MARYLAND TEACHER PROFESSIONAL DEVELOPMENT EVALUATION GUIDE

M. Bruce Haslam

October 2008

This guide was prepared under a contract with the Harford County (Maryland) Public Schools. The guidance and observations presented here are those of the author and do not necessarily reflect those of the leaders and staff of the district. No endorsement should be inferred.

Users are encouraged to contact Bruce Haslam at Policy Studies Associates with questions and comments about the guide. He can be reached at bhaslam@policystudies.com or by phone at (202)939-5333.

Acknowledgments

I am pleased to thank Carol Williamson (Superintendent, Queen Anne's County Public Schools), Jeff Maher (Director of Professional and Organizational Development, St. Mary's County Public Schools), Margaret Pfaff (Director of Curriculum, Supervision, and Staff Development, Carroll County Public Schools), and Jackie Tarbert (Professional Development Coordinator, Harford County Public Schools) for their invaluable advice and suggestions at every step in the development of this guide. Their deep interest in teacher professional development and intensely pragmatic perspectives on what is possible and necessary to improve professional learning helped keep the guide grounded in the reality of districts and schools.

A very special thanks to Jackie Haas, Superintendent of the Harford County Public Schools, who has been a very good critical friend of this work.

MBH
Washington D.C.
September 2008

Contents

Introduction

Each year, teachers in Maryland and across the country participate in a wide range of professional development activities, which together reflect substantial investments of time and money. Despite the widespread reliance on professional development as a key component of efforts to improve education for all children, there is little systematic information on the quality of professional development or its contributions to teaching and learning.

Rigorous, ongoing evaluations can help close this gap in several ways:

- Early or formative evaluations gauge teacher satisfaction with professional learning activities, whether the activities took place as planned, and whether teachers mastered new knowledge and skills.
- Formative evaluations help professional development providers and sponsors determine whether the activities are on track or whether some mid-course changes are necessary to achieve the intended outcomes.
- Summative evaluations focus on whether the professional development achieved the intended outcomes as reflected in changes in teachers' professional practice and changes in student learning.
- Ongoing evaluations yield information about changes in school organization and culture that may result from teacher participation in professional development.

This guide is intended to help staff in school district central offices, schools, the Maryland State Department of Education (MSDE), and faculty and staff in institutions of higher education and other professional development providers work together to plan, conduct, and report on evaluations of teacher professional development. The guide complements and should be used in conjunction with the *Maryland Teacher Professional Development Planning Guide.* (Visit www.marylandpublicschools.org to review the planning guide and related materials.)

The guide rests on three assumptions:

- *There is no single "best" approach to evaluation, although the suggestions presented in this guide are applicable to a broad range of professional development.* Evaluations should be tailored to the professional development being evaluated. Evaluation questions, data collection strategies, and reporting will all vary depending on the nature of the activity and the purpose of the evaluation. In addition, the availability of resources (money,

people, time) that can be allocated to the effort will very much influence what can be accomplished.

- *Evaluation planning should be an integral part of professional development planning.* Individuals who will be responsible for the evaluation should be on professional development planning teams from the beginning of the planning process. Planning teams should begin thinking about evaluation as they identify the need for the professional development (Step 1 of the planning process described in the *Maryland Teacher Professional Development Planning Guide)*, identify the intended participants (Step 2 of the planning process), determine the professional learning outcomes and related indicators (Step 3 of the planning process), and determine the kinds of professional learning activities that are most likely to result in the intended outcomes (Step 4 of the planning process).

- *Evaluations of teacher professional development should be separate and distinct from teacher performance appraisals.* Evaluation teams should establish the safeguards necessary to ensure that there is no overlap between data collection and reporting evaluation results and teacher performance appraisals. The safeguards should be clearly visible to teachers and others who are involved in the professional development and subsequent evaluations.

Section I of the guide poses five questions to help shape the evaluation plan. Section II presents suggestions and options for evaluation design and data collection. Section III discusses monitoring data quality and data analysis. Section IV offers guidelines for preparing evaluation reports. Appendix A includes an annotated bibliography of evaluation resources, including data collection instruments. Appendix B includes sample items for surveys of teacher perceptions of their professional learning experiences.

I. Five Questions to Shape the Evaluation Plan

Evaluation planners should answer each of the following questions as they begin planning their evaluation. The answers will define the basic parameters of the evaluation.

1. To evaluate or not to evaluate?

Evaluations should focus on:

- Large-scale professional development activities (those that include large numbers of teachers, extend over relatively long periods of time, and represent significant investments of professional development resources)
- Professional development that is a key component of state, district, or school improvement initiatives
- Pilot professional development activities that, if determined to be successful, will be taken to scale

Despite some potential payoffs, it probably ***does not make sense to evaluate***

- Short-term professional development activities with few participants and no prospects for scale-up
- Professional development activities for which there are limited or unclear expectation for teacher outcomes
- Professional development activities that are not clearly and explicitly aligned with district or school priorities

In most cases, the decision about whether to conduct an evaluation will be fairly easy. Nevertheless, given the cost and time necessary to conduct rigorous evaluations, planning teams should think carefully about whether or not to proceed with an evaluation.

2. What are the key features and guiding assumptions of the professional development that will be evaluated?

The answer to this question emerges as the first four steps of the planning process in the *Maryland Teacher Professional Development Planning Guide* are completed.

If the professional development planning team has not already done so, the evaluation planners should suggest developing a logic model to help identify key components of the professional development, the underlying assumptions, the

timeline, and the expected outcomes. Creating a logic model is especially helpful in planning long-term professional development that includes a variety of professional learning activities. Creating a logic model is equally helpful in planning an evaluation. The logic model is, in effect, the road map for the evaluation.

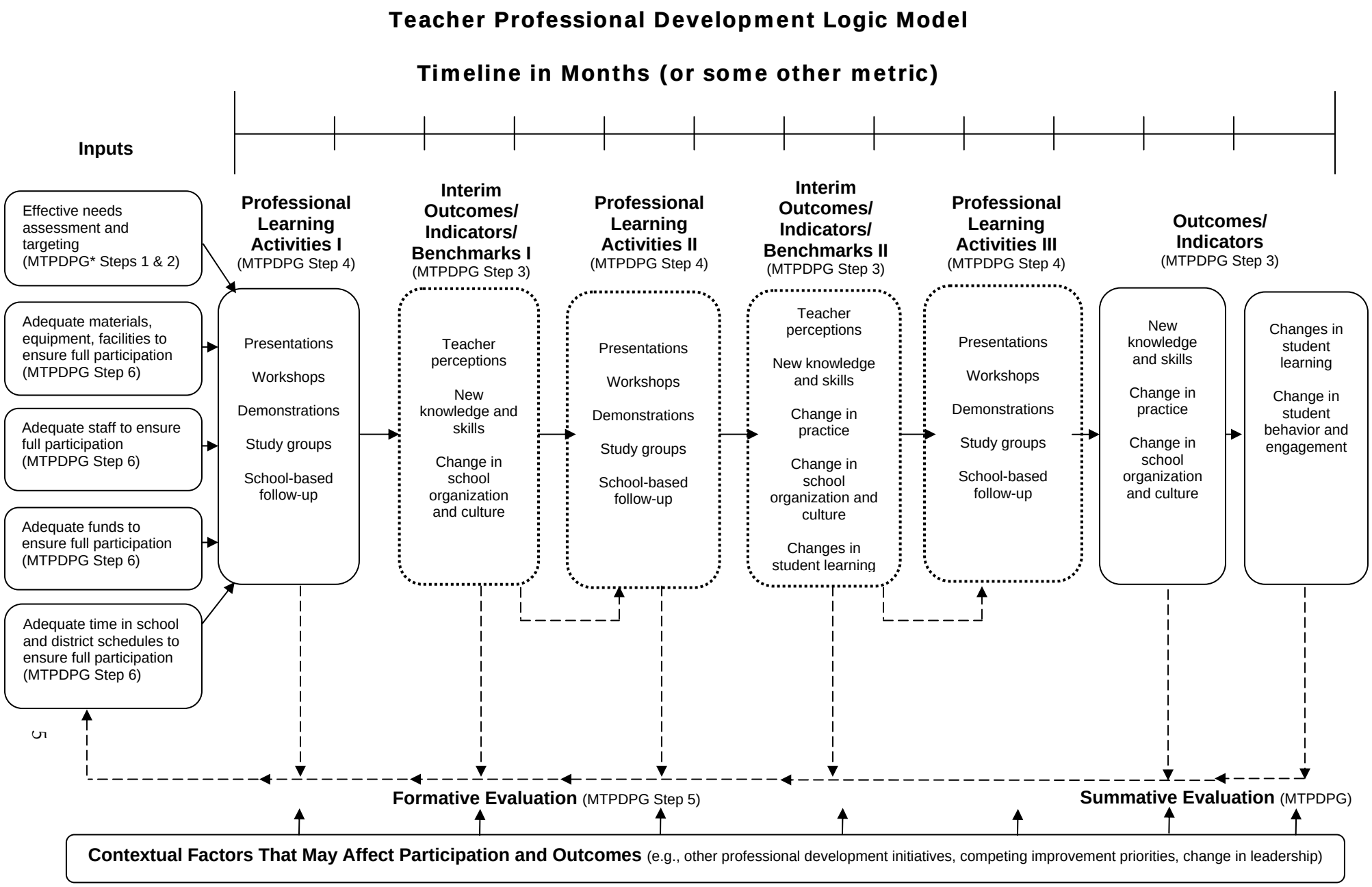
The exhibit on the next page illustrates what a professional development logic model could look like, although each professional development activity requires a logic model that reflects the unique design, underlying assumptions, and expected outcomes and indicators.

Each professional development activity requires a logic model that reflects the unique design, underlying assumptions, and expected outcomes and indicators.

The boxes on the left side of the logic model list the inputs necessary for teacher professional development. Evaluation planners should also recognize that identifying teacher learning needs and who will be targeted for participation in the professional development (Steps 1, 2, and 3 in the planning guide) will go a long way toward determining baselines against which an evaluation can gauge improvements in teacher knowledge, skills, and practice as well as student outcomes.

The inclusion of the three boxes labeled Professional Learning Activities I, II, and III indicates that the professional learning may be ongoing and extend over a number of months or even several years. The activities included in each of the boxes may be different as the professional development unfolds or some may be repeated several times. Note that some professional development may require only one of the learning activity boxes while other activities may require several of these boxes.

The smaller boxes, labeled Interim Outcomes, Indicators, Benchmarks I and II, can include various outcomes, indicators, and benchmarks that planners expect to be able to *observe and/or measure* at different times as the professional development continues. For example, the first set of interim outcomes could include participants' perceptions of the usefulness of the professional development, initial mastery of new knowledge and skills, and changes in school organization to accommodate later classroom applications of new knowledge and skills. As the professional learning activities progress, the second set of interim outcomes could include more extensive mastery and application of new knowledge and skills and an early look at whether changes in student learning are occurring as expected.



**Maryland Teacher Professional Development Planning Guide (MTPDPG)*

The timeline across the top of the logic model not only describes the schedule of the professional development, it also helps determine the schedule for the evaluation. The broken lines across the bottom of the logic model define the overall focus of the evaluation and suggest that the results from early or formative stages of the evaluation can be used to inform modifications to the design of the professional development. Of course, these results can also be used to report on early outcomes to interested stakeholders. The final or summative phase of the evaluation is completed at a reasonable point after the learning activities have taken place and when it is expected to be possible to observe the expected outcomes and related indicators.

The box at the bottom of the logic model hypothesizes that there are many contextual factors that influence the professional development and the outcomes. Planners may not be able to identify all of the relevant contextual factors or predict their influence. Nevertheless, trying to identify them may help avoid problems as the professional development takes place.

A word of caution is in order for planning teams, including the evaluators, as they set outcomes and indicators: The outcomes and indicators should be specific and not overly ambitious. In addition, while the ultimate goal of teacher professional development is to improve student learning, the more immediate goal (as reflected in the outcomes and indicators) is improved teacher knowledge, skills, and practice.

While the ultimate goal of teacher professional development is to improve student learning, the more immediate goal (as reflected in the outcomes and indicators) is improved teacher knowledge, skills, and practice.

It is extremely difficult to empirically establish clear causal relationships between teacher participation in professional development and changes in student learning. This is because of all of the intervening and mediating variables that affect student learning, especially as it is measured by standardized assessments such as the Maryland School Assessment (MSA) and the High School Assessment (HSA). Therefore, planners are well-advised to focus on outcomes for teachers and proximal learning outcomes for students, with the latter being reflected in student work samples and results on locally developed benchmark assessments.

3. Who is likely to be interested in the evaluation and what do they want to know about the professional development?

Potential audiences and their interests include:

- *Teachers* will want to see their feedback on the professional development and its benefits as well as those of their colleagues reflected in the evaluation results.

Five Questions to Shape the Evaluation Plan

- *District professional development staff, especially professional development coordinators, supervisors, and curriculum coordinators,* will want to know whether the activities take place as planned, including whether they attract the intended participants. They will also want to know what the participants think of the activities as a possible indicator of (a) whether they will apply the new knowledge and skills and (b) changes that may be required as the activities moves forward. Later, they will want to know whether the activities achieved the intended outcomes and perhaps whether the evaluation results suggest payoffs in continuing the activity or taking it to scale.

- *Principals and other school leaders* will want to know whether the activities produce the intended changes in teachers' knowledge and skills and whether and when these changes are likely to result in positive changes in student outcomes.

- *Providers, including college and university faculty, consultants, and vendors,* will want to know what participants think of the activities and whether the activities achieve the intended results.

- *Funders and program managers* will want to know the evaluation results, either to satisfy their own reporting requirements or to inform decisions about additional or follow- up funding.

- *District leaders, including school board members,* will join other stakeholders in wanting to know whether the activities achieve the intended changes in teachers' knowledge and skills and, consequently, lead to positive changes in student learning. They might also want to know how these results stack up against the results from other professional development activities.[1]

- *Parent and community groups* will want to know whether and how the activities contribute to changes in instruction and student learning outcomes, especially when those activities require a substantial investment of district resources and/or require teachers to be out of the their classrooms for extended periods.

[1] District leaders are also likely to want to know about the cost of the professional development, especially if the evaluation findings and report suggest that it might be a good idea to continue or expand the activity. Examining spending on professional development is outside the scope of the evaluations described in this guide. The evaluators can, however, examine the budget prepared as part of the planning process (Step 6 in the *Maryland Teacher Professional Development Planning Guide*). The evaluation might also ask whether funds were spent according to the plan and, if not, how the spending varied from what was planned.

Evaluation planners are well-advised to learn (a) what key stakeholders would like to know or need to know about the activities and (b) when they would like to have or need to have the information. It is possible, indeed likely, that the evaluation will not address all stakeholder questions. When information needs exceed what is possible given the resources and staff available for the evaluation, reaching some sort of a compromise about the focus and scope of the evaluation will be necessary.

4. What resources are available to support the evaluation?

Because the resources available to support evaluations are often limited, it is helpful to know what is available and to plan accordingly. Moreover, it is important to recognize that the resource constraints may make it difficult, if not impossible, to address all of the information needs and questions posed by various stakeholders. It is, however, possible that pointing out that the resource constraints will make it difficult to answer key stakeholder questions will be a strategic lever for garnering additional resources for the evaluation.

5. Who will work on the evaluation?

In many cases, the answer to this question may simply be that the people who are reading this guide will work on the evaluation. Knowing who will be able to work on the evaluation is a critical part of the resource issue discussed above. Specifically, evaluation planners should decide (1) who will be responsible for overseeing or managing the evaluation and (2) who else will be available to work on the evaluation (e.g., develop data collection instruments, collect and analyze data, prepare reports). This may also be a good time to begin thinking about whether and how the participants (teachers, in this case) and other school and district staff might be involved in the evaluation. (Options for teacher involvement in evaluations of professional development are discussed in more detail in later sections of the guide.)

As the planning process proceeds, the planners may identify additional staff needs or, if additional staff are not available, adjust the evaluation plan accordingly. Because work on the evaluation is likely to begin soon after the professional development gets underway, it is important to have staff in place as early as possible.

Reviewing staff requirements is a good time to decide whether or not it is advisable or necessary to seek the help of an evaluator or evaluation consultant. These individuals can contribute to discussions about professional development outcomes and indicators, and they can provide suggestions about appropriate approaches to data collection and data analysis. Skilled evaluators and data analysts can also help with more complicated design and analytic tasks, especially those associated with (1) choosing samples of participants for inclusion in various

data collection activities, (2) linking participation in professional development to changes in student learning outcomes over time, (3) analyzing large amounts of quantitative data, and (4) experimental and quasi-experimental designs.

District research offices and college and university faculty are two good sources of evaluation expertise. Evaluation firms and contractors are also good sources. When considering the costs of hiring an evaluator or evaluation consultant, planners should keep in mind that doing so may result in increased efficiency, thus reducing the overall cost of the evaluation while significantly improving the technical quality. In addition, relying on an external evaluator can add objectivity and credibility to the evaluation.

II. Evaluation Design and Data Collection

As already noted, evaluations of teacher professional development should be explicitly tailored to the activities being evaluated. Once the professional development plan has been developed, the evaluation planners can determine the questions that the evaluation will answer and how best to answer them. As they think about the evaluation questions and how to address them, the evaluation planners should also consider whether to (1) include all of the participants in the evaluation or identify a sample and (2) add a comparative dimension to the evaluation design.

Evaluation Questions

Evaluations of teacher professional development should focus on three questions:

- *Did the professional development take place as planned?*
- *What were teachers' perceptions of the professional development?*
- *Did the professional development achieve the intended outcomes?*

In deciding whether and how to address each of these questions, the evaluation planners should review the plans for the professional development and the logic model to determine what the evaluation should focus on and the kinds of data that should be collected. If the professional development plans are unclear or inconsistent about (1) the inputs, (2) the professional learning activities, and/or (3) the expected outcomes and indicators, the evaluation planners should discuss their concerns with the professional development planning team to clear up any ambiguities.

Collecting Data on Professional Development Implementation[2]

Examining implementation, including problems and impediments to implementing the activities as planned, can inform decisions about future professional development as well as decisions about necessary modifications or mid-course corrections to the professional development being evaluated. Specifically, evaluations should collect data on:

[2] Examining implementation and early outcomes is sometimes referred to as formative evaluation or process evaluation. Gauging participants' reactions to and perspectives on the evaluation, which is discussed later, is often included in formative evaluations.

- *Individual teacher participation* to determine whether those teachers who were targeted to participate actually did participate and whether they participated in all of the key professional learning activities specified in the professional development plan and/or logic model

- *The availability of the supplies, materials, and equipment*, especially in schools and classrooms, as specified in the professional development plan and/or logic model

- Whether the *professional learning activities occurred at the intended levels of frequency and duration and included the content* specified in the professional development plan and/or logic model

- The extent to which all of the *key actors* (e.g., presenters, facilitators, school-based professional development staff, principals) *carried out their responsibilities* as specified in the professional development plan and/or logic model

- The extent to which *contextual factors* (e.g., changes in school or district leadership, changes in school or district priorities, changes in resources, changes in teacher assignments, changes in student characteristics) *influenced implementation*, including teacher participation

Collecting data on the implementation of professional development is relatively straightforward, but it does require careful record-keeping that, in turn, requires reliable procedures to record the various kinds of data necessary to track implementation. These might include sign-in sheets to track teacher participation in centralized activities and activity logs maintained by school-based professional development staff to track various kinds of follow-up support for individual teachers and groups of teachers. Using online systems for maintaining activity logs and similar kinds of records is very efficient and inexpensive, especially with well-designed reporting forms.

Data on professional development implementation should be quantitative as well as qualitative. For example, when a professional development plan calls for weekly observations and feedback on classroom applications of new instructional strategies, the evaluation should collect data on the frequency and duration of these weekly sessions as well as on their substantive focus. This can be accomplished with logs maintained by the people doing the observations and providing feedback.

Although it will almost certainly not be a central focus of data collection on implementation, the evaluation should be sensitive to changes in the overall school and district context that are likely to affect implementation. For example, a change in district priorities for instructional improvement could have a

significant effect on teacher participation, especially if the change is somehow inconsistent with the purpose and goals of the professional development being evaluated. Similarly, unanticipated changes in the availability of resources to support the activities could impede implementation and participation. If changes in these and other contextual factors do occur, the evaluation team should document the changes and learn as much as possible about how they affect the implementation and outcomes of the professional development.

Collecting Data on Participant Perceptions of the Professional Development

Professional development planners and sponsors may want to know what participants thought of the activities as an early indicator of whether and how participants are likely to apply new knowledge and skills in their professional practices. Participant feedback can also yield useful information on whether the activities were implemented as planned and they pinpoint both components of the activities that went well and those that may require modification.

Usually the easiest and most efficient way to collect data on participants' perceptions of professional development activities is through a survey. Depending on the nature of the activity and the reasons for examining participant perceptions, surveys can address some or all of the following topics. (Sample survey items related to these topics are included in Appendix B):

- Understanding of the purpose of the professional development
- Ratings of the usefulness or relevance of key components of the professional development to current assignments/responsibilities, with special attention to perceptions of usefulness for working with their own students
- Perceptions of the extent to which professional development met individual professional learning needs
- Ratings of the alignment of the content or focus of the professional development with district or school improvement priorities, plans, and goals
- Perceptions of the kinds of support and encouragement they receive to actively engage in the professional development
- Perceptions of the kinds of support and encouragement they receive to apply new knowledge and skills in their classrooms
- Ratings of the likelihood of applying new knowledge and skills in the classroom

- Overall ratings of the usefulness of the professional development compared with other professional development

In addition, if the survey is about an activity or series of activities with multiple components and/or that extend over a relatively long period of time, the survey should include items that ask about participation in each component of the activity to help determine overall participation patterns and whether they appear to be related to attainment of the intended outcomes.

If possible, surveys should be administered three to six weeks after the professional development is completed, instead of the standard end-of-the-session administration used in many local evaluations. The advantage of this arrangement is that it provides an opportunity for participants to reflect on the professional development and how they benefitted from it. The primary disadvantage is that it may be difficult to administer the survey after participants have returned to their schools and classrooms.

Evaluation planners should also consider the option of administering surveys online. Online surveys are inexpensive, and available software packages make survey development, administration, analysis, and reporting much easier than using traditional paper-and-pencil instruments. These surveys can be administered quickly and efficiently through district email systems or by hosting them on district websites.[3] District information technology staff can provide the best advice on how to administer online surveys and how to meet local requirements related to online security and personal privacy.

The evaluation can also collect data on teacher perceptions of the professional development through focus group interviews conducted at a reasonable interval after the activities have been completed. Note that these focus groups can also be included as part of ongoing data collection that follows major components of the professional development.

The primary advantage of focus group interviews is that they yield rich information about teachers' perceptions of professional development and can complement the survey data. The primary disadvantages of focus group interviews are that they are more time-consuming than surveys and typically include relatively small samples.

Evaluations that focus only on teacher perceptions of some components of professional development, typically the workshop or training components that take place at the beginning of a longer series of professional learning activities, are of limited use because they provide incomplete information. Evaluations that examine teacher perceptions of all components (especially follow up and school-

[3] *Survey Monkey* is an example of an easy-to-use survey software package. The basic survey package is free, and more comprehensive and versatile versions are available at very low monthly rates. For more information on this software and how to use it, visit www.surveymonkey.com.

based activities) and that include a careful look at implementation and outcomes are much more useful in understanding what happened and how it paid off for teachers and their students.

Collecting Data on Indicators and Outcomes

Evaluations of teacher professional development collect data on (1) outcomes for teachers, including mastery of new knowledge and skills and changes in professional practice and (2) changes in student learning outcomes, behavior, and engagement in school that are associated with the teacher outcomes. In addition, these evaluations may examine changes in school organization and culture that are linked to the professional development.

Plans for data collection should include strategies for collecting data on all of the outcomes and related indicators included in the plan for the professional development being evaluated.

As a general rule, the plans for data collection should include strategies for collecting data on all of the outcomes and related indicators included in the plan for the professional development being evaluated. For example, if the expected outcome of professional development is participant mastery of the content of one of the subject areas included in the Voluntary State Curriculum (VSC), an indicator could be scoring at a specified level of proficiency on a written test on the content area upon completion of the professional development. In a more ambitious professional development activity, one outcome could be mastery of the content of one of the subject areas in the VSC, with a second outcome being understanding of and ability to apply appropriate pedagogical strategies necessary to help students master the content. The indicator for the first outcome could be passing the written test on the content. An indicator for the second outcome could be demonstration of the appropriate instructional practices in simulated or actual classroom settings. A third set of outcomes and indicators could focus on understanding and applying appropriate classroom assessments to gauge student mastery of content in one of the content areas of the VSC, and a fourth set could focus on changes in student learning associated with changes in instruction.

Not only do these examples suggest progressively more extensive professional development activities, they also suggest increasingly complex evaluation efforts and clearly illustrate the need for multiple data collection strategies. More important, they illustrate the relationship between choices about indicators included in the plan for the activity and decisions about the kinds of data that will need to be collected to determine whether the outcomes were achieved. Thus, the data collected to determine whether the first outcome (mastery of VSC content) was achieved will include participant scores on a test. The data collected to determine whether teachers can apply appropriate pedagogical practices could come from direct observations of instruction, from written responses to prompts calling for descriptions of instructional activities that would help students master

various parts of the relevant VSC content area, or from teacher logs in which they report on classroom activities. Data on whether teachers are able to apply appropriate assessment strategies could come from a review of teacher-developed classroom assessments and assignments. Data on changes in student learning could come from a review of samples of student work prepared in response to teacher-developed prompts related to the relevant VSC content, a review of student scores on relevant sections of local benchmark assessments, or, less frequently, a review of student scores on state assessments.

Just as the evaluation should be sensitive to changes in the local context that affect implementation of the planned professional development, it should also be sensitive to how changes in context affect the outcomes. Here, again, changes in school or district priorities could greatly influence teacher application and use of new knowledge and skills.

Evaluations should also be sensitive to unanticipated outcomes that may be attributed to the professional development. For example, professional development that relies on math resource teachers to help math teachers implement new content and instructional practices could result in other teachers and school administrators understanding the benefits of this kind of support. This new understanding could, in turn, lead to an increased demand for school-based professional development provided and facilitated by skilled resource teachers in other content areas. Alternatively, poorly planned and incompletely implemented school-based professional development in one subject area could seriously undermine interest and willingness to participate in similar kinds of activities in other subject areas. Both of these outcomes are important, and the evaluation team should do what it can to learn about how and why they occurred.

Focusing on a Sample of Participants Versus All Participants

For professional development that includes relatively large numbers of participants and/or that includes multiple professional learning activities that extend over time, it may make sense to collect data from a sample or samples of participants for at least portions of the evaluation. The advantage of looking at a sample is that it becomes more feasible to gather more detailed or in-depth information, especially about things like changes in classroom practice.

There are several potential challenges associated with focusing on a sample or samples of participants. First, there is the sampling task itself. One option is to select a random sample. This can be done with a table of random numbers or something similar. A second option is to draw a sample that is somehow representative of all of the participants. This requires having information about individual characteristics that might somehow affect participation patterns, application of new knowledge and skills, or both. For example, if the participants in a professional development activity focused on understanding and using instructional practices to foster reading comprehension are divided between

teachers who are reading specialists and those who are not, or teachers with extensive experience and those who are new to the profession, the sample should reflect the overall distribution of these characteristics.

A second challenge is communicating about the samples and sampling procedures. How and why were some participants and not others selected for the evaluation? Despite the evaluators' very good intentions to the contrary, identifying samples might somehow suggest that some participants are being singled out for special attention or scrutiny. The evaluators must address these concerns from the outset and make every effort to allay them to ensure that teachers will be willing participants in the evaluation.

In most cases, it will make sense to consult with a sampling statistician for advice on how to select appropriate samples.

Adding a Comparative Dimension to the Evaluation

A potentially powerful approach to looking at professional development outcomes is to compare changes in participants' (the treatment group) knowledge and skills with changes in the knowledge and skills of teachers who did not participate in the professional development or who participated in other professional development (the comparison group). Alternatively, the evaluation could compare the learning outcomes of the students of the teachers in the treatment group with those of the students of the teachers in the comparison group.

When done well, both kinds of comparisons help gauge the impact of the professional development on teacher knowledge and skills or classroom practice, and on student learning. Adding a comparative dimension to an evaluation also poses substantial challenges and may add significantly to the cost. In most cases, adding a comparative dimension to the evaluation will require seeking assistance from a skilled evaluator who can help with sampling issues as well as complex data analysis.

At a minimum, this design option requires identifying comparison groups of teachers who share many if not all of the characteristics of the participating teachers. For example, the comparison group should include teachers who teach at the same grade level and in the same content areas as the participating teachers. In addition, the comparison group should teach students with similar characteristics and work in schools that are similar to those of the participants.

Conceptually, identifying comparison groups, especially when they are matched to treatment groups on a number of variables, helps to focus on the treatment, in this case, professional development, as the "cause" of the observed outcomes and to eliminate other explanations of causality. An even more powerful evaluation design is random assignment of teachers to treatment and control groups and to compare the two groups on the outcome variables of interest at the end of the

professional development or at the point at which planners expect to be able to observe the intended outcomes. Random assignment designs are recognized as the "gold standard" in studies of program effectiveness and impact.

For most districts and many evaluations, both of these design options have some serious practical disadvantages. First, they can be very costly and time-consuming. Second, they can require relatively large samples of teachers and students, and, in some cases, there may simply not be enough teachers available to form either a comparison or control group.[4] In some situations, a third challenge may be having to justify the random assignment of teachers and/or students to treatment and control groups to the teachers and to parents and others in the local education community.

Preparing for Data Collection

For any evaluation of teacher professional development, preparing for data collection involves:

- *Selecting appropriate instrument(s)*
- *Preparing staff who will collect the data*
- *Gaining access to people and data*

In some evaluations, completing these preparations will be fairly simple; in others, the preparations will be more complicated. Some preparations will be very inexpensive while others may be quite expensive and time-consuming. In some evaluations, costs will trump all other factors in deciding how to complete these preparations. In these cases, evaluation planners need to recognize and be prepared to explain the trade-offs being made. In every evaluation, preparation is critical to the success of the effort.

Selecting appropriate instruments. (Evaluation resources included in Appendix A provide extensive information about a variety of instruments that can be used in evaluating teacher professional development.) Typically, evaluations of teacher professional development will employ one or more of the following kinds of instruments:

- *Written tests* to assess participants' mastery of professional knowledge (including knowledge of curriculum content) and skills

[4] One solution to this problem is to identify comparison groups or control groups from other districts. While this strategy can add rigor to the evaluation design, it also adds to costs and it may not be feasible for other reasons.

- *Surveys* to collect data on (1) participants' experiences and perspectives on the professional learning activities and (2) participants' attitudes (including changes in their attitudes about teaching and their careers)

- *Protocols for observations and interviews* to collect data on (1) participants' perspectives on the professional learning activities, (2) implementation of professional development activities, and (3) application of new knowledge and skills in their classrooms and other settings

In some cases, evaluations will collect data on students. Most evaluations that focus on student outcomes associated with teacher participation in professional development will rely on student work samples, results on state and local assessments, and administrative records of student attendance, behavior, and discipline.

A practical first step in selecting instruments is to see if appropriate instruments already exist. For example, some districts have developed generic survey instruments to gauge teacher perspectives on their professional development experiences. These instruments may require only a bit of tailoring to be appropriate for the evaluation being planned. Some curriculum materials include assessment instruments to measure content mastery, and locally developed observation protocols and guides may be useful in gauging implementation of new practices. Finally, a number of instruments have been developed to measure teacher attitudes.[5]

Adopting and/or tailoring existing instruments can be relatively inexpensive and require very little time. In some cases, there may be the added advantage of formal guidance for administration and use of these instruments. In addition, results from other evaluations using these instruments may be available that could be applied in some sort of a comparative analysis of the results from the professional development evaluation being planned.

Here are some criteria to guide selection and/or development of instruments:

- Instruments should be designed to *minimize the burden* on respondents (surveys, interview protocols, written tests) and users (observation and interview protocols). Overly long surveys and interviews tend to generate bad data and may result in respondents' unwillingness to participate in other kinds of data collection activities. Observation protocols should not require more than 30 minutes per observation, and less if possible. Surveys should require 15-20 minutes at most and, again, less if possible.

[5] Visit http://www.coe.ohio-state.edu/ahoy/researchinstruments.htm for links to a number of survey instruments and detailed information about their use and psychometric properties.

- The content or substantive focus of the instruments should be *aligned with the content of the professional development.* Observation protocol prompts should call for observations and data that explicitly reflect key elements of the content of professional development (e.g., frequency of teacher use of specific questioning strategies, teacher use of appropriate curriculum materials, teacher use of appropriate strategies to gauge student learning).

- As a general rule, observation protocols *should call on observers to describe* classroom activities.[6] These protocols *should not call on observers to judge or rate* the activities (e.g., rate teacher proficiency on a scale of 1 to 5). Protocols that collect descriptive data could, for example, call on observers to count or report on:

  - The number of times that teachers use various questioning strategies
  - The number of students to whom the teacher directs questions and feedback
  - Teacher use of academic language versus less formal language
  - The presence and use of certain instructional materials and equipment
  - The particulars of how a teacher uses prompts to assign student work (e.g., time spent on discussing the assignment, opportunities for students to collaborate and collect information)

  In contrast, protocols that do not rely on well-defined rubrics or that call for anecdotal reports or subjective judgments yield data that are difficult to analyze or data that are simply not very useful. Finally, protocols that call for observers' judgments or ratings require extensive training to ensure consistency within and across the observations.

- Survey instruments and other self-report forms *should include close-ended items.* Responses to these items are easier and less expensive to analyze, and they ensure a degree of consistency

[6] See "Professional networks and school improvement," by Richard Elmore (*School Administrator,* April 2007), also available online at www.aasa.org/publications. Elmore discusses the use of protocols in describing and discussing classroom practice and highlights the importance of careful preparation for effective use of the protocols. See also *The power of protocols: An educator's guide to better practice,* by McDonald et al. (2007), for extensive guidance about using protocols for collecting data and discussing the results.

among respondents. Responses to open-ended responses require careful coding and are therefore more difficult and more expensive to analyze.

A note of caution about adopting locally developed instruments—especially instruments designed for purposes other than evaluating teacher professional development—is in order here: Planners should review instruments very carefully to be sure that they are, in fact, suitable for the evaluation being planned. For example, as noted above, protocols developed for classroom observations, walk-throughs, and teacher performance appraisals may be useful instruments for evaluating the impact of professional development on teacher performance. At the same time, planners should be sure that these instruments are designed to collect data on the specific indicators that are included in the evaluation. In other words, these instruments should meet reasonable standards for validity. Instruments that are generic or that focus on indicators other than those included in the evaluation will not yield data that will help gauge the effectiveness of the professional development. In short, selection and/or development of appropriate instruments depend in large part on very clear specification of what the instrument is intended to help measure. What are the key indicators of successful implementation of new practices? How often should they be present and under what circumstances? The evaluation planners should ask these and other questions. The answers should be provided by those who are planning the professional development.

Selection and/or development of appropriate instruments depend in large part on very clear specification of what the instrument is intended to help measure. What are the key indicators of successful implementation of new practices? How often should they be present and under what circumstances? Evaluation planners should ask these and other questions. The answers should be provided by those who are planning the professional development.

For evaluations that focus on student outcomes to gauge the long-term impact of professional development, it is important to be sure that assessment tools, including locally developed tests and assignments or prompts used to generate student work samples, are aligned with the knowledge and skills that teachers are expected to demonstrate as a result of their participation in professional development.

When using student work samples, the evaluation planners and the professional development planners should also think very carefully about how the assignments and prompts will be administered and whether and how the evaluation will monitor the process. Differences in how assignments are made or how prompts are used may result in considerable differences in student work. These differences could, in turn, lead to differences in judgments about instructional practices. Consider the following example:

Using Student Work Samples in Evaluating Teacher Professional Development: A Cautionary Example

The professional development planners and the evaluation planners agree that one source of data about the student outcomes that follow from teacher participation in a long-term professional development designed to help them improve instruction in 20th-century American history will be student essays on the causes of the Korean War. Teacher A and Teacher B have recently completed the three-month professional development activity, and their students' essays will be reviewed to gauge whether the teachers have successfully implemented the instructional strategies and curriculum content that were the focus of the professional development.

In reviewing the completed essays, the evaluation team observed that the essays of the students of Teacher A were much shorter and less well-developed than those of the students of Teacher B. The team also observed that the essays of the students of Teacher A reflected very limited understanding of the economic and political factors that preceded the conflict. Because the two teachers gave their students the same assignment, and because classroom observations of the two teachers that took place after the professional development indicated that both had mastered the new content and related instructional strategies, the evaluation team initially concluded that the students of Teacher B were more able than the students of Teacher A. However, because there was no other obvious evidence of these differences in indicators of students' ability, the evaluation team was puzzled by the result and decided to look at the observation data a second time. What they discovered was that although instruction was quite similar in the two classrooms, Teacher A had written the prompt on the board at the beginning of a class period and, after responding to a few questions, had given the students two days to complete their essays. In contrast, Teacher B had spent the better part of a class period discussing the assignment and responding to a variety of student questions. Next, she organized students into small groups to brainstorm possible explanations for the causes of the Korean War. After that, she encouraged them to go to the school library to collect additional information before completing their essays. Not surprisingly, the result of these different approaches to the same assignment was essays of very different quality.

Preparing the evaluation team. No matter what the evaluation design or which data collection instruments are used, a critical step in conducting good evaluations is to thoroughly prepare individuals who will be responsible for data collection. Ideally, this preparation should include:

- *A thorough orientation to the evaluation plan, with special attention to data collection tasks and responsibilities and the amount of time required.* This is particularly important when data collection includes interviews, observations, and other direct contact with participants. These forms of data collection can be quite time-consuming, so it is important that data collectors understand the amount of time that they will need to devote to the task. This is especially important if district staff are responsible

for data collection. Finally, it will be important to review confidentiality issues that may arise in the course of data collection. Data collectors need to understand that it is imperative to protect the identity of participants in the evaluation and how and with whom they are expected to share information.

- *A detailed review of the data collection instruments and how to use them.* This should include an explanation of the purpose of using the instrument or instruments, strategies for introducing and explaining it to participants, and expectations for how data are to be recorded and/or reported in preparation for data analysis. Preparation for using observation protocols, interview guides, and similar instruments should also provide very clear guidelines for the frequency and duration of use. For example, if an observation protocol is to be used to measure changes in instruction or application of new instructional strategies and content, observers need to know how many times they are expected to observe each teacher, appropriate intervals between observations, and how the data are to be recorded and reported.

- *Opportunities to practice using data collection instruments, such as observation protocols and interview guides.* Practice in using these instruments ensures that data collectors understand their responsibilities and that data collection is consistent and thorough. Practice in using observation protocols may also afford opportunities to test the reliability of the instruments and to make refinements as necessary. Examples of practice in using data collection instruments include role-play situations in which members of the evaluation team practice using interview protocols with other members of the team and team members completing observation protocols as they view videos of vignettes of classroom instruction.

Securing participation in the evaluation. Following three basic strategies can greatly facilitate ensuring the willing participation of teachers and others in the evaluation. The strategies are: (1) be transparent about the evaluation plans and how the results will be communicated, (2) ensure confidentiality, and (3) make participation voluntary.

- *Transparency.* Evaluators need to be clear about (a) the purposes of the evaluation, (b) what participation will entail in terms of time and other possible commitments, (c) what data will be collected and how, and (d) how the evaluation results will be communicated and to whom. Teachers may have concerns about the extent to which the evaluation will somehow be an evaluation of them and their work. The evaluation team needs to reassure them that the evaluation is assessing the professional development and not

teacher knowledge and skills and/or performance in the classroom. At the same time, the evaluators should also make it clear that teacher knowledge and skills and performance are included among the expected outcomes and, therefore, the focus of the data collection activities.

- *Confidentiality.* Evaluators need to assure teachers and others who participate in the evaluation that all data collected from and about them will be maintained in strict confidence and that they will not be identified by name in any reports or other communications about the evaluation. As appropriate, these assurances should explain the procedures for maintaining confidentiality. For example, the evaluation team will probably store data in secure files accessible only to members of the team. Similarly, teachers and others who participate in the evaluation may be assigned an identification number so that all names and other identifying information can be eliminated from data files.

- *Voluntary participation in the evaluation.* As a general rule, teachers and others who may be involved in the professional development as participants or in other ways should be *invited but not required* to participate in the evaluation. In addition, they should be permitted to opt out of the evaluation at any point in the process. To ensure that participants understand what is being asked of them, the evaluation team should prepare a short, written description of the evaluation. This description should, at a minimum, explain the overall purpose of the evaluation, the amount of time required to participate, and how the results will be reported and to whom.

Gaining access to students and student data. If an evaluation focuses on changes in student learning or in other student outcomes (e.g., attendance, truancy, disciplinary referrals, suspensions, expulsions) or on student perceptions of changes in teacher performance, it may be necessary to seek parental consent to gain access to student records, to interview students, or ask them to complete a survey. Many districts have well-defined procedures for gaining access to students and student data, and it is incumbent on evaluators to pay close attention to these procedures. Here, again, written communications explaining the purposes of the evaluation, the kinds of data that are necessary, procedures for protecting student privacy and confidentiality, and how the data will be used are important. In addition to obtaining parental consent, it will be necessary for the evaluation team to work closely with district staff who maintain district data systems. Indeed, it may be necessary to call on them to extract data from student files.

A Note About Classroom Observations

Classroom observations are a very good way to collect information on professional development outcomes related to teachers' application of new knowledge and skills. As already noted, the success of these data collection activities is dependent, in part, on the availability of good instruments and careful training in using them. In addition, observers should:

- *Conduct multiple observations* of each teacher over a period of several weeks or perhaps even a few months
- *Complete data collection reports as soon as possible* but no later than 48 hours after the observations
- *Avoid scheduling observations on the days before or after school holidays or on "special days,"* which may have shortened or alternative schedules
- *Immediately report any problems*, including concerns from teachers who are observed, to the person who is leading the evaluation and/or managing the observations

Evaluation planners should also think about the option of following up classroom observations with short interviews. These interviews can elicit teachers' views of what happened, how things worked, and why. These interviews can also examine teachers' reasons for using various content and instructional strategies. As appropriate, these interviews should be scheduled to avoid additional disruption of classroom activities.

One of the important issues in planning the observational components of an evaluation is to decide what role district staff will play in data collection, if any. School-based professional development staff, supervisors, principals and assistant principals typically spend a lot of time in both formal and informal observations. Therefore, it may make sense to recruit them as data collectors for the evaluation. At the same time, there are some additional factors to consider. First and foremost, it will be difficult for principals and others who are responsible for teacher performance appraisals to separate their role in an evaluation of professional development from their performance appraisal role. In addition, teachers who are included in the evaluation may not understand or appreciate the different roles and may therefore be uncomfortable in observations related to the evaluation. (To be sure, they may be uncomfortable being observed as part of a formal performance review, but that is another matter.) One option for involving principals and assistant principals in data collection is to have them conduct interviews and observations with teachers in other schools.

School-based professional development staff and other staff who spend time working with teachers but who are not involved in performance appraisals may be good choices for collecting evaluation data.

In addition to deciding about recruiting building administrators and district to help with data collection, there is also the question about using information collected for other purposes (e.g., performance reviews, monitoring implementation of new practice) for an evaluation of professional development. Here, again, it may be tempting to take advantage of information that has already been collected. Indeed, it is possible that the professional development planning team used some of these data to determine the need for the professional development and to identify the teachers who should participate.

At least three factors will influence the decision about using these data. First, district personnel/human resource policies and negotiated agreements may prohibit the use of data from performance reviews for any other purpose or they could require that teachers give their permission for other uses. Second, the evaluation planners should review the alignment of the data collection process (e.g., formal observations, walk-throughs) with the content and purpose of the professional development. Misalignment means that the data collected for these other purposes will be of limited use for the evaluation. Third, the evaluation planners should review the quality of the data collection process and the data. Inadequate data collection and incomplete or weak data are of no use to the evaluation. (Of course, these same considerations should apply to the use of these data for determining the need for the professional development and who should participate.)

Involving Teachers in Data Collection

The *Maryland Teacher Professional Development Planning Guide* encourages including teachers in planning professional development. Teachers can also play key roles in evaluating professional development, especially professional development that is designed by school improvement teams as part of their school improvement plans and that takes place in the school as part of regular school activities.

Involving teachers in data collection and, later, in data analysis will almost certainly extend and enrich their professional learning. In addition, involving teachers in evaluating their professional development is worthwhile because it establishes their ownership of the effort as well as the results.

One good way of involving teachers in data collection is through peer observations of classroom practices. When done well, peer observations generate a lot of information on teaching practices and help school faculties develop and

use a common language about instruction and how to improve it.[7] In addition, teachers can help develop observation protocols and the preparations for using them. Consistent with earlier suggestions for preparing for data collection, preparations for peer observations should include practice in using the protocols and clear guidance and explanations about how to record and share data.

Despite the fact that there are some real benefits to teacher involvement in this form of data collection, there are several important challenges that need to be addressed. First and foremost, preparing for and conducting the observations requires time and may also require making arrangements for covering the classes of teachers who are conducting the observations. Ideally, the observations can be conducted during preparation time or during non-instructional periods. When this is not possible, the principal and teachers will need to work together to find ways of covering classes to minimize disruptions in instructional activities. Second, it is incumbent upon principals and other school leaders to establish an atmosphere of trust among teachers, to help them recognize the payoffs of the observations, and to promote discussions of the results to improve practice and to determine the focus of future professional learning activities.

[7] For a more extensive discussion of how to plan and use peer observations, visit the Annenberg Institute's "Tools for School Improvement Planning" at www.annenberginstitute.org/Tools.

III. Data Quality and Data Analysis

Data analysis begins early with ongoing monitoring of data quality and continues with the application of appropriate analytic procedures. Just as data collection reflects decisions on the outcomes and indicators that will be of interest in the evaluation, data analysis anticipates the reporting task and sets the stage for presenting the evaluation findings. This section of the guide provides suggestions for monitoring data quality and general approaches to data analysis. The guide does not, however, discuss specific statistical procedures that evaluators can use to analyze evaluation data.

Monitoring Data Quality

As noted earlier, a key to a successful evaluation is having solid data. In addition to selecting appropriate instruments and preparing the evaluation team to use them, it is necessary to monitor the quality of data as they are being collected. Although there will be exceptions, it is almost always difficult to collect additional data after the planned data collection has been completed. Therefore monitoring data quality while data collection is underway is essential to the success of the evaluation. Here are some tips:

Monitor response rates on surveys and other quantitative data collection. When administering surveys, pre-test/post-test instruments to measure changes in participant knowledge, or similar kinds of instruments, the evaluation team should check to see that participants complete and return the instruments. Ideally, there will be a procedure for tracking individual survey responses that will, in turn, make it possible to follow up with non-respondents. If surveys are administered online, the evaluation team can send email reminders to non-respondents. Alternatively, if it is not possible to track individual responses, blanket email responses can be sent to all participants. If the survey permits identifying subgroups of participants (e.g., elementary school reading teachers, elementary special education teachers), it is possible to send email reminders to the subgroups with low response rates.

The higher the survey response rates, the better. Time and resource constraints permitting, the evaluation team should aim for response rates of at 80 percent for all participants or 80 percent for each identifiable sub-group of participants. As response rates drop below 80 percent, it becomes more difficult to analyze the data or to draw any meaningful conclusions from them.

Monitor record-keeping necessary for tracking participation and implementation of the professional learning activities specified in the plan for the activity. Developing and maintaining record-keeping systems is necessary for the evaluators to be able to report on participation patterns and implementation of key components of the professional development being evaluated. This task becomes both more important and more complicated as activities increase in size and scope. Note also that sections of evaluation reports that present

findings about participation and implementation should describe what actually happened as opposed to making assertions about what was supposed to have happened.

As noted earlier, sign-in sheets are good sources of data on participation in large group activities. Evaluators should work with professional development providers and facilitators to ensure that participants record their attendance at these sessions. The evaluators should also work to ensure that the sign-in sheets collect information at an appropriate level of detail. For example, if the professional development begins with five days of workshop sessions, there should be a sign-in sheet for each workshop session. This permits tracking overall participation as well as variations in participation patterns. Data from the sign-in sheets should be entered in the evaluation database as soon as possible after participants sign in.

For professional development that includes multiple sessions that take place over several weeks or several months, the data system should be organized to track individual teacher participation in all of the activities. Later, these data can help profile participation patterns and help evaluators determine whether the variations are sufficiently large to compare and contrast various participation patterns as they may be related to teacher outcomes. In the shorter term, interim evaluation reports may pinpoint gaps in participation, thus making it possible for professional development providers and facilitators to identify ways to improve participation as the professional development continues over time.

Similarly, if the professional development includes follow-up activities such as observations, feedback, and assistance from school-based professional development staff, evaluators should work with providers and facilitators to develop ways of keeping track of these activities. One approach is to ask school-based professional development staff to maintain simple activity logs to record information about their interactions with participants in school or classroom-based follow-up. If the evaluators are maintaining master files, data from the logs can easily be combined with data from sign-in sheets to generate more complete participation records.

Monitor the accuracy and completeness of entries in observation protocols and similar kinds of data collection instruments. If data collection includes structured or unstructured observations, interviews with teachers and others, or similar kinds of activities, the evaluators should monitor the processes to ensure that the various data collection activities are taking place as planned. They should also check completed reporting forms to make sure that data are being recorded correctly. As with the other monitoring suggested here, tracking these efforts can help ensure that there are no gaps in the data and that the data are of high quality. It is relatively easy to solve data collection and data quality problems if they are identified early while data collection is still underway; it is almost impossible to solve them after the fact.

Convene the data collectors to review data collection efforts. When an evaluation includes several cycles of data collection, it is a good idea to convene the data collectors after each cycle to review the data collection activities. Specifically, data collectors could be invited to discuss any problem that they are experiencing in using data collection instruments, especially observation and interview protocols. Are there important things that they are seeing or hearing that are not included in the protocols, and, if so, how can the protocols be modified to capture these things? What findings are emerging? Should data collection instruments and plans be modified to examine these findings in more depth?

Data Analysis

Learn about the data. A real benefit of monitoring data quality in the ways suggested previously is that the process helps familiarize the evaluators with the data and what the data suggest about the professional development. Indeed, an important first step in data analysis is to get to know the data and to see what they appear to be suggesting even before conducting more careful and systematic analysis.

Start with the basics. In many evaluations of teacher professional development, much of the data analysis will involve counting. Here is a partial list of questions that can be answered by counting various kinds of data:

- How many teachers participated in *all* of the learning activities and how many teachers participated in only a few of the learning activities?
- How many participants gave the activities high marks for usefulness and relevance?
- How many participants reported that they received adequate help in applying new knowledge and skills in their classrooms?
- How many participants viewed the activities as "a good start" versus "providing them with all the information they needed to apply new knowledge and skills?"
- How many participants scored at or above proficient on a written test of content area knowledge administered one month after the professional development? Are there consistent gaps in content knowledge?
- Based on completed observation protocols, how many participants could be rated as proficient in applying new instructional strategies in their classroom? What gaps in applications are suggested by the observational data?

Depending on the design of the professional development, the kinds of data that are available, and the quality of the data systems developed for the evaluation, it may also be possible and useful to compare and contrast the perceptions and experiences of various groups of participants. Consider the following example:

The Case of the Accidental Comparison Group and the Benefits of Looking at the Evaluation Data a Second Time

The activity being evaluated focused on helping participants understand and apply instructional strategies to increase reading fluency. The first phase included a concentrated series of workshops on these strategies, and the second phase included ongoing observations and coaching from school-based reading resource teachers. The evaluation design called for (1) documenting participation in the two phases of the professional development, (2) using a written test to assess participant understanding of instructional strategies associated with helping students develop reading fluency, and (3) assessing participant application of the instructional strategies through a series of structured observations scheduled to take place 4-8 weeks after the completion of all of the professional learning activities, including the school-based supports.

Initial evaluation results indicated that almost all participants participated in all of the workshop activities and that most gave the activities high marks for the potential usefulness of the content. Most participants also agreed that the strategies would work well with their students. In addition, the scores on the written test clearly suggested that a large majority of the participants understood the new instructional strategies.

To the evaluators' surprise, however, the initial review of the observational data revealed rather large differences in classroom implementation and use of the new instructional practices. Because they were able to link the data on coaching support with the data on classroom instruction, the evaluators were able to examine the extent to which more extensive coaching appeared to be related to more extensive implementation of new instructional practices. Subsequently, when the evaluators reviewed the activity logs maintained by the reading resource teachers, they found considerable variation in the amounts and kinds of school-based follow-up to support implementation of the new strategies. Some teachers received help several times a week while others received help once a week or even less. Later, when the evaluators re-examined the data in the observation protocols used to measure the extent to which teachers were applying the new practices, they found that teachers who had received more follow-up help were better able to use the new instructional practices than teachers who had received less help.

In looking at the observation records a second time, the evaluators also found that there were school-level differences in the amount of coaching support that teachers received. Teachers in some schools received considerably more support than did teachers in other schools. Curious, the evaluators returned to the schools to talk with the principals and reading resource teachers about the arrangements for working with individual teachers. These conversations revealed that there were important differences in how the principals and reading resource teachers worked together and in the amount of support and encouragement that the principals provided. This led the evaluators to conclude that stronger working relationships and higher levels of encouragement and support were clearly related to the reading resource teachers providing more classroom support.

Quantify the qualitative data. Qualitative data, including data collected through observations, interviews, self-reports, and student work samples yield rich descriptions of professional learning activities, classroom instruction, and student learning outcomes. These descriptions are especially useful when they include some sort of quantification to illustrate the extent to which they are typical or representative of the experiences of all of the participants. Thus, observation protocols that yield detailed descriptions of instructional practices (i.e., the number of students engaged in class discussions, the number of students to whom teachers direct particular kinds of questions, the availability and arrangement of classroom supplies and materials, etc.) also make it possible to report on how many teachers are actually implementing new practices. Reporting observers' judgments about the extent to which teachers are implementing new practices makes it incumbent on the evaluators to explain the extent to which the observations may reflect observer bias.

When an evaluation collects more open-ended data, such as data from semi-structured observations or interviews, the analytic task is to review the write-ups from the various data collection activities for keyword and key themes and to count their frequency. Because of the costs associated with preparing observers and interviewers to carry out these data collection activities as well as the costs associated with data analyses, evaluation planners should think very carefully about pursuing this design option. One solution to the problem of analyzing large amounts of qualitative data, especially data from structured and semi-structured observations and interviews, is to use software programs designed specifically for this purpose.[8] The advantage of using this software is that it helps organize and analyze large amounts of qualitative data.

Anecdotal data, including self-reports of implementation and use of new instructional practices, can be useful in analyzing changes in teacher practice. Once again, counting is important. How many teachers actually reported applying a new instructional strategy and provided a concrete example? How many teachers actually reported changes in student outcomes that resulted from the application of a new strategy and provided a concrete example of change in student work or assessment results?

Involving Teachers in Data Analysis

Just as they can be involved in data collection, teachers can play a role in data analysis. Teachers can, for example, share the results of peer observations and examine what they learned about the implementation of new instructional strategies. Teachers can also

Teacher involvement in data analysis is, by itself, a valuable professional learning opportunity.

[8] Visit www.qsrinternational.com for more information about software available for use in qualitative data analysis. Note that some versions of the software are available at no cost.

participate in reviewing samples of student work that serve as indicators of the implementation of new instructional strategies.[9] As with involvement in data collection, teacher involvement in data analysis is, by itself, a valuable professional learning opportunity.

Keys to involving teachers in data analysis are similar to those related to involving them in data collection, and they include:

- *Setting aside adequate time* in the regular school schedule
- *Orienting them to the task,* especially strategies for discussing the data (both observational data and student work samples) objectively
- *Establishing an atmosphere of trust* in which observations and comments can be freely shared

Explaining Cause and Effect in Evaluations of Teacher Professional Development

Perhaps the greatest challenge in evaluating teacher professional is to determine patterns of causality. The bad news is that the evaluation designs necessary to empirically determine causality are beyond the scope of most state and local evaluation efforts. The good news is that there are options for looking at causality in evaluations such as the ones discussed in this guide.

The evaluation planners should look carefully at the data used by the professional development planners to assess the need for the professional development. What do these data suggest about teacher knowledge, skills, and practice prior to the professional development? For planning and design purposes, these data can be considered as baseline or pre-intervention data. In general, the goal of professional development is for teachers to move from the baseline to higher levels of knowledge, skills, and performance, and the purpose of evaluation is to determine whether these changes did occur and, if so, whether they are attributable, at least in part, to teacher participation in the professional development.

Next, as discussed in earlier sections of this guide, the evaluation should determine whether the professional development was implemented or occurred as planned. If it did take place as planned, the evaluation should look for evidence that the intended changes in teacher knowledge, skills, and professional practice (as specified in the outcomes and indicators included in the original plan) did, in fact occur. If these changes did occur and if they occurred after participation in the professional development, the evaluation should focus on explaining the

[9] For more ideas about involving teachers in the analysis of student work and using protocols in this process, visit www.lasw.org/protocols.html.

reasons for those changes. If the professional development did not take place as planned and/or if the intended changes in teacher knowledge, skills, and professional practice did not occur, possible claims about causality are weak at best. At the same time, if implementation was incomplete (including participation patterns that did not meet the expectations included in the original plan for the professional development) but at least some of the intended outcomes were achieved, the evaluation should look for patterns of causality.

In the kinds of evaluations discussed in this guide, one of the best ways to examine causality is to ask teachers directly about whether and how their professional development experiences contributed to changes in knowledge, skills, and professional practice. For example, after a series of classroom observations that yield clear and consistent evidence of appropriate application of instructional strategies that were the focus of the professional development being evaluated, the evaluators should interview teachers to ask where and/or how they learned about these strategies and what motivated them to try the strategies in their classrooms.[10] Follow-up questions should probe for concrete examples of the various factors that influenced teacher use of new strategies. Follow-up probes should also explore the relative salience of various factors in teacher decisions. It is possible—perhaps even likely—that teacher responses will suggest an array of factors that contributed to these changes. Careful analysis of these responses will help evaluators understand whether and how participation in professional development "caused" changes in teacher knowledge, skills, and performance.

Extending the causal links to student outcomes related to teacher participation in professional development is even more difficult. This difficulty also underscores the importance of examining indicators of student learning that are closer (proximal indicators) to the focus of the professional development than those that are more distant (distal indicators). Student work samples produced in response to prompts given by teachers and scores on locally developed assessments are examples of proximal indicators. Scores on state assessments, such as the HSA and MSA, are examples of distal indicators. Conceptually, the advantages of using proximal indicators are that they (a) are observable and measurable sooner than distal indicators and (b) typically reflect a narrower range of student learning. A related advantage is that because they are observable sooner—and closer to teacher participation in professional development—they are less likely to be influenced by other variables. Nevertheless, the evaluators need to be very careful about linking teacher participation in professional development to changes in student learning.

In the end, findings and conclusions are almost certainly to be somewhat tentative and speculative. The task of the evaluation is to present the data clearly and to

[10] It is generally not a good idea to rely on surveys or other self-report forms to ask teachers about the factors that influenced their use of new strategies, unless there is an opportunity for some sort of direct follow-up that permits probing the initial responses for concrete examples. Open-ended survey questions are especially difficult to analyze and are therefore not good sources of information.

provide a compelling case for linking participation in professional development to changes in teaching and learning. The case is most compelling when it recognizes the complexity of the causal chain and acknowledges that a variety of factors contribute to changes in teaching and learning. The case is least compelling when it simply asserts that teacher participation in professional development resulted in improved teaching and learning.

IV. Reporting

The last step in any evaluation is the preparation of the evaluation report. There is no single formula for evaluation reports, although they should follow a few basic principles to help readers understand the professional learning activities that took place and the extent to which they contributed to changes in teacher knowledge and skills, and how these changes led, in turn, to changes in student outcomes. In general, evaluation reports should:

There is no single formula for evaluation reports, although they should follow a few basic principles to help readers understand the professional learning activities.

Anticipate readers' information needs and interests. Recall that one of the issues to be addressed in planning the evaluation was to anticipate who would be interested in the results of the evaluation. Possible audiences include local and state policymakers, funders, principals, and, of course, the participants. Each group is likely to have different interests, although there will almost certainly be some overlap. For example, all of these groups and others who may read the report will want to know whether the professional development achieved the intended results as they were specified in the original design. Audiences are likely to vary in their interest in whether the activities unfolded as planned, although principals, participants, and staff who worked on the activities are likely to be interested in problems that may have arisen as the activities unfolded.

Most readers will welcome short, non-technical reports. Writers can help readers by:

- *Including an executive summary and clear, concise summaries* at the end of each major section of a report
- *Using advance organizers* to provide a map of the report and the key points
- *Defining important terms* (and possibly including a glossary)
- *Providing concrete examples* to amplify descriptions and observations
- *Using charts and exhibits* to present data and illustrate important findings
- *Including technical appendices* to provide additional detail about the evaluation design, data collection, and data analyses for interested readers

Describe the approach to the evaluation. These descriptions should explain what data were collected, how the data were collected, from whom they were collected,

and the timeline for data collection. These descriptions can also include copies of key data collection instruments or the instruments can be included as appendices to the report. Finally, these descriptions should present information about response rates, the extent to which complete data are available for all participants included in the evaluation, and any problems that were encountered in data collection.

Explain the purpose of the evaluation and address each of the evaluation questions. In describing the purpose of the evaluation, the report should introduce the logic model (or other conceptual framework) that guided the professional development and that informed the evaluation. Alternatively, the description of the purpose of the evaluation should discuss the key assumptions that guided the professional development and the evaluation.

After discussing the purpose(s) of the evaluation, the report should list the evaluation questions and explain why they are important. The answer(s) to each question should include as much detail as necessary to help the reader understand the evaluation results, but these discussions should also include a clear, concise summary of the answers.

Explain data analysis procedures in sufficient detail to permit interested readers to draw their own conclusions about key findings and results. These explanations should also describe the kinds of analyses that were conducted and, as appropriate, describe why these procedures were selected. Depending on the complexity of the analyses and readers' interest, it may make sense to discuss the analyses in greater detail in a technical appendix to the evaluation report.

Present all of the news—both the good and the bad. It is important for evaluation reports to discuss all of the key findings about each of the evaluation questions, whether the findings are positive or negative. Evaluation reports that present only positive results and gloss over or neglect negative findings often have little or no credibility among readers. Indeed, most readers understand that professional development activities—especially those activities that involve large numbers of participants and that extend over long periods of time rarely take place exactly as planned. Similarly, professional development does not always achieve the intended outcomes.

Presenting evaluation results objectively and in a straightforward manner helps readers understand what happened and what results were achieved. It is also important to report on any problems that affect the evaluation itself. For example, evaluation reports should be clear about problems such as low survey response rates that make it difficult to interpret survey results. Similarly, if classroom observations or teacher interviews included as part of the evaluation design were not conducted as planned, the evaluation report should discuss these problems and provide a clear explanation of how data collection problems or gaps in the data affected analysis and the confidence that readers should have in the evaluation findings.

While it is important to report negative findings and problems that may have affected evaluation results, it is equally important that these reports not overemphasize the negative findings. To illustrate how a few words change the "meaning" of an evaluation finding, consider the following examples of a description of survey results:

How a Few Words Can (Re)Define the Evaluation Findings

- A *scant* 25 percent of teachers who responded to the survey reported that they found the activity useful compared with other professional development that was available to them.
- *Twenty-five percent* of teachers who responded to the survey reported that they found the activity useful compared with other professional development that was available to them.
- An *impressive* 25 percent of teachers who responded to the survey reported that they found the activity useful compared with other professional development that was available to them.

The first and third examples convey judgments about the findings and may reflect the evaluator's bias. The second example simply reports the finding. Both the first and third examples call for an explanation of what criteria, either implicit or explicit, led the evaluator to make such a judgment. The second example does not require an explanation, although the evaluator may choose to provide one in a summary or concluding section of the report.

In general, when an evaluation report presents judgments about the quality of professional development or about its impact or effectiveness, the report should also be clear about the criteria or standards used to make the judgment. For example, if the evaluation report concludes that the professional development did not achieve the intended results, the report should remind the reader about what the expected outcomes were. Further, if the report attempts to explain why the professional development did not achieve the intended results, the report should be very clear about the extent to which the proffered explanations are based on findings from the evaluation, especially findings about implementation and participation, or the extent to which they are based on speculation.

Reflect careful fact-checking. Few things undermine the credibility of an evaluation report more than errors in descriptions of the activities and errors in reporting findings. Therefore, authors of these reports are well-advised to review drafts with great care and, whenever possible, to ask someone who was not involved in preparing the report to review and comment on the draft.

A final consideration: One evaluation report or several reports?
Comprehensive, long-term professional development activities and initiatives probably suggest the need for several reports. If the professional development includes several cycles of learning activities spread out over four to six months or longer, and if there are clearly defined interim benchmarks specified in the overall plan, it is a good idea for the evaluators to prepare several reports. For example, if the activity extends over two school years and there is an expectation that teachers will begin using new instructional practices during the first year, an interim report prepared over the summer between the first and second year of professional development could examine the first-year professional development activities and initial implementation of new practices. A final report, which continues to track the professional development in the second year and examines ongoing changes in instruction, could be completed in the summer after the second year.

Preparing multiple evaluation reports has several advantages. First, interim reports can help providers and managers determine whether the activities are taking place as planned and whether they appear to be achieving the intended outcomes. Findings in these reports can also inform mid-course corrections that increase the likelihood that the overall outcomes will be achieved. Interim reports make it possible to provide information about results—albeit incomplete information—to policymakers and funders sooner. Depending on the circumstances, these results may be useful in leveraging additional resources for the professional development. To be sure, interim reports that indicate that the activities are not proceeding smoothly or that they are not producing the expected results could undermine support for the professional development. At the same time, knowing what does not work or contribute to improved practice is almost as important as knowing what does work.

Appendix A: Resources for Evaluating Teacher Professional Development

The resources listed here provide conceptual models, guidance on developing instruments and conducting evaluations, and samples of data collection instruments (especially instruments for collecting various kinds of qualitative data.) These resources should be considered as a starter set; they by no means represent a comprehensive sample. In addition, as evaluation planners review the various examples of instruments and data collection activities, they should recall the importance of tailoring the instruments and data collection activities to meet the needs of the evaluations that they are planning.

Administration on Children, Youth, and Families. (no date). *The program manager's guide to evaluation.* Washington, DC: U.S. Department of Health and Human Services.

This guide and a companion guide, *The Headstart Bureau Evaluation Handbook,* are intended to help program staff plan and conduct local evaluations. Both guides include tools that can be used in local program evaluations. The guides also explain how programs can use evaluation results to improve program design and services to children and their families.

Blank, R.K., de las Alas, N., & Smith, C. (2008, February). *Does teacher professional development have effects on teaching and learning? Analysis of evaluation findings from programs for mathematics and science teachers in 14 states*. Washington, DC: Council of Chief State School Officers. Retrieved August 19, 2008, from http://www.ccsso.org/content/pdfs/cross-state_study_rpt_final.pdf

This report examines findings from a number of evaluations of professional development for teachers across the country. It provides extensive information about evaluation designs.

Council of Chief State School Officers, & National Science Foundation. (2008, April). *Bridging the gap between research & practice in professional development: Selected research and resources*. Washington, DC: Authors.

This report is a companion to the review of evaluation results. It includes extensive lists of resources that are available for evaluating professional development as well as information about how to locate these resources.

Curriculum Services Canada. (2007, September). Teacher moderation: Collaborative assessment of student work (The Literacy and Numeracy Secretariat).

This publication describes how teachers can work together to assess student work as part of larger efforts to understand and improve instruction. A 90-minute video is also available to illustrate various parts of the process.

Dirr, P.J. (no date). *Classroom observation protocols: Potential tools for measuring the impact of technology in the classroom* (Policy and Planning Series #104). Alexandria, VA: Appalachian Technology in Education Consortium with support from The CNA Corporation. Retrieved August 19, 2008, from http://www.eed.state.ak.us/edtech/pdf/ATEC-PP104Tools.pdf.

This report explains how to use various kinds of observation protocols in collecting information on the impact of interventions in classrooms. The report also contains information about a number of well-designed protocols that can be used for this purpose, including their use in pre-/post evaluation designs.

Elmore, R. (2007, April). Professional networks and school improvement. *School Administrator*. Available online from http://www.aasa.org/publications.

This article describes how observations that yield descriptions rather than judgments are important in understanding classroom practice.

Frechtling, J., Frierson, H., Hood, S., & Hughes, G. (2002, June). *The 2002 user-friendly handbook for project evaluation*. Washington, DC: The National Science Foundation.

This handbook lives up to its title. It explains various approaches to program evaluation in easy-to-understand language.

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

Prepared for the National Staff Development Council, this book describes five "levels" of professional development evaluations and explains how to conduct evaluations at each level and the relative benefits of the results that are generated. The book also contains an exhaustive list of references of other resources for professional development.

Henry, M.A., Murray, K.S., & Phillips, K.A. (2007, November 30). *Meeting the challenge of STEM classroom observation in evaluating teacher development projects: A comparison of two widely used instruments*. St. Louis, MO: M.A. Henry Consulting, LLC.

This report discusses the challenges associated with using classroom observations in evaluations of teacher professional development and it offers concrete suggestions for overcoming the challenges. It also reviews two observational tools, including their reliability and validity.

Killion, J., & Harrison, C. (2006, June). *Taking the lead new roles for teachers and school based coaches*. Oxford, OH: National Staff Development Council.

This book, which focuses on the multi-faceted role of school-based professional development, includes detailed suggestions for evaluating their impact in schools and classroom.

King, J.A., Morris, L.L., & Fitz-Gibbon, C.T. (1987). *How to assess program implementation*. Newbury Park, CA: SAGE Publications, Inc.

Part of the nine-book *Program Evaluation Kit,* this small volume provides detailed guidance on how to plan and conduct evaluations of program implementation. Discussions of data collection and data analysis are especially helpful and there are a number of sample data collection instruments that can be adapted for evaluations of teacher professional development.

Kirkpatrick, D. (1994). *Evaluating training programs*. San Francisco, CA: Berrett-Koehler Publishers.

One of the classics on program evaluation, this book introduces the notion of various levels of evaluation of training programs and explains the uses of each level of evaluation.

McDonald, J.P., Mohr, N., Dichter, A., & McDonald, E.C. (2007). *The power of protocols: An educator's guide to better practice*. New York: Teachers College, Columbia University.

The authors explain how carefully crafted protocols can be used to structure discourse about practice, to collect, organize, and analyze data, and to examine student work.

Patton, M.Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA: SAGE Publications, Inc.

Another in the *Program Evaluation Kit,* this book explains how qualitative methods can and should be used in evaluations. In addition to guidance on designing qualitative evaluations, it offers extensive guidance on conducting observations and interviews and analyzing qualitative data.

Puma, M., & Raphael, J. (2001). *Evaluating standards-based professional development for teachers: A handbook for practitioners*. Washington, DC: The Urban Institute. Retrieved August 27, 2008, from http://www.urban.org/publications/410432.html.

This detailed guide provides step-by-step directions for evaluating what the authors call standards-based teacher professional development. The guide also develops cases of two district evaluations to illustrate different approaches to evaluation.

Singer, N.R., & Scollay, D. (2007). *Increasing student achievement in writing through teacher inquiry: An evaluation of professional development impact*. Retrieved August 8, 2008, from http://www.nwp.org/cs/public/download/nwp_file/10561/Gateway_LSRI_Cohort_ll.pdf?x-r=pcfile_d.

This report is a very good example of a rigorous local evaluation of teacher professional development in the area of writing instruction.

Taum, A.K.H., & Brandon, P.R. (2006, April 9). *The iterative process of developing an inquiry science classroom observation protocol*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

This essay offers clear guidance and examples for developing classroom observation protocols.

W.K. Kellogg Foundation. (1998, January). *W.K. Kellogg Foundation evaluation handbook*. Battle Creek, MI: Author. Retrieved September 3, 2008, from http://www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf.

Long a leader in the field of program evaluation, the W.K. Kellogg published this guide to help program leaders and staff plan and conduct rigorous evaluations. The guide also includes case studies of how programs have used evaluation results to enhance program quality and outcomes.

Resources for Evaluating Teacher Professional Development

Planners should conduct their own searches for instruments and data collection strategies that may be appropriate for their evaluations. Use the following search terms and others like them to locate helpful documents on the Internet.

- Looking at student work
- Assessing/evaluating program impact
- Measuring/assessing teacher knowledge **(Add a content area to the search terms)**
- Evaluating training (and/or training programs)
- Program evaluation/impact evaluation
- Logic model/theory of change

Appendix B: Sample Items for Surveying Participants' Views of Professional Development

The following items can be adapted for use in surveys of participants' views of their professional learning experiences.

1. Understanding the purpose of professional development

Sample item: *Which of the following statements best describes the primary purpose of* **(Insert name of professional development activity)**? (*Select one.)*

The purpose of the activity was:

(a) To communicate some new ideas for me to consider using in my classroom
(b) To provide an opportunity for me to learn from other teachers
*(c) To help me understand (**Insert content of professional development**)*
*(d) To help me apply/implement (**Insert content of professional development or other descriptor**) in my classroom*
(e) Not clear
(f) Other (Specify)

Note: In using this or a similar item, be sure that one response option includes the intended purpose of the activity.

2. Ratings of the usefulness of key components of the professional development

Sample item: *Which of the following statements best describes the usefulness of* ***(Insert the name of the professional development or a specific component)*** *(Select one.)*

(a) It was a good start.
(b) It was a good start, but I have a lot of questions.
(c) It was a good start, and I look forward to using the new ideas in my classroom.
(d) It provided everything I need to use the new ideas in my classroom.
(e) I don't think that these ideas will work very well in my classroom.
(f) It's too soon to tell.

Sample Items for Surveying Participants' Views of Professional Development

3. Perceptions of the extent to which the professional development met participants' needs

Sample item: *Indicate the extent to which* ***(Insert name of professional development)*** *met your professional learning needs. (Select one.)*

(a) It addressed my professional learning needs completely.
(b) It addressed some of my professional learning needs.
(c) It did not address my professional learning needs.
(d) This activity did not help much because I was already familiar with this topic.

4. Ratings of the alignment of the content of the professional development with improvement priorities

Sample item: *To what extent was* **(Insert the name of the professional development)** *aligned with your* **(school/district) (goals/priorities)** *for improving instruction? (Select one.)*

(a) The activity was very closely aligned with ***(goals/priorities)*** *for instructional improvement.*
(b) The activity was somewhat aligned with ***(goals/priorities)*** *for instructional improvement.*
(c) The activity was not aligned with ***(goals/priorities)*** *for instructional improvement.*
(d) The activity was inconsistent with ***(goals/priorities)*** *for instructional improvement.*
(e) I don't know.

5. Perceptions of support and encouragement to participate in the professional development

Sample item: *Which of the following statements best describes the support that you received from your principal* ***(or other school leader or school-based professional development staff)*** *to participate in* ***(Insert name of professional development)****? (Select one.)*

(a) The principal strongly encouraged me to participate.
(b) The principal encouraged me to participate.
(c) The principal tried to discourage me from participating.
(d) I did not discuss the activity with the principal prior to participating.

Sample Items for Surveying Participants' Views of Professional Development

6. Perceptions of support and encouragement to apply new knowledge and skills

Sample item: *Which of the following statements best describes the support that you received from your principal to apply what you learned in* ***(Insert name of professional development)*** *in your classroom? (Select one.)*

(a) The principal has encouraged me to apply what I learned in my classroom.
(b) The principal has encouraged me to apply what I learned in my classroom and has offered to help.
(c) The principal has not encouraged me to apply what I learned in my classroom.
(d) I have not discussed what I learned with the principal.

Note: Additional items and response can focus on encouragement from other school leaders, school-based professional development staff, and other teachers.

7. Ratings of the likelihood of applying new knowledge and skills in the classroom

Sample item: *Which of the following statement best describes the likelihood that you will apply what you learned in* ***(Insert the name of the professional development)*** *in your classroom? (Select one.)*

(a) I have already (practiced/applied) (skill/practice) in my classroom
(b) I have already (practiced/applied) (skill/practice) in my classroom, and it seemed to work well
(c) I have already (practiced/applied) (skill/practice) in my classroom, but it was not appropriate for my students
(d) I look forward to (practicing/applying) (skill/practice) in my classroom in the next few weeks.
(e) I look forward to (practicing/applying) (skill/practice) in my classroom sometime later this year.
(f) I would like to (practice/apply) (skill/practice) but I don't have the materials that I need.
(g) I don't think that these things will work with my students.

8. Overall ratings of the usefulness of the professional development compared with other professional development

Sample item: *Which of the following statements best describes how* ***(Insert the name of the professional development)*** *compares with other professional development in which you have participated during the past six months* ***(or other period)****? (Select one.)*

(a) This activity was more useful than other professional development that I have participated in.
(b) This activity was about the same as other professional development that I have participated in.
(c) This activity was less useful than other activities that I have participated in.
(d) I don't have an opinion.
(e) I don't have an opinion because I haven't participated in any other professional in the last six months.