

Evaluating Science and Mathematics Professional Development Programs

Iris R. Weiss

The design of a professional development program can be considered a series of hypotheses. Based on their understanding of the needs of a particular group of teachers, and their knowledge of what “works” under a particular set of circumstances, professional development providers plan activities that they believe will result in the desired outcomes: increased teacher knowledge, improved pedagogical skills, etc. Evaluation provides a way to test those hypotheses, providing information that can be used both to improve the program during its implementation, and to judge the effectiveness of the program.

This essay provides a brief overview of some of the issues involved in evaluating professional development programs for science and mathematics teachers, and describes a number of resources that can be used for further investigation.

Why evaluate professional development programs?

One reason to evaluate professional development programs is that you have no choice; funders are increasingly likely to expect that the programs they support provide evidence of their quality and effectiveness. A more satisfying reason for evaluating professional development programs is because you want to make the program the best it can be, and to provide information that others can use to improve their programs as well.

When should the evaluation take place?

Evaluation tools and techniques can be applied at various stages in a program. In fact, one of the most important uses of evaluation is the one most often overlooked: helping project staff refine the plan before the project begins in order to increase the likelihood of success. People who design professional development programs for science and mathematics teachers are trying to address a great many needs with limited resources, and it is easy to overlook things in the process. By asking project staff what they are trying to accomplish, and having them describe the strategy for reaching those goals in detail, evaluators can help identify areas of fuzzy or wishful thinking. It is very common, for example, for projects to list as goals both improving teacher content knowledge and improving teacher understanding of pedagogy, but then to design a program that addresses one of them well and the other only superficially. Or a program can call for teacher leaders to assist other teachers in their instruction, but not make adequate provision for released time for them to do so. This “design critique” function can be an extremely cost-effective use of evaluation, leading to improvements in the design before commitments have been made and resources have been expended on implementation.

A second, often under-utilized function of evaluation is to assess the quality of project activities and provide feedback to project staff. Sometimes called formative evaluation, and other times implementation evaluation, this process allows staff to make mid-course corrections, improving

the likelihood that the project will achieve its goals. For example, in a program with teams of scientists and lead teachers working with groups of teachers, the evaluation may identify a great deal of inconsistency in the quality of implementation, and even in the vision of what the various professional development providers are trying to accomplish. Armed with this information, project staff may decide to bring the teams together to discuss the purposes of the program and guidelines for effective implementation, and to design in a more extensive orientation for teams in the future.

A third function of evaluation, and the one people typically think of as “evaluation”, is to assess the extent to which a project has in fact achieved its goals. For example, evaluators might look for evidence that teachers have increased their knowledge and skills or improved their classroom practice as a result of participating in the professional development. To address this question, evaluators need to be able to compare the end-of-program status to what it would have been if the program had not been implemented, which often involves collecting quantitative and qualitative data both before and after participation.

What information is needed in order to evaluate a professional development program?

Armed with an understanding of the project goals, and the strategies to be used to achieve them, evaluators can begin to design an evaluation to provide information about the quality of the program’s implementation and its success in achieving its goals. Project staff need to be involved in this process as well, to ensure that the evaluation addresses their most important questions, and that the evaluation process will result in information that will be useful both to them and to other stakeholders.

For example, many professional development programs work with teachers who volunteer to participate. While project staff’s primary interest may be in providing a high-quality experience to meet the needs of participating teachers, the district that is supporting the program might be even more interested in finding out whether the teachers who needed the most help were signing up, as opposed to those who were already most skilled.

Possible questions to be used in the evaluation of professional development programs

How well does the program design adhere to standards of best practice in mathematics/science professional development?

To what extent is the program implemented as planned?

What is the extent of teacher involvement in the professional development program? Are the teachers most in need of the program participating fully?

How do teachers feel about the professional development activities? Do they perceive them as relevant to their needs? Have they received the kinds and extent of support they feel they need in order to implement changes in their classrooms?

What is the impact of the professional development program on teachers' attitudes, knowledge, and skills?

What is the impact of the professional development on classroom practice?

When teachers change their practice as advocated in the professional development, what is the impact on student attitudes, knowledge, and skills?

Once the big questions to be addressed are determined, the next step is to decide what to look at in order to answer each question, and then how to collect the necessary data. For example, in assessing the quality of efforts to deepen teachers' understanding of mathematics/science content, an evaluator might consider whether appropriate time and emphasis was given to disciplinary content; the extent to which the content was matched with teacher needs; and whether the content was presented accurately and accessibly. Answering those questions might require reviewing the project plans and session agendas, as well as observing a number of professional development sessions to see how teachers were engaging with the content. In contrast, questions about the impact of the professional development on classroom practice might best be addressed by administering surveys, observing classrooms, and interviewing teachers.

It is important to note that there is no one "best way" to evaluate professional development programs. The basic "building blocks" of any evaluation are similar -- data collected by document review, surveys, interviews, and observations -- but they can be combined in many different ways to answer the same evaluation questions. For example, in order to assess the impact of a professional development program on teacher knowledge of science one could:

- (1) administer pre- and post-tests (multiple choice, essay, and/or performance tasks, with responses provided in writing or orally);

- (2) ask a teacher to listen to student dialogue or review student work to identify areas where student understanding of content was incorrect or incomplete;
- (3) use surveys to ask teachers about their understanding of particular content areas, both before and after the program;
- (4) ask teachers to describe how the program had impacted their content knowledge;
- (5) review teacher lesson plans to assess whether the teacher understood which are the “big ideas” and which are the supporting detail; and/or
- (6) observe classes for the same purpose as well as to assess the depth of the teachers’ understanding of the content.

Working in concert with project staff, the evaluator needs to choose a set of activities that will provide the necessary information within the constraints of time and available resources. Since each type of data collection has inherent limitations, it is usually a good idea to use multiple methods in addressing important questions. For example, surveys can tell you “how many” teachers say they are using a particular approach or set of materials, but typically can not tell you how well. In addition, many people have concerns about the accuracy of self-report data. In contrast, observations can provide richer, more “convincing” data, but, given resource limitations, typically target far fewer people. As a result, it is unlikely an evaluation would be able to report observation results separately for subgroups of teachers in various grades, or rural, urban, suburban schools.

For the results to be useful and credible, each data collection method has to be used appropriately and well. For example, if surveys are used, items need to be worded clearly and unambiguously; samples need to be both representative and large enough to support the intended analyses; and response rates have to be high enough to provide confidence that the findings are not biased. Observations need to be carried out by people who can be trusted to do so objectively, and who have the knowledge and experience needed to assess the quality of what they are seeing and hearing.

Who should conduct the evaluation?

Depending on the purposes, evaluation can be carried out by people who are considered part of the project, or people who are completely external to the project. The most important criterion is their competence to design and implement an evaluation; the evaluators need to understand what the project is trying to accomplish, and be able to collect, analyze, and interpret data to determine how well the project is progressing and the extent to which its goals are being met. For example, if a project is aimed at helping teachers provide mathematics instruction consistent with the NCTM Standards, and the evaluation calls for classroom observations, the evaluators will need to have a deep enough understanding of mathematics content and how the Standards define quality instruction to be able to recognize when a lesson is and is not “standards-aligned.” Another evaluation plan might require sophisticated statistical expertise to determine the impact of the professional development on student achievement, but not require the evaluators to be able to recognize standards-based instruction when they see it.

The second important criterion is credibility. If the primary audience for the evaluation is the program itself, then having people associated with the program provide the information can work very well and at a lower cost than hiring consultants. On the other hand, if the evaluation is aimed at providing information to external audiences, such as funders, findings provided by people closely associated with the program may not be credible. Some professional development programs address these issues by having a team of internal and external evaluators, providing the advantages of both continuous feedback and credibility of findings.

What does it cost to evaluate a professional development program?

Depending on the complexity of a professional development program, and the extent to which new data collection instruments must be developed, evaluations can be pretty costly, especially if you plan to deploy teams of evaluators collecting a variety of data from multiple sources. At the same time, it is important not to go overboard in designing the “perfect” evaluation. Just as it makes little sense to provide professional development services without a way of knowing whether they are effective, it makes little sense to devote a lion’s share of the program’s resources to finding out if it is working and have little left to provide the designated services to teachers.

A general guideline is that 5-10 percent of a project budget should be devoted to evaluation, although small projects may find it impossible to do even a cursory evaluation for that amount (10% of a little is very little), while large projects that are providing essentially the same kind of professional development services to multiple groups of teachers can sometimes have an excellent evaluation for less than 5 percent of their budget.

Where would I find additional information about evaluating professional development programs?

There are a number of resources for learning more about evaluating professional development projects. Some of these will be helpful to project staff in becoming more savvy consumers of evaluation; others may be more helpful to evaluators, people who are actually engaged in designing and implementing evaluations.

The “User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering and Technology Education” was developed for Principal Investigators and project evaluators working with the National Science Foundation’s Directorate for Education and Human Resource Development. The authors note that the Handbook “builds on firmly established principles, blending technical knowledge and common sense to meet the special needs of NSF’s programs and those involved in them.” In addition to descriptions of different types of evaluations and data collection methods, the Handbook provides examples of project evaluations, including an in-service program for elementary science teachers, and guidelines for selecting project evaluators.

A website maintained by Horizon Research, Inc. (www.horizon-research.com) describes the evaluation of NSF's Local Systemic Change (LSC) initiative, a program that involves more than a hundred school districts nationally. The evaluation includes observations of classrooms and professional development sessions, as well as teacher and principal questionnaires and teacher interviews. All of the instruments used in the LSC evaluation may be used for other education and research purposes, with appropriate attribution; they may not be used for commercial purposes.

SRI International has developed an Online Evaluation Resource Library (www.oerl.sri.com) for evaluators of science and mathematics education projects. Materials include sample evaluation plans, instruments, and reports; criteria for judging the quality of these materials; and guidelines for their use. Users can search the database for particular kinds of projects/materials using pre-selected categories or with a keyword search.

Horizon Research, Inc.
326 Cloister Court
Chapel Hill, NC 27514

1999