CHAPTER

# 1

# INTRODUCTION

*The role of the program director is to believe,*
*the role of the evaluator is to doubt.*

—Carole Weiss



## LEARNING OBJECTIVES

After reading this chapter, you should be able to

- Explain the primary and secondary goals of evaluation
- Describe the steps of the evaluation framework

An executive director of a medium-sized, nonprofit social service agency once told me, "I know we're doing good—I can see it in our clients' eyes." I wish that this was a unique encounter, but I have repeatedly had people responsible for running programs and delivering services to clients tell me some version of the above statement. Sometimes there is a variation on the statement—"Funders wouldn't fund us if we weren't doing good," or from the funders' perspective, "They must have done something good since they spent all of the money in the way that they said they were going to." Sometimes, administrators view increases in program enrollment as evidence of program success.

Unfortunately, none of these statements tells us what would have occurred to clients if the program hadn't been in existence. Clients may have been just as successful, or even more successful, if left to their own devices.

In all fairness, if those who deliver services didn't believe that they were effectively improving people's lives, they might have trouble delivering the services. It is difficult to devote oneself to a job if one doesn't believe that one's efforts have a positive effect.

The primary job of the evaluator is to examine rigorously the impact of interventions. However, a secondary job is to persuade those who believe that they can determine their effectiveness by looking into their clients' eyes to be more critical of the role that their services play in their clients' lives. Although clients' eyes may reflect gratitude, they cannot reveal whether a program reaches its intended **target population,** or whether clients would have been just as well off without the program. A critical perspective is needed if the efforts are to continually improve over time.

We are at a cusp of change in the way that programs are funded and delivered. Governmental funding, even for the most basic needs such as housing and food, is drying up, and more reliance is being placed on the nonprofit sector. Increasingly, government is relying on the nonprofit sector, with its army of volunteers and relatively low-paid staff, to compensate for gaps in governmental programs. This became painfully evident during the aftermath of Hurricane Katrina, when short-term federal assistance seemed ineffective, and government officials asked Americans to donate to a host of nonprofit organizations for hurricane aftermath relief initiatives because the federal government did not provide enough support to meet the tremendous need that developed.

However, with more reliance placed on the nonprofit sector comes the demand for all nonprofit organizations (also known in the international context as nongovernmental organizations or NGOs), and especially those that compete with other nonprofits for "market share," to demonstrate their effectiveness. This demand derives from many sources. First, government and consumers would like to know which nonprofits are most effective at providing comparable services. In this respect, evaluation is viewed as essentially a means of providing information to rank programs and organizations. Second, there exists curiosity about the ultimate effectiveness of services and programs; in this regard, evaluation is used to determine which approaches are the most appropriate for remedying a problem or for bringing about change.

Although well trained in service delivery, many in the nonprofit world lack training in the nuances of evaluation research. Likewise, many in the funding world who

have "caught the evaluation bug" lack the background to determine how to form fair criteria by which to judge and reward effective nonprofit organizations and programs.

Working in the background behind service deliverers and funders are people who call themselves evaluators. Many evaluators are primarily trained in a field other than evaluation, such as public health, sociology, economics, statistics, or public policy. They may have had some formal training in evaluation, but they often have moved into evaluation over time (sometimes because of job and market demands) and many have acquired mostly on-the-job training. For this reason, although the majority of evaluators have in common a desire to learn the "truth" about interventions, they bring to the table a divergent set of skills and may place varying emphases on different aspects of programs, policies, and evaluatory practices.

My perspective is aligned with that of Michael Quinn Patton (1996)—the goal of evaluation is to assist with continuous programmatic improvement and introspection. I readily acknowledge that how to achieve this goal is more of an art than a science. However, there is a rigor to the art and there exist tools that the artist needs to have in order to create the most appropriate evaluatory approach for the intervention at hand. Like a painter, the evaluation artist needs to understand the context and size of the canvas. Then, the artist needs to apply colors to the canvas that will result in a pleasing picture.

## THE EVALUATION FRAMEWORK

The box below displays (in a very condensed form) the broad steps one must go through to successfully develop an evaluation of a program. All of these issues are discussed throughout the remainder of this book.

### THE EVALUATION FRAMEWORK

1. Are the evaluation activities formative or summative?
2. What questions are in the universe of possible evaluation questions?
3. Rigorously describe the program by developing the program theory and using a program logic model.
4. Revisit and narrow the universe of evaluation questions.
5. Develop the evaluation plan.
6. Develop a data collection and data analysis plan.
7. Analyze the data.
8. Write evaluation report.

**1. Are the evaluation activities formative or summative?** Summative evaluations answer the question "What was the impact of what was done?" whereas formative evaluation addresses the question "How can the program improve what it does?"

**2. Determine the possible universe of evaluation questions.** Before engaging in any evaluation activities, you should determine the universe of possible evaluation questions, based on informal interviews conducted with program stakeholders.

**3. Rigorously describe the program.** Next, you must describe the program by diagramming its *program theory* (also known as the *theory of change*) and charting out its *program logic model*. These two tools force evaluators to become acquainted with the program, both theoretically and operationally. Many evaluators skimp on learning what the program is all about. I contend that any evaluation that is done without a thorough understanding of the program can be of no constructive use. For this reason, I devote a considerable portion of this book to getting to know the program.

With respect to the program theory, helping program personnel to articulate how they expect change to occur as a result of their program's activities can be quite revealing. Often, personnel have differing views on how and to what extent these changes will occur. Getting everybody on the same page can be a very time-consuming endeavor, and it can take a considerable amount of time for consensus to be reached, but the explication of the program theory ultimately allows the program to operate in a more focused way.

Diagramming the causal chain of the program theory helps you to consider whether, theoretically, the program can have an impact. The program theory also encourages better understanding of which parts of the chain are unknown and thus ripe for research, and which parts of the chain are so based in fact or on proven relationships that it is not necessary to examine or prove that particular link. For example, even though the goal of a program that distributes prescription drugs to low-income people might be improved health, the program does not need to show that taking the drugs under a prescribed protocol improves health—that relationship is known and proven. Instead, the evaluation of the program might focus on the processes utilized to increase drug access to the low-income population.

A well-constructed program logic model reveals strengths and potential weaknesses within the program. Although some program administrators are reluctant to streamline or focus a program, there is value in tightening up a program based on the program logic model. By tightening up a program, I mean slashing anything extraneous from the program logic model. All goals should have activities tied to them; all assumptions should reflect reality; the target population should be described as narrowly as possible; appropriate activities should be in place so that goals are achievable; and the program should have access to enough resources in order to function.

Sometimes, program directors roll up their sleeves and approach an evaluation intending to conduct a full-blown outcome-based quasi-experimental evaluation. However, once the program logic model and program theory are displayed, administrators may want to and sometimes should be encouraged to tweak or revamp the program if they see room for improvement. Rethinking the program should not be perceived negatively.

Seen under the harsh lights of a program logic model, a program may need to be reworked in order to be effective and the outcome-based evaluation may be put off until the program is stable.

**4. Revisit and narrow the evaluation questions.** After rigorously describing the program, evaluators should revisit the universe of evaluation questions and narrow them accordingly. Program administrators and funders should be open to the idea that outcome-based evaluation is not the only type of evaluation that could bring about program improvements. Process-oriented evaluation can also be fruitful. In fact, for unstable programs, process-based rather than outcome-based evaluation will be appropriate.

**5. Develop an evaluation plan.** It is not until the program is rigorously described that one can plan an appropriate evaluation. Certainly, the appropriate evaluation framework cannot be determined from the program title alone. I once became familiar with a program known as the "Class-Size Reduction Program." From the title, one might have expected the number of children in the classroom at any given time to have been reduced by the program. However, this interpretation of the program's activities would have been incorrect. Instead of reducing class size, the program provided a participating classroom with a teacher's aide for about four hours per week. At no point was the number of students in the classroom reduced. In fact, the program increased the number of people in the classroom. After seeing a program rigorously described using a program logic model and program theory, the appropriate evaluation questions and approaches become readily apparent.

Before deciding on an outcome-based evaluation, you should ask the following questions:

- Is the program stable?
- Are the services delivered as intended?

If the answer to either of these questions is "no," then you would certainly examine the processes of the program and put an outcome-based evaluation on hold, at least for now. There would be no point in examining outcomes other than, perhaps, collecting baseline data on the population served.

If you choose to examine *processes* (which would be the case in the event that the program is not stable), then you need to develop a feedback system that would determine not only the extent to which the program is being implemented appropriately and as intended, but also how to use the information on program processes.

If using an outcome-based evaluation approach, you must then decide which outcome measures are most appropriate. After consensus is reached on the relevant measures, the next consideration is which *quasi-experimental design* would best demonstrate that the program actually *causes* changes to occur. The question of what would have occurred to participants if the program did not exist is paramount in this decision; evaluators must be thinking of a "but for" statement—"but for this program, how participants would have fared." The quasi-experimental stage considers relevant "control groups" and the ethics of collecting information from those who are not receiving the benefits of a program.

**6. Develop a data collection plan.** This part of the evaluation plan describes how data will be collected, whether by using primary data collection techniques such as focus groups, interviews, or surveys, or by relying on secondary or pre-existing sources, such as census or archival program data.

If a survey is chosen to collect data, then you need to design and pre-test the survey, come up with a sampling schema, and then implement the survey. If other data collection approaches are used, then you need to consider a data collection protocol so only relevant data are extracted. For example, if client records are available but only a few outcome measures will be used, then the evaluation needs to develop a protocol so that only the relevant information is extracted from the records.

**7. Decide how data will be analyzed.** You also should consider which types of statistical techniques are most appropriate. There are countless decisions to be made when collecting and analyzing data. Sometimes, findings that result from an evaluation are not robust to the analysis approach and methods used. That is, if a different statistical approach was used, the results could change.

This overall decision model must be followed in order to determine the appropriate approach and criteria for evaluation, and I contend that it is impossible to determine these ahead of time. For example, I was once presented with a program that provided health care services to an urban, homeless population. At first glance, and from reading the program's literature, it seemed that appropriate evaluation criteria might involve considering the health status of the homeless. However, after getting to know the program, it became clear that the program's primary goal was to educate physicians and other health care workers about the plight of the homeless and the health ramifications of being homeless. Providing the homeless with on-the-spot medical treatment was a tool used to achieve the outcome that health care professionals would become more empathetic to treating the homeless.

Without a developed blueprint for the evaluation, there is a risk that the evaluation activities will continue ad infinitum. Separating the evaluation activities into seven discrete "tasks" allows evaluators to pause and to think about the direction that their activities should take.

The remainder of this book is devoted to more fully explaining each of these seven steps. By the end of the book, you should be able to carry out many evaluation activities independently. However, if you are faced with a particularly complex challenge that you cannot evaluate on your own, this book will have helped you develop enough expertise to become "educated consumers" of evaluation consultants, and you will be able to determine which consultants are most likely to carry out successful evaluation activities.

The remainder of this book is organized as follows. The first half focuses on developing the appropriate background for evaluation, and the second half addresses the use of quasi-experimentation in an evaluation context. Unlike many evaluation texts, this one emphasizes that one must rigorously describe the program prior to evaluation.

Chapter Two introduces the program logic model and program theory as tools to use when describing the program. The chapter closes with examples of program logic

models and program theories from various types of programs, hoping to make you comfortable enough with these tools so that you will be able to apply them to one of your own programs.

Chapter Three discusses the framing of evaluation questions, elaborating upon the pre-evaluation steps that a good evaluator will complete.

Because all programs exist to cause change in the trajectory of those whom they serve, it is necessary to consider what "cause" means. Chapter Four delves into the meaning of causation in an evaluation context.

Chapter Five explicates "validity" and what it means to do valid evaluation research. Chapter Six introduces the reader to the quasi-experimental designs used, and how trade-offs are made between quasi-experimental designs and types of validity. Quasi-experimental design is used to attribute changes to a program.

Chapter Seven presents issues that arise when collecting data and when designing surveys, and considers sampling approaches. While the chapter does not substitute for a good course in statistics, it does discuss some of the more pertinent points.

Chapter Eight concludes the text, discussing the similarities between grant proposals and evaluation plans, both of which are essential to a good proposal.

## SUMMARY

The primary job of the evaluator is to examine the impact of program interventions. A secondary job is to provide a critical perspective to improve programs. The goal of evaluation is to assist with continuous program improvement and introspection. To successfully evaluate a program, the evaluator must follow the steps of the evaluation framework.

### KEY TERMS

evaluation
evaluators
formative evaluations
program logic model
program theory
quasi-experimental design
summative evaluations
target population

### DISCUSSION QUESTIONS

1. What are the goals of program evaluation?
2. Why is it important to have an evaluation plan?
3. What is the evaluation framework?