examples are why evaluators always should include measures of a broad array of student learning outcomes (Joyce, 1993).

Measures of student learning typically include cognitive indicators of performance and achievement, such as assessment results, portfolio evaluations, marks or grades, and scores from standardized tests. In addition, affective measures (attitudes and dispositions) and psychomotor outcomes (skills and behaviors) may be considered as well. Examples include students’ self-concepts, study habits, school attendance, homework completion rates, and classroom behaviors. Schoolwide indicators such as enrollment in advanced classes, membership in honor societies, participation in school-related activities, disciplinary actions, and attendance or dropout rates might also be considered.

Student and school records provide the majority of such information. Results from questionnaires and structured interviews with students, parents, teachers, and administrators could be included as well. Evaluators use Level 5 information summarily to document a program’s or activity’s overall impact. This formatively, this information can help guide improvements in all aspects of professional development, including design, implementation, and follow-up. In some cases, information on student learning outcomes is used to examine the true effectiveness of professional development, sometimes referred to as “return on investments,” or “ROI evaluations” (Parry 1996; Phillips, 1997; Teislins & Warner, 1993).

IMPLICATIONS FOR IMPROVEMENT

Three important implications stem from this model for evaluating professional development. First, each of the five levels is important. Although evaluation at any level can be done well, particularly or laughingly, the information gathered at each level provides vital data for professional development leaders to improve the quality of professional learning programs and activities.

Second, tracking the program’s effectiveness at one level tells you nothing about the impact of the program at the next. Although success at an early level may be necessary for positive results at the next level, earlier success on its own is not sufficient to cause higher-level success (Cody & Gouskey, 1997). Breakdowns can occur at any point along the way. Sadly, few educators and policy makers recognize the difficulty involved in moving from professional development experiences (Level 1) to improved student learning (Level 5). Most are also aware of the complexity of this process, not of the time and effort required to build this connection (Gouskey, 1997; Gouskey & Sparks, 2004).

The third implication, and perhaps the most important, is that in planning professional development to improve student learning, the goals of these levels must be targeted. In other words, educators must plan backward (Gouskey, 2001a & b, 2002a), starting with the final goals and then working backward.

In backward planning, one first considers the student learning outcomes to be achieved (Level 5). For example, the goal to improve students’ reading comprehension, enhance their skills in problem solving, develop their sense of confidence in learning situations, improve their behavior in class, or persistence in school, or their collaboration with classmates? Critical analyses of rele
DECISIONS MADE AT EACH LEVEL PROFOUNDLY AFFECT THOSE MADE AT THE NEXT.

The next step is to determine, on the basis of pertinent research, what instructional practices and policies will most effectively and efficiently produce these outcomes (Level 4). Key questions to ask are: What evidence verifies that these particular practices and policies will lead to the desired results? How reliable is that evidence? Does the evidence meet the requirements for being "scientifically based"? Was it gathered in contexts similar to ours? By particularly mindful of innovations that are more opinion-based than research-based, promoted by people more concerned with what sells than with what works. Before jumping on any educational Bandwagon, always make sure that trustworthy evidence validates the selected approach.

Next, consider what aspects of organizational support need to be in place for those practices and policies to be implemented (Level 3). Many valuable improvement efforts fail miserably due to a lack of active participation from school administrators (Guirky, 2003b). Others prove ineffective because the resources required for implementation were not provided. The lack of time, personnel, materials, or necessary technology can severely impede teachers' attempts to use the new knowledge and skills they acquired through a professional learning experience.

Sometimes aspects of the organization actually pose barriers in implementation. A big part of planning, therefore, involves ensuring that organizational elements are in place to support the desired practices and policies.

Then decide what knowledge and skills the participating professionals must have to be able to implement the prescribed practices and policies (Level 2). What must they know and be able to do to successfully adapt the innovation to their specific situations and bring about the sought-after change?

Finally, consider what set of experiences will enable participants to acquire the needed knowledge and skills (Level 1). Workshops and seminars, especially when paired with collaborative planning and structured opportunities for practice with feedback, can be highly effective means of sharing information and expanding educators' knowledge. Action research projects, organized study groups, and a wide range of other activities all can be effective, depending on the specified purpose of the professional learning experience.

What makes this backward planning process crucial is that the decisions made at each level profoundly affect those made at the next. For example, the desired student learning outcomes influence the kinds of practices and policies that will be implemented. Likewise, the practices and policies to be implemented influence the kinds of organizational support or change required, and so on.

The context-specific nature of this work further complicates matters. Even if we agree on the student learning outcomes we want to achieve, what works best in one context with a particular community of educators and a particular group of students might not work equally well in another context with different educators and different students. The influence of these varied contexts is what makes developing universal best practices in professional development so difficult. What works always depends on where, when, and with whom.

Regrettably, professional development leaders too often fall into the same trap in planning that teachers do when they plan their lessons. They plan what they are going to do rather than what they want their students to know and be able to do. Their planning tends to be event-based or process-based, not results-based. Planning this way not only diminishes the effectiveness of their efforts, it also makes evaluation much more difficult.

Beginning professional development planning by considering learning goals to be achieved makes planning more efficient and results much easier to evaluate.

CONCLUSION

Many good things are done in the name of professional development. But so are a lot of empty things. What professional development leaders haven't done is provide evidence to document the difference.

Today, in the context of the No Child Left Behind Act and its increased demands for accountability, presenting that evidence is more important than ever.

Evaluation provides the key to distinguishing between the good and the empty. Backward planning provides the foundation for developing meaningful and effective evaluations that demonstrate the difference. Using the five levels of evaluation for both will not only make professional development programs more successful, but also will make evaluation efforts simpler and more purposeful.

REFERENCES


NATIONAL STAFF DEVELOPMENT COUNCIL
research on the Kentucky Education Reform Act (pp. 191-209). Frankfort, KY: The Kentucky Institute for Education Research.


