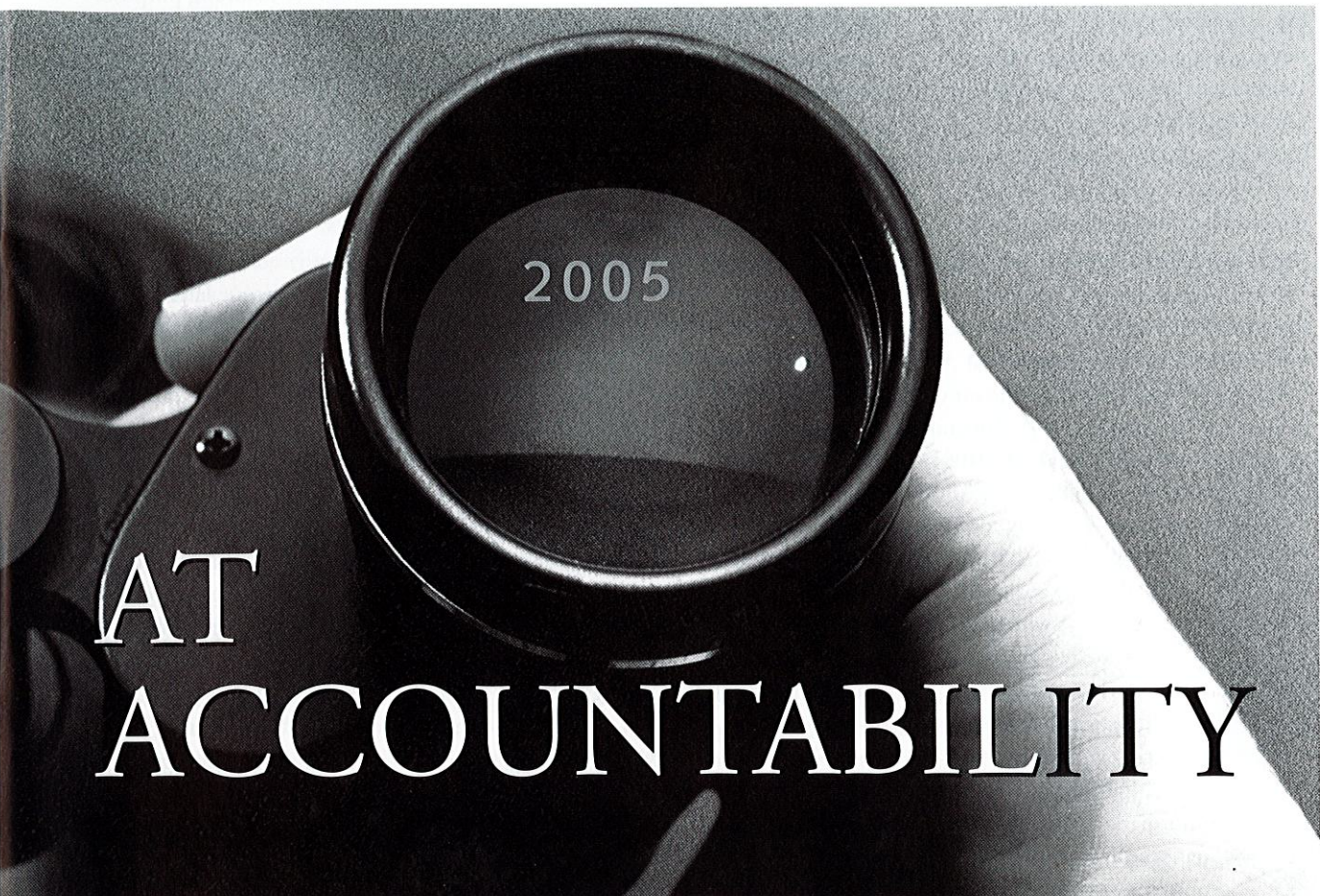


BY THOMAS R. GUSKEY

When 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 presume that their professional development programs are state-of-the-art efforts that turn teachers and school administrators into reflective, brain-based, consensus-seeking, multiple-intelligent capacity builders. They remain naively confident that what they do brings priceless benefits to students, teachers, parents, board members, and

THOMAS R. GUSKEY is a professor of education at the University of Kentucky. You can contact him at the College of Education, Taylor Education Building, University of Kentucky, Lexington, KY 40506, (859) 257-8666, fax (859) 257-4243, e-mail: guskey@uky.edu.



AT ACCOUNTABILITY

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

appropriate only for formal workshops and seminars, but not for the wide range of other less formal, ongoing, job-embedded professional development activities. Regardless of its form, however, professional development is not a haphazard process. Professional learning is, or should be, purposeful and results- or goal-driven (Schmoker, 2004). The objectives of professional development activities remain clear: to make a difference in teaching, to help educators reach high standards, and ultimately to have a positive impact on students. These objectives apply to workshops and seminars, as well as study groups, action research, collaborative planning, curriculum development, structured observations, peer coaching and mentoring, and individually guided professional development activities. Determining whether the goals of these activities are met or whether progress is being made requires systematic evaluation.

If the evidence isn't there, heads may roll and programs may get axed.

“Investigation” means collecting and analyzing appropriate and pertinent information. While no evaluation can be completely objective, the process isn't based on opinion or conjecture. Investigation requires acquiring specific, relevant, and valid evidence examined through appropriate methods and techniques.

Using “merit or worth” in our definition implies appraisal and judgment. Evaluations are designed to determine something's value. They help answer questions such as:

- Is this program or activity leading to the intended results?
- Is it better than activities done in the past?
- Is it better than another competing activity?
- Is the professional learning activity worth the costs?

Answers to these questions require more than a statement of findings.

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 in 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, differentiating their intent helps clarify our understanding of evaluation procedures (Stevens, Lawrenz, & Sharp, 1995).

Planning

Planning evaluation occurs before a program or activity begins, although certain aspects may be continual and ongoing. It gives those involved in program development and implementation a precise 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 “evaluability assessment” prior to planning to determine if a program or activity is “evaluable” (Wholey, Hatry, & Newcomer, 2004), most experts contend that planning evaluation done well makes 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 collecting relevant baseline information.

Evaluation for planning purposes is sometimes referred to as “preformative evaluation” (Scriven, 1991) and may be thought of as preventive 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 operation 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 (Scriven, 1967).

The most useful formative evaluations focus on the conditions for success. They address issues such as:

- What conditions are necessary for success?
- Have the conditions for success been met?
- 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 (Fitzpatrick, Sanders, & Worthen, 2003).

To keep formative evaluations efficient, Scriven (1991) recommends using them as “early warning” evalua-

tions, designed to offer an early version of the final, overall evaluation. As such, formative evaluations typically consider intermediate benchmarks of success to determine what is working as expected and what difficulties must be overcome. They help identify flaws and locate weaknesses in time to make the adaptations necessary for success.

Summative

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 evaluations is the bottom line.

An excellent description of the distinction between formative and summative evaluation 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 Scriven, 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 much help. Thus, while the relative emphasis on planning, formative, and sum-

mative 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 considering the five critical stages or levels of information shown on page 14. (Guskey, 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 seen 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, 1996).

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 participants liked the experience. 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 all professional development activities, regardless of the format.

Some educators refer to these measures of participants' reactions as "happiness quotients," insisting 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.

An excellent description of the distinction between formative and summative evaluation is offered by Robert Stake: "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summative."

Five levels of professional development evaluation

EVALUATION LEVEL	WHAT QUESTIONS ARE ADDRESSED?	HOW WILL INFORMATION BE GATHERED?	WHAT IS MEASURED OR ASSESSED?	HOW WILL INFORMATION BE USED?
1 PARTICIPANTS' REACTIONS	<ul style="list-style-type: none"> • 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 fresh and tasty? • Was the room the right temperature? • Were the chairs comfortable? 	<ul style="list-style-type: none"> • Questionnaires or surveys administered at the end of the session. 	<ul style="list-style-type: none"> • Initial satisfaction with the experience. 	<ul style="list-style-type: none"> • To improve program design and delivery.
2 PARTICIPANTS' LEARNING	<ul style="list-style-type: none"> • Did participants acquire the intended knowledge and skills? 	<ul style="list-style-type: none"> • Paper-and-pencil instruments. • Simulations. • Demonstrations. • Participant reflections (oral and/or written). • Participant portfolios. 	<ul style="list-style-type: none"> • New knowledge and skills of participants. 	<ul style="list-style-type: none"> • To improve program content, format, and organization.
3 ORGANIZATIONAL SUPPORT AND CHANGE	<ul style="list-style-type: none"> • 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? 	<ul style="list-style-type: none"> • Minutes from follow-up meetings. • Questionnaires. • Structured interviews with participants and district or school administrators. • District and school records. • Participant portfolios. 	<ul style="list-style-type: none"> • The organization's advocacy, support, accommodation, facilitation, and recognition. 	<ul style="list-style-type: none"> • To document and improve organizational support. • To inform future change efforts.
4 PARTICIPANTS' USE OF NEW KNOWLEDGE AND SKILLS	<ul style="list-style-type: none"> • Did participants effectively apply the new knowledge and skills? 	<ul style="list-style-type: none"> • Questionnaires. • Structured interviews with participants and their supervisors. • Participant reflections (oral and/or written). • Participant portfolios. • Direct observations. • Video or audiotapes. 	<ul style="list-style-type: none"> • Degree and quality of implementation. 	<ul style="list-style-type: none"> • To document and improve the implementation of program content.
5 STUDENT LEARNING OUTCOMES	<ul style="list-style-type: none"> • What was the impact on students? • Did it affect student performance or achievement? • Did it influence students' physical or emotional well-being? • Are students more confident as learners? • Is student attendance improving? • Are dropouts decreasing? 	<ul style="list-style-type: none"> • Student records. • School records. • Questionnaires. • Structured interviews with students, parents, teachers, and/or administrators. • Participant portfolios. 	<ul style="list-style-type: none"> • Student learning outcomes. • Cognitive (performance and achievement). • Affective (attitudes and dispositions). • Psychomotor (skills and behaviors). 	<ul style="list-style-type: none"> • To focus and improve all aspects of program design, implementation, and follow-up. • To demonstrate the overall impact of professional development.

SOURCE: Adapted from *Evaluating Professional Development* by Thomas R. Guskey (Corwin Press, 2000).

Level 2:**Participants' learning**

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 (Can participants describe the critical attributes of mastery learning and give examples of how these might be applied in common classroom situations?) 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 usually require more than a standardized form. And because the evaluation typically focuses on specific learning goals, evaluators must work before the session begins 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 (Supovitz, 2002). But in some instances, individuals

collaborate to block change or inhibit advancement (Little, 1990). Recent investigations further show that collaborative efforts sometimes run headlong into enormous conflicts over professional beliefs and practices, conflicts that can impede progress (Achinstein, 2002). Thus even the best planned professional development occasionally yields unanticipated 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 vital 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 at the individual level encouraged and supported at all levels (Corcoran, Fuhrman, & Belcher, 2001)? Were sufficient resources made available, including time for sharing and reflection (Langer & Colton, 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). Suppose, for example, that a group of high school educators participates 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 valiant efforts to have students cooperate and help each other learn (Guskey, 2000b).

The lack of positive results in this case doesn't reflect poor training or inadequate learning, but rather organizational policies that are incompatible with implementation efforts. Problems at Level 3 essentially cancelled the gains made at Levels 1 and 2 (Sparks & Hirsh, 1997). Identifying such conflicts is one reason professional development evaluations must include information on organizational support and change.

Procedures for gathering information at Level 3 differ depending on the goals of the program or activity. They may involve analyzing district 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 recognizing 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 initiatives.

LEVEL 4:**Participants' use of new knowledge and skills**

Level 4 evaluations include questions such as: Did the new knowledge

Careful evaluators also consider the possibility of unintended learning outcomes, both positive and negative.

and skills that participants learned make a difference in their professional practice? The key to gathering relevant information at this level rests 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 the 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 struc-

tured interviews with participants and their supervisors. Evaluators might consider participants' oral or written personal reflections and also might examine participants' journals or portfolios. The most accurate information typically comes from direct observations, either by trained observers or using video and/or audiotapes. These observations, however, should be kept as unobtrusive as possible (for examples, see Hall & Hord, 1987).

Analyzing this information gives evaluators evidence about the extent to which participants are using the new knowledge and helps professional development leaders restructure future 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 evaluations 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. But 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, memberships in honor societies, participation in school-related activities, disciplinary actions, and retention 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 summatively to document a program or activity's overall impact. But 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 estimate the cost effectiveness of professional development, sometimes referred to as "return on investment," or "ROI evaluation" (Parry 1996; Phillips, 1997; Todnem & 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 or poorly, convincingly or laughably, the information gathered at each level provides vital data for professional developer 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 & Guskey, 1997). Breakdowns can occur at any point along the way. Sadly, few legislators and policy makers recognize the difficulties involved in moving from professional development experiences (Level 1) to improved student learning (Level 5). Most also aren't aware of the complexity of this process, nor of the time and effort required to build this connection (Guskey, 1997; Guskey & Sparks, 2004).

The third implication, and perhaps the most important, is that in planning professional development to impact student learning, the order of these levels must be reversed. In other words, educators must plan backward (Guskey, 2001a & b, 2003a), starting with the final goals and then working back.

In backward planning, one first considers the student learning outcomes to be achieved (Level 5). For example, is 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, their persistence in school, or their collaboration with classmates? Critical analyses of rele-

vant data from assessments of student learning, examples of student work, and school records are especially useful in identifying these student learning goals.

The next step is to determine, on the basis of pertinent research, what instructional practices and policies will most effectively and efficiently produce those 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? Be 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 and clear support from school administrators (Guskey, 2003b). Others prove ineffective because the resources required for implementation were not provided. The lack of time, instructional 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 to 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 situation 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 so 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 learners’ needs and the specific 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 rotten 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 rotten. 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

- Achinstein, B. (2002, April).** Conflict amid community: The micropolitics of teacher collaboration. *Teachers College Record*, 104(3), 421-455.
- Alliger, G.M. & Janak, E.A. (1989, Summer).** Kirkpatrick’s levels of training criteria: Thirty years later. *Personnel Psychology*, 42(2), 331-342.
- Cody, C.B. & Guskey, T.R. (1997).** Professional development. In J.C. Lindle, J.M. Petrosko, & R.S. Pankratz (Eds.), *1996 Review of*

Decisions made at each level profoundly affect those made at the next.

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

Corcoran, T., Fuhrman, S.H., & Belcher, C.L. (2001, September). The district role in instructional improvement. *Phi Delta Kappan*, 83(1), 78-84.

Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2003). *Program evaluation: Alternative approaches and practical guidelines (3rd ed.)*. Boston: Pearson Education/Allyn & Bacon.

Guskey, T.R. (1997, Spring). Research needs to link professional development and student learning. *Journal of Staff Development*, 18(2), 36-40.

Guskey, T.R. (1998, Fall). The age of our accountability. *Journal of Staff Development*, 19(4), 36-44.

Guskey, T.R. (2000a). *Evaluating professional development*. Thousand Oaks, CA: Corwin.

Guskey, T.R. (2000b, December). Grading policies that work against standards ... and how to fix them. *NASSP Bulletin*, 84(620), 20-29.

Guskey, T.R. (2001a). Backward planning: An outcomes-based strategy for professional development. *Curriculum in Context*, 28(2), 18-20.

Guskey, T.R. (2001b, Summer). The backward approach. *Journal of Staff Development*, 22(3), 60.

Guskey, T.R. (2002a, March). Does it make a difference? Evaluating professional development. *Educational Leadership*, 59(6), 45-51.

Guskey, T.R. (2002b, August). Professional development and teacher change. *Teachers and teaching: Theory and practice*, 8(314), 381-391.

Guskey, T.R. (2003a, Fall). Scooping up meaningful evidence. *JSD*, 24(4), 27-30.

Guskey, T.R. (2003b, January-February). *Designing and evaluating effective professional development*. Presentation to the California Staff

Development Council.

Guskey, T.R. & Sparks, D. (2004). Linking professional development to improvements in student learning. In Guyton, E.M. & Dangel, J.R. (Eds.), *Teacher education yearbook XII: Research linking teacher preparation and student performance*. Dubuque, IA: Kendall/Hunt.

Hall, G.E. & Hord, S.M. (1987). *Change in schools: Facilitating the process*. Albany, NY: SUNY Press.

Holton, E.F. (1996, Spring). The flawed four-level evaluation model. *Human Resources Development Quarterly*, 7(1), 5-21.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards (2nd ed.)*. Thousand Oaks, CA: Sage Publications.

Joyce, B. (1993, Summer). The link is there, but where do we go from here? *Journal of Staff Development*, 14(3), 10-12.

Kirkpatrick, D.L. (1959). Techniques for evaluating training programs [A four-part series beginning in the November issue]. *Journal for the American Society of Training Directors*, 13(11).

Kirkpatrick, D.L. (1998). *Evaluating training programs: The four levels (2nd ed.)*. San Francisco: Berrett-Koehler Publishers.

Langer, G.M. & Colton, A.B. (1994, Winter). Reflective decision making: The cornerstone of school reform. *Journal of Staff Development*, 15(1), 2-7.

Little, J.W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record*, 91(4), 509-536.

Parry, S.B. (1996, May). Measuring training's ROI. *Training & Development*, 50(5), 72-75.

Phillips, J.J. (1997). *Return on investment in training and performance improvement programs*. Houston, TX: Gulf Publishing.

Phillips, P.P. (2002). *The bottom line on ROI: Basics, benefits, & barriers to measuring training & performance improvement*. Atlanta, GA: CEP Press.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.), *Curriculum evaluation. American Educational Research Association Monograph Series on Evaluation, No. 1*. Chicago: Rand McNally.

Scriven, M. (1991). *Evaluation thesaurus (4th ed.)*. Newbury Park, CA: Sage Publications.

Schmoker, M. (2004, February). Tipping point: From feckless reform to substantive instructional improvement. *Phi Delta Kappan*, 85(6), 424-432.

Stevens, F., Lawrenz, F., & Sharp, L. (1995). *User-friendly handbook for project evaluation: Science, mathematics, engineering, and technology education*. Arlington, VA: National Science Foundation.

Sparks, D. (1996, February). Viewing reform from a systems perspective. *The Developer*, pp. 2, 6.

Sparks, D. & Hirsh, S. (1997). *A new vision for staff development*. Alexandria, VA: ASCD and NSDC.

Supovitz, J.A. (2002, December). Developing communities of instructional practice. *Teachers College Record*, 104(8), 1591-1626.

Todnem, G. & Warner, M.P. (1993). Using ROI to assess staff development efforts. *Journal of Staff Development*, 14(3), 32-34.

U.S. Congress. (2001). *No Child Left Behind Act of 2001*. Washington, DC: Author.

U.S. Department of Education. (2002). *Scientifically based research and the Comprehensive School Reform (CSR) Program*. Washington, DC: Author.

Wholey, J.S., Hatry, H.P., & Newcomer, K.E. (Eds.). (2004). *Handbook of practical program evaluation (2nd ed.)*. San Francisco: Jossey-Bass. ■