# EVALUATION 101

**Written by**

## REBECAH BUSSELLE

The process of program evaluation for professional development (PD) can be broken into two main phases — the planning phase and the implementation phase. The planning phase is a time for rich collaboration between the evaluator and the key program stakeholders. In this phase, the evaluator provides the scaffolding and the program team provides the substance — the deep program knowledge required to map out a meaningful evaluation. The evaluator completes much of their work during the implementation phase, with the continued input and advice from the program team.

### PLANNING PHASE

- Form an Evaluation Team
- Connect stakeholders
- Build a logic model
- Identify goals & evaluation questions
- Determine the evaluation design

### IMPLEMENTATION PHASE

- Identify and group participants
- Collect & analyze data
- Report findings
- Respond to findings

**Terms**

**Instrument:** general term for a measurement device (e.g., survey, test, questionnaire, etc.)

**Validity:** the extent to which an instrument is measuring what it is supposed to measure

**Reliability:** the extent to which an instrument consistently measures what it is intended to measure

**External validity:** the extent to which the results of a study can be generalized from a sample to a population

**Outcomes:** measurable changes in individuals or systems

**Impacts:** broad, long-term changes to individuals or systems

**Indicator:** a measure of implementation and/or results (often called a metric or measure)

**Target:** the value of an indicator expected to be achieved at a specific point in time

01

## FORM AN EVALUATION TEAM & SELECT ADVISORS

**The first step in planning** a PD program evaluation is to form an Evaluation Team. The Evaluation Team should be led by someone with the skills and experience to conduct a good evaluation and to interpret the findings. However, the best program evaluations do not occur in isolation. During the course of planning and implementing program evaluations, questions and challenges invariably arise. The expertise of a PD program's stakeholders can make or break the success of an evaluation. The Evaluation Team should engage stakeholders during the Planning Phase and keep them actively updated and consulted throughout the evaluation process.

Because many PD programs don't have the technical expertise to conduct a good program evaluation, you might need to hire external evaluators, which can be costly. Or, depending on your needs and internal staff's capacity and expertise, you may be able to rely on internal evaluators or, internal evaluators with the support of an external consultant. The decision

of whether to hire external evaluators might be a decision the program managers make independently, or it may be dictated by the funding requirements of a grant. The Resources for Evaluation Team Selection toolbox is a useful guide for making these decisions. If you wish to use evaluation findings to seek future program funding, your best option is to use an external lead evaluator. These costs should be considered when budgeting for professional development programs. For example, it's common in grant proposals to see 15–20% of the program's budget allocated to program evaluation.

In addition to members with program and technical expertise, your Evaluation Team might also include representatives from groups with a vested interest in the PD program (e.g., funders, professional learning providers, district leaders, policymakers, education departments, teacher unions, school officials). Engaging these stakeholders in the work of determining evaluation goals, questions, logic, and methodology is critical to the success of an evaluation. Anyone who could help (or disrupt!) your evaluation — before, during, or after it is conducted — should be engaged in the planning process. It is essential for the Evaluation Team to know why stakeholders care about the program, what they hope the program will accomplish, what they would need to see to say the program was successful, and what resources they can bring to the table.

## Resources for Evaluation Team Selection

| Y | N | |
|---|---|---|
| ☐ | ☐ | 1. Does your program have funds designated for evaluation purposes? |
| ☐ | ☐ | 2. Have you successfully conducted previous evaluations of similar programs? |
| ☐ | ☐ | 3. Are existing program practices and information collection forms useful for evaluation purposes? |
| ☐ | ☐ | 4. Can you collect evaluation information as part of your regular program operations? |
| ☐ | ☐ | 5. Are there program staff who have training and experience in evaluation-related tasks? |
| ☐ | ☐ | 6. Are there advisory board members who have training and experience in evaluation-related tasks? |

If you answered:

- **NO to all of the questions,** consider postponing your evaluation until you have sufficient funds to hire an outside evaluator.

- **YES to Question 1, but NO to the rest,** hiring an outside lead evaluator with support from an in-house team is your best option.

- **NO to Question 1 but YES to most of the rest,** an in-house evaluation team is an appropriate choice.

- **YES to Question 1 and the remainder of your answers are mixed YES/NO** then you should hire an outside lead evaluator and could use in-house or external partners to complete the team.

Adapted from the Administration of Children and Families.
(https://www.acf.hhs.gov/opre/resource/the-program-managers-guide-to-evaluation-second-edition)

## 02

# BUILD A LOGIC MODEL

**The Evaluation Team should be clear** on what they think the PD program will accomplish and how these accomplishments will come to be. A common and effective way of communicating this information is to create a logic model. A logic model is a graphical representation of the logic behind what a program does and how it works. For example, imagine a school district wants to implement a PD program for teachers of grades 3–5 that will increase implementation of Next Generation Science Standards (NGSS) and integration of NGSS with Common Core State Standards (CCSS) in their classrooms. The logic model for this program might look like this:

## Explanation of Terms

**Inputs** are the resources needed to implement a program. **Activities** are what the program actually does. **Outputs** are the results expected from the inputs and the activities. The outputs drive **outcomes,** which are measurable changes in individuals or systems. Outcomes might be observed close in time to the activities of interest **(short-term outcomes)** or they might be observed much later **(intermediate outcomes).** Over time, a program's outcomes contribute to **impacts,** which are broader, long-term changes to individuals or the system.

| Inputs | Activities | Outputs | Outcomes Short-Term | Intermediate | Impacts |
|---|---|---|---|---|---|
| Substitute Teachers | Full Day (6 hrs.) of PD instruction focused on NGSS implementation & NGSS-CCSS integration in grades 3–5 | All teachers in grade 3–5 classrooms in the district participate in 48 hours of NGSS-CCSS professional development each academic year | Increased teacher understanding of NGSS practices | Increased NGSS implementation in grade 3–5 classrooms | Higher student achievement in science on state standardized tests |
| PD Instructors | | | Increased teacher science content knowledge | Integration of CCSS and NGSS in grade 3–5 classrooms | Increased teacher retention |
| PD Materials | PD offered four times each academic year | | Increased understanding of connections between NGSS and CCSS | More class time devoted to science instruction in grade 3–5 classrooms | Greater student engagement in middle school science courses |
| Participation Incentives | | | Increased teacher confidence in implementing NGSS | | |

An evaluation can't move forward without clarity around a program's goals, resources, services, and measurable outcomes. Creating a logic model together with stakeholders helps build consensus around program and evaluation goals and beliefs. For example, one stakeholder might be singularly focused on student achievement on standardized tests as a metric for program success. Having a logic model helps this stakeholder see the many outcomes that must first be achieved before changes in student achievement could be reasonably expected as a result of the PD program. Logic models are also often critical for communicating the rationale behind the evaluation design and the interpretation of the results.

## 03

# IDENTIFY THE EVALUATION GOALS AND QUESTIONS

**There are many reasons** for conducting a PD program evaluation. Being explicit about what your most important goals are keeps the evaluation focused. Evaluation goals are the reason you are conducting the evaluation. Your goals might be related to understanding how (or if) the PD program is working or to identifying what the PD program is accomplishing.

Some examples of evaluation goals include:

- To determine a way to measure teachers' understanding of connections between NGSS and CCSS
- To determine if participating teachers devote more time to science instruction
- To determine if participation in 48 hours of NGSS-CCSS PD each academic year leads to increased teacher science content knowledge
- To determine how often PD should be offered to achieve desired levels of NGSS-CCSS integration in classrooms
- To determine if the PD program increases teacher understanding of NGSS practices and teacher science content knowledge
- To determine the average number of hours of NGSS-CCSS PD teachers are participating in over the course of the academic year

Your evaluation questions determine what data will be collected about which aspects of the program. Attempt to answer no more than 5–7 specific questions. Our **Evaluation: Methods** paper provides greater insights into the kinds of questions most often asked when evaluating teacher professional development.

You should know that your PD program meets basic standards of high quality professional development and that it can reliably deliver its activities before trying to evaluate its outcomes and impacts. If program activities are not reliably delivered, there will be variations in the outputs produced and thus variations in the resulting outcomes. If Evaluation Teams do not know that programs can be reliability implemented they should first conduct a **program monitoring evaluation** to determine if sufficient resources have been allocated to implement PD effectively; the best ways to deliver the PD; how faithfully the PD program is being implemented; and the obstacles to successful program implementation. Program monitoring evaluation reports are used to refine a program as it is being implemented. This feedback loop between evaluation and program improvement drives PD programs to maximize their efficiency and efficacy.

Once a PD program is confirmed to be working as intended, it is time to figure out what impact it is having. **Impact evaluations** help determine if PD programs move the needle on student achievement, student engagement in advanced courses, student beliefs, teacher efficacy, and teacher retention. It's important to note that impact evaluations also often yield data that's useful for PD program improvement.

## 04
# DETERMINE THE EVALUATION DESIGN

**The last step in the Planning Phase is to map out the evaluation design.** This means defining what data is needed to answer each evaluation question and planning the data collection. There are two main approaches — an experimental design or a descriptive design — and countless variations on these approaches. An experienced evaluator will know how and when to best employ these approaches to answer your evaluation questions. Evaluation designs often require splitting program participants into groups, only some of whom participate in the PD program. This can seem unfair to teachers or schools who do not receive PD. So it's helpful for all stakeholders and everyone on the Evaluation Team to have a basic understanding of the rationale behind these evaluation designs.

| Experimental design | Descriptive design |
| --- | --- |
| Experimental design is chosen when an evaluation question asks if a causal relationship exists between an input/activity and an outcome. For example, consider the question, "Does participation in 48 hours of NGSS-CCSS PD each academic year lead to increased teacher understanding of NGSS practices?" In order to answer this question, you need to compare teacher understanding of NGSS practices between teachers who participated in 48 hours of this PD with teachers who didn't. This allows you to make a causal claim about the relationship between participation in the PD and understanding. That is, you would have solid evidence to say that participation led to increased understanding. Experimental evaluation design typically requires random assignment, two different program implementation plans, and data collection for both teacher populations. | With descriptive design the goal isn't to make causal claims, but rather to describe what's happening. Descriptive evaluations are most useful for program monitoring. For example, descriptive design is appropriate for evaluation questions such as: are teachers attending the PD, why aren't teachers attending the PD, and are teachers scores on content assessments higher after the PD compared to their scores before? The first question would simply require collecting attendance data. The second would require asking non-participating teachers why they didn't attend. The third question would require administering content tests before and after the PD and comparing the results. None of these evaluation questions is asking about causal factors, so using an experimental design is unnecessary. |

## 05
# IMPLEMENTATION PHASE

**The final phase of a PD program evaluation plan** is the enactment of the evaluation design. Participants are identified and assigned to different groups (if needed), instruments for collecting data are identified or developed, data is collected and analyzed, and reports are written and distributed. While stakeholders are often consulted during this process, the Evaluation Team leads the majority of this work.

For program monitoring, evaluation reports should be presented to the stakeholders who can use the results to make improvements to the program. With impact evaluation reports, stakeholders are often tasked with determining the fate of the program — to continue allocating resources to implementation, to improve the program, or to discontinue funding. Being clear about how evaluation results will be used in the Planning Phase helps everyone involved understand what they stand to gain (or lose) from conducting a thoughtful, well-planned program evaluation.