Program Evaluation

An Overview

Suppose you are working for the Student Counseling Center at Western Tech in a year or two. The center plans to offer a Sexual Assault Prevention Program to women and men in the following month, and you have been asked to develop a plan that will permit the center’s staff to decide whether (a) to offer the program again because it is successful, (b) to alter the program in order to make it more useful to participants, or (c) to drop the program because it fails to meet a need. You might stop reading for a moment and write down the steps you think are necessary to measure the success of such a program. Among other points, consider what evidence would indicate the program is successful, what would mean that it should be improved, and what would show that it should be ended.

Because the center wishes to maintain and promote useful services to students, evaluating a program that might reduce sexual assaults on campus is an important assignment. The staff of the center would not want to devote resources for a program that does not meet the needs of students on campus; in such a case, something more helpful should be offered instead. If the program is good, you would want your plan to detect its strengths; where it needs improvement, you would want to detect its limitations. When you are finished with this text, you will be ready to tackle this project. To begin, what is program evaluation?

The practice of evaluating one’s own efforts is as natural as breathing. Cooks taste their gravy and sauce and basketball players watch to see whether their shots go in. Indeed, it would be most unwise after turning on the hot water to neglect to check the water temperature before stepping into a shower. At a basic level, program evaluation is nothing more than applying this commonsense practice to settings in which organized efforts, called “programs,” are devoted to helping people who need education, medical treatment, job training, safe streets, welfare assistance, safety while traveling in airplanes, recreational services, or any of the thousands of services provided in a modern society. Program evaluation is a methodology to learn the depth and extent of need for a human service and whether the service is likely to be used, whether the service is sufficient to meet the unmet needs identified, and the degree to which the service is offered as planned and actually does help people in need at a reasonable cost without unacceptable side effects. Utilizing research methods and concepts from psychology, sociology, administration and policy sciences, economics, and education, program evaluators seek to contribute to the improvement of programs.

There are several crucial differences between the natural, almost automatic, evaluations we carry on as we work and play and the practice of program evaluation in organizational settings. These differences make program evaluation harder than the self-evaluations that we all do. First, organizational efforts are nearly always carried out by a team—whether that be a school team of teachers, coaches, and administrators; a hospital team of physicians, nurses, and technicians; or a welfare department team of social workers, counselors, and file clerks. This specialization means that the responsibility for program effectiveness is shared among many people. Furthermore, the results of such programs—an educated student, a well patient, or an effective young mother—are not the sole responsibility of any individual.

Second, most programs attempt to achieve objectives that can only be observed sometime in the future rather than in a matter of minutes, as is the case with a cook adding a spice to a simmering pot or a painter repairing a damaged wall. In other words, as the time between an activity and the desired outcome of that activity lengthens, it becomes less clear what we are to observe in order to decide that the activity is being carried out appropriately or not and what could be done to improve the result of the activity. In fact, the choice of criteria to use in deciding whether a program is working well creates debates among evaluators, program personnel, clients, and funding agencies. How to respond when a problem is detected is often just as controversial.

Third, in contrast to evaluating our own efforts, there are many different parties involved in the evaluations of programs. For example, the effectiveness of teaching in a school may be assessed by evaluators who work for the central administration or the school board. It would not be surprising that individual staff members might feel threatened if the evaluators do not have their trust because, in an extreme case, the livelihood of the teachers might be harmed by an incompetently done program evaluation.

Last, programs are usually paid for by parties other than the clients of a program. Funds for the salaries of nurses are obtained from payments to hospitals by insurance companies and government agencies, and schoolteachers are paid through taxes collected by school districts. Although nearly all nurses and teachers want to do effective work, their job security is at least in the short term more dependent on keeping the program funders satisfied than serving their patients and students well.

Such complications as these make program evaluation considerably more difficult to do than the evaluations of our own activities that we perform so easily and naturally. This text focuses its discussion on evaluations of programs in human services, such as education, medical care, welfare, rehabilitation, or job training. Program evaluation is one form of the general discipline of evaluation that includes employee assessment, quality control in manufacturing, and policy analysis, among other activities (Scriven, 2003). The various kinds of evaluation are sometimes confused with each other. All forms of evaluation involve discovering the value and worth of something; however, the focuses, purposes, and methodologies are quite different.

EVALUATION AREAS THAT NEED TO BE CONSIDERED

A common assumption by people who are not familiar with program evaluation is that all evaluators do is measure the outcome of a program—perhaps counting how many of the clients in a substance abuse treatment program no longer use the substance, either at the end of the program or at a later time, or determining the typical change in knowledge (hopefully an increase) for those who have attended a series of workshops. As will be seen below, measuring outcomes is a very important part of program evaluation. But it is only one part. We will make the case briefly here and then repeatedly throughout this text that program evaluation is a much broader set of activities than just measuring the outcome of a program and that good evaluators consistently think about this wider range of elements. The following paragraphs introduce the seven general areas that program evaluators need to consider. It should be noted that some evaluations may not explicitly include all these elements, and at times a single one of the components may be a complete evaluation (for example, a needs assessment). But choosing not to include one or more elements should be a deliberate decision based on a clear rationale, not a matter of focusing so intently on a limited set of areas that one forgets other considerations.

Meeting Needs

Most often, program evaluations are conducted on existing programs where there is a reasonable likelihood that the programs will continue in somewhat similar fashion for the foreseeable future. As a result, the original need that led to the program may not be raised as an important point of the evaluation; nearly everyone involved may just assume that doing something to address that need is a worthwhile activity. In fact, many people may be unable to imagine not having certain programs. Most of us assume that education is necessary for children and that we must have police and fire departments in our communities without needing to document precise details of what those children do not yet know or of the potential for crime and fires. As you will see in Chapter 6, “The Assessment of Need,” there are a number of important aspects of this component of program evaluation beyond just documenting that there is a legitimate problem that the program addresses.

One overarching point, however, is a focus on unmet needs. Although we all need to breathe air, we do not have large numbers of programs providing air nor evaluations to discuss them. Rather obviously, programs are developed when one or more needs are not already addressed. So evaluators consider unmet needs—whether in an area that has never been addressed or the extent to which current needs are not fully met by existing programs. Various other aspects can also be considered, such as how the need might be expected to change in the future or how the need is perceived differently by different people.

Implementation

A common problem with programs is that some are either never implemented as planned or are implemented in such a way that people in need receive no or minimal benefit. Consequently, it is essential that evaluators confirm points ranging from the fact that program personnel have been hired and offices have been opened to documenting how many clients have found the program and the nature of the services provided. When the primary focus addresses such issues, the project may be called an implementation evaluation, in contrast to an outcome evaluation that also considers the results of the interventions.

Although the monitoring function of program evaluation may at first strike the reader as unnecessary, there are many examples of programs that, although funded, were never put into practice. Even more common, descriptions of programs in agency materials may be accurate only with regard to what was intended, but not what is currently happening. In fact, a recent estimate notes “that failure rates for organizational change implementation typically range from 28% to as high as 93%” (Durand, Decker, & Kirkman, 2014, p. 404). To be fair, the actual practices could be better than originally planned, rather than worse. In any case, an essential area for evaluators is monitoring the details of how the program has been implemented.

Stakeholders

In the process of conducting an evaluation, evaluators might generally expect to have some contact with the clients who are served by the program, especially indirect contact, by considering data from the clients such as information on their changed condition or their responses to a questionnaire. But it is essential that evaluators take into account the many people who are involved with a program in one way or another, those who provide the service as well as those who receive it, and those who are supportive from a distance as well as those whose connection is either mixed or even opposed to some degree. The idea that all of these people have a stake of some form in the program has led to the term “stakeholders” as a general way to talk about them.

Staff providing the services are obviously highly involved both because they deal day-to-day with the clients and because they generally are paid to provide the services. Paying attention to staff as stakeholders involves many points such as recognizing that therapists often know important details about clients beyond what is contained in agency records, that staff may either resist change or be eager for it, and that most employees have some level of apprehension about the results of an evaluation in terms of their own future. Of course, there are a range of staff positions in most agencies—therapists, administrators, support staff, and others. In addition, most agencies have a board that oversees general operations. Often there are others in the community who care about the program in some way—typically including family and perhaps friends of clients, people in the community who are passionate about the specific need being addressed, and occasionally those who are affected by financial aspects of the program. At times, those who live near a treatment facility may have reactions to details of the operation such as traffic or clients’ behavior or still other points.

On the one hand, it is impossible for evaluators to know everything about all stakeholders and completely unrealistic to try to make all stakeholders happy. But on the other hand, it is essential that evaluators make reasonable efforts to learn about the major stakeholders and any important current issues. A common element is that the staff providing services, administrative staff, and board members have different expectations for the evaluation itself. If administrators want to improve efficiency, board members want to lower costs, and therapists want more freedom to extend treatment, there are important implications for the progress of an evaluation and what evaluators should expect regarding cooperation from the stakeholders.

Recently, evaluation researchers have explored the principles that appear to support successful collaboration with stakeholders, such as clarifying the motivation for collaboration (Shulha, Whitmore, Cousins, Gilbert, & al Hudib, 2016) as well as what can be learned about the involvement of stakeholders in evaluation more generally, such as noting that studies have found substantial evidence both for benefits and for liabilities (Brandon & Fukunaga, 2014).

Side Effects

Efforts to solve problems usually create some undesirable side effects (Sieber, 1981). For example, medications that relieve one health problem may create problems in other systems of the body; sometimes these side effects are minor, sometimes they require medical attention or a change in medication, and sometimes, though rarely, they are fatal. Patients and physicians must therefore weigh the dangers of side effects, neither minimizing their importance nor avoiding helpful treatments. Even when the benefits appear clearly to outweigh the costs to medical professionals, others may perceive these effects differently. An important current challenge is that some parents come to different conclusions about the comparative value of vaccinations than physicians (Harmsen, Mollema, Ruiter, Paulussen, de Melker, & Kok, 2013). A parallel situation occurs in the area of educational and social service programs. Welfare policies can help people during crises to reestablish their productive roles in society; on the other hand, welfare policies can create long-term dependency (McKnight, 1995). Programs to provide social support to isolated people can have negative effects that have typically been ignored (Lincoln, 2000). Special education classes permit children with unique needs to learn at a comfortable pace, yet being in such a class carries a stigma. Consequently, program planners seek to develop services that provide benefits but minimize negative side effects. On the other hand, unforeseen side effects are occasionally positive: When people learn new skills, self-esteem often increases (Dawes, 1994; Felson, 1984).

Improvement Focus

Most of these seven areas cover elements of programs that good evaluators consider, measure, and even analyze with formal statistical procedures. The area of improvement focus, in contrast, describes an attitude or approach that evaluators take throughout the course of an evaluation. This pervasive aspect involves considering not just how good the program is now, but how it can become better in the future. The improvement focus is especially important when the evaluation occurs early in the life of a program and is intended to catch and fix any problems. Such evaluations are often called formative evaluations in contrast to summative evaluations that may come at or even after the conclusion of the program. Formative evaluations explicitly look for ways to improve the program, whether by correcting actual mistakes or unanticipated problems or by building on particularly successful components. Even with summative evaluations, however, an improvement focus will make an important contribution whether the program is directly replicated or adapted to fit new settings, circumstances, or methods.

To be clear, this improvement focus does not mean that evaluators are never critical of problems or shortcomings, let alone of failures in a program. Remember that the primary purpose of an evaluation is to determine the value or worth of a program, so if a particular intervention is wasting time or money, or, worse, actively harming clients, evaluators have an obligation to report on these problems clearly. But even with substantial limitations, most programs have value that can be identified and used as a basis for better work in the future. Although it is possible that an evaluation may find there is nothing redeemable in a given program, such conclusions are rare especially when evaluators take the improvement focus seriously.

Outcomes

It should be fairly obvious that an evaluation of a program would consider the results such as how clients improve or other details about how those involved in the program have changed because of their participation. If the proof of the pudding is in the eating, the proof of the program is commonly seen in those who have gone through the program. Those who attend educational programs are expected to know more. Clients in treatment for problems should show fewer (or no) problems; they might also show higher levels of functioning—increased strengths, not just fewer weaknesses. Broader programs such as in public health might be expected to lead to changes at the community level—lower incidence of particular diseases or other problems among the entire population, not just with identified individual. The concept of outcomes is a general one, but the details vary considerably depending on the particular program.

As noted above, for many people examining outcomes is the most obvious, if not the only, aspect they look for in an evaluation. Although it is essential that good evaluators think about the other areas, outcomes need substantial attention, too. Improvement in clients could be seen in a reduction of problems, in an increase of skills, or in both. Clients’ own perception of skill may be important as may observations by others that their abilities have increased. The most appropriate time frame for documenting changes might be at the end of treatment, six months after treatment, or six years later. Increasingly, it is not enough for evaluators to show that those who participated in programs are different following the program—there is an emphasis on showing that the program contained the “active ingredients” that caused the change (Gates & Dyson, 2017). Chapters 8 through 11 will address the many ways that evaluations consider evidence, not just that there are good outcomes, but that the programs are responsible for those good outcomes. As you will see, at times there are at best gentle hints that the program may have led to the changes whereas under the best conditions, there may be very clear reasons to believe that the program caused the good effects. Evaluators examine outcomes and the evidence that those outcomes are caused by the program.

When the specifics about outcomes are clear, the task may still not be finished. There are a number of ways to compare aspects of the outcomes that are important in some evaluations. Although many agencies have one program with a single focus and the goal is simply to have good results from it, others have multiple programs with more sophisticated criteria and where programs are compared with one another in terms of the value of the benefit overall or in terms of the cost for a given benefit. A program with moderately good outcomes may not have good enough results to justify its continuation if resources are threatened. The larger area of outcomes can thus be subdivided into various kinds of comparisons—by cost, by value to the stakeholders, or by other constraints such as how long it takes to achieve a given result. All these ways of examining outcomes are potential areas for evaluators to consider, and will be addressed in a variety of ways later in this text.

Nuances (Mechanisms)

Although evaluations have often aimed simply to document that participants have benefited sufficiently, another growing trend is to examine more sophisticated questions such as specific reasons why the good outcomes have occurred—both in terms of what mechanisms were involved as well as considering how those in different groups or conditions experience better or worse results (Solmeyer & Constance, 2015). Were all the services provided (and funded) essential ingredients in the good results or were some of the activities pleasant for the clients but otherwise irrelevant ways to use staff time, effort, and the funders’ money? Elements of therapy that continue out of habit or tradition rather than because they have been shown to lead to improvements are good candidates to be reconsidered and, potentially, eliminated. Another important point is that some of those receiving services may not benefit like the rest of the group either because of their existing characteristics or because they react to the treatment differently. A treatment that works well with older adolescents but is not effective with the youngest clients might appear to yield good outcomes overall, but understanding the different effects could lead to using a different, but better, treatment for the younger children.

One important aspect of examining mechanisms in an evaluation is that measures of the suspected mechanisms must be included. If therapists or evaluators believe that cognitive restructuring is an essential part of useful therapy, the evaluators will need to assess how much each client successfully restructured thoughts. With such measures, the evaluators can analyze the degree to which clients with greater restructuring tended to improve more. If the effect is consistent and large enough, that would provide evidence to support the use of that component. In addition to specific techniques, a somewhat broader construct such as a measure of the degree to which the planned intervention is followed can also address the question of mechanisms (Abry, Hulleman, & Rimm-Kaufman, 2015). More detailed discussions of these specifics will appear in the sections on data analysis, especially Chapter 11 on experimental approaches.

MISSION

Many evaluators new to the profession find it challenging to remember these aspects of good evaluations. To help students and others, “MISSION” is an acronym for these seven areas—Meeting Needs, Implementation, Stakeholders, Side Effects, Improvement, Outcomes, and Nuances. These seven elements provide an all-purpose set of categories for evaluators to consider regardless of the evaluation. As noted above, it is possible that in a particular case, the evaluator might choose to minimize or even ignore the issue of needs if all stakeholders were already so thoroughly convinced of their reality. But not addressing one of these areas should be a deliberate choice. And more than one evaluator has forgotten to pay attention to side effects only to regret the oversight later. To help students remember these elements, the MISSION acronym will be a theme throughout this text, explicitly used in the case studies to illustrate both the various forms these components can take as well as the times when good evaluations may pay little or even no attention to some of them.

COMMON TYPES OF PROGRAM EVALUATIONS

The primary goals of program evaluation, just discussed, can be met using a number of different types of program evaluations; the major ones involve studies of need, process, outcome, and efficiency.

Assess Needs of the Program Participants

An evaluation of need seeks to identify and measure the level of unmet needs within an organization or community. Assessing unmet needs is a basic first step before any effective program planning can begin. Program planning involves the consideration of a variety of alternative approaches to meet needs. In selecting some alternatives and discarding others, planners engage in a form of program evaluation, one that occurs before the program is even begun. The close association between program planning and program evaluation is indicated in the title of the journal Program Planning and Evaluation.

As part of the assessment of need, evaluators may examine the socioeconomic profile of the community, the level of social problems within the community, and the agencies and institutions currently serving the community. Through close contact with residents and local leaders, evaluators can learn which aspects of a program are likely to be useful and which might be unacceptable, thus adding depth to the inferences drawn from statistical summaries suggesting critical unmet needs. Sometimes program evaluators believe that there is a pressing need for a sense of self-determination or empowerment (Fetterman & Wandersman, 2004).

Examine the Process of Meeting the Needs

Once a program has been developed and begun, evaluators turn to the task of documenting the extent to which implementation has taken place, the nature of the people being served, and the degree to which the program operates as expected. The activities making up the program, such as hours of counseling, number of classes held, and hours police officers patrolled, are called “outputs.” Evaluations of process involve checking on the assumptions made while the program was being planned. Do the needs of the people served or the community match what was believed during planning? Is there evidence to support the assessment of needs made during the planning stage? Do the activities carried out by the staff match the plans for the program? What evidence can be found that supports the theoretical assumptions made by the program planners? It is crucial to learn how a program actually operates before offering the same services at additional locations or with other populations.

Measure the Outcomes and Impacts of a Program

If a study of implementation shows that a program has been implemented well and that people seek and secure its services, a measurement of the program's outcomes may become a focus of an evaluation. An evaluation of outcome can take on several levels of complexity. The most elementary level concerns the condition of those who have received services. Are program participants doing well? Do they possess the skills being taught? A more challenging evaluation would compare the performance of those in the program with those not receiving its services. An even more challenging evaluation would show that receiving the program’s services caused a change for the better in the condition of those participating in the program (Boruch, 1997).

Although program managers hope that their programs will elicit positive changes in people, the causes of behavioral changes are difficult to pin down. For example, many people begin psychotherapy during a life crisis. If after several months of counseling they feel better, the change could be due to the counseling, the natural resolution of the crisis, or something else entirely. In a work setting, a change in procedures may result in better morale because of increased efficiency or because the workers feel that management cares about their well-being. Or perhaps an improved national economic outlook reduced the workers’ anxiety about possible job loss. Discovering the causes of behavioral changes requires sophisticated techniques since an organization must continue to provide services while the evaluation is being conducted. Experienced evaluators are not surprised by tensions between gathering information and providing services.

Besides the limitations on the choice of research design, evaluators seeking to assess the outcome of a program often discover that people hold different opinions about what constitutes a successful outcome. A job-training program paying unemployed people to learn job skills may have been planned so that they can later obtain jobs with private companies. City officials may view the training as a reward for people who have worked in local election campaigns, and the trainees may view the program as a good, albeit temporary, job. An evaluator would not adopt just one of these views of the desired outcome of the program.

Assessing the maintenance of improvement creates another problem when evaluating outcomes. People leaving a drug rehabilitation program typically return to the same community that contributed to their problems in the first place. Despite their good intentions, people treated for alcoholism are often unable to withstand the influence of their peers. Changing long-standing behaviors is difficult. Although positive changes may be observed after a person’s participation in a program, the changes may be only superficial and disappear in a matter of months, weeks, or days. If so, was the program a failure? When positive outcomes do occur, their effects can influence other behaviors and even improve the condition of other people (e.g., children); such long-term outcomes are called '‘impacts.”

Integrate the Needs, Costs, and Outcomes

Even when evaluators can show that a program has helped participants, they must also deal with the question of costs. Just as a family must make choices in how a budget is spent, governments and agencies must select from among those services that might be supported. A successful program that requires a large amount of resources simply may not be a good choice if a similar outcome can be achieved with lower cost. If an evaluator is asked to compare two or more programs designed to effect similar outcomes, efficiency can be assessed in a fairly straightforward manner; this would be called a cost-effectiveness analysis (Levin & McEwan, 2001). Real difficulties occur when evaluators are asked to compare the efficiency of programs designed to achieve different outcomes, sometimes for different groups of people. For example, should a university spend funds to reduce alcohol abuse among students or to increase the number of tutors available? Although competing purposes are seldom contrasted in such a stark fashion, no organization can do everything that might be worthwhile; choices have to be made. Evaluations that incorporate high level integration of multiple elements have the potential to provide information to administrators to help them make such choices (King, 2017).

Note that there is a logical sequence to these four general types of evaluations. Without measuring need, programs cannot be planned properly; without a plan, implementation would be haphazard; without effective implementation, outputs are not provided adequately and good outcomes cannot be expected; and without achieving good outcomes, there is no reason to worry about efficiency. A premature focus on an inappropriate evaluation question is likely to produce an evaluation with little value (Wholey, 2004).

IMPORTANT ISSUES IN EVALUATIONS

Time Frames of Needs

Short Term Needs

Many programs are set up to help people in crises. For example, medical care is needed for injuries or illnesses, and emotional and financial support are often needed after a crime, a natural disaster, or a home fire. Some educational services are designed to meet specific needs of employees required to learn new skills. To be effective, such services should be available on short notice.

Long-Term Needs

To meet some needs, services must be available for a long period of time. Children need education for many years. Psychotherapy, treatment for chronic illnesses, training for prison inmates, and rehabilitation for alcoholism and drug abuse are examples of programs that focus on problems that cannot be resolved in a short time period.

Potential Needs

Ideally, potential problems can be averted. Preventive programs are supported so that problems can be avoided or at least postponed. The effectiveness of immunization, health education, or business security programs is judged by the degree to which problems do not occur. Clearly, evaluations of prevention programs must differ from evaluations of programs developed to relieve current problems.

Extensiveness of Programs

Some programs are offered to small groups of people with similar needs, others are developed for use at many sites throughout a county or a state, and yet others are written into federal laws to be offered throughout a nation. Although tools useful in program evaluation apply to evaluations carried out at all levels, there are considerable differences between an evaluation of a day hospital program in Memorial Hospital and an evaluation of psychiatric services supported by Medicaid. Although it would be quite reasonable for these evaluations to use some of the same measures of adjustment, the decisions faced by the administrators of the local program differ so greatly from those faced by Medicaid managers that the evaluations must differ in scale, focus, and type of recommendations.

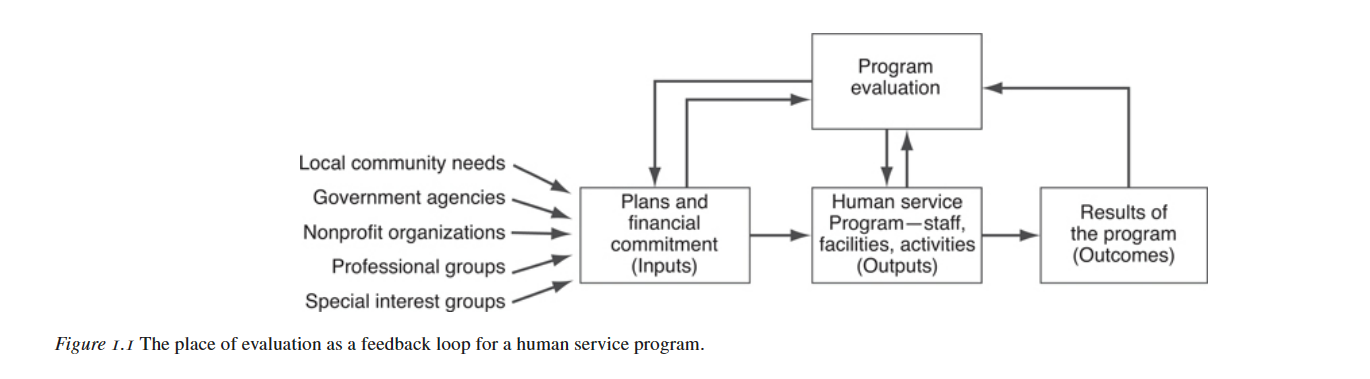
The discussions throughout this text focus on evaluations of smaller programs likely to be encountered by evaluators working for school districts, hospitals, personnel departments, social service agencies, city or state governments, and so on. Although some national programs are mentioned, most readers of an introductory text are more familiar with local government or educational programs than they are with national programs. Even when a program is supported by federal funds, implementation at a local agency differs markedly from those in other localities. The characteristics of programs are not like medications, which can be given in easily specified 20 mg doses. Thus, organizations with support from local governments, foundations, and federal agencies are all expected to carry out program evaluations and submit them to the funding source.

For studies done at a local level, statistical procedures seldom need to be particularly complicated partially because many administrators find statistics to be an incomprehensible—if not intimidating—topic. Complicated analyses require data from larger numbers of people than are usually available to evaluators working with local organizations. Evaluators can often show how to improve data summaries and presentations so that program information can be used to answer questions faced by program administrators.

PURPOSE OF PROGRAM EVALUATION

There is only one overall purpose for program evaluation activities: contributing to the provision of quality services to people in need. Program evaluation contributes to quality services by providing feedback from program activities and outcomes to those who can make changes in programs or who decide which services are to be offered. Without feedback, human service programs (indeed, any activity) cannot be carried out effectively. Our bodily processes require feedback systems; similarly, feedback on behavior in organizations is also crucial for the success of an organization. Delayed feedback, not clearly associated with the behavior being examined, is not very informative.

Figure 1.1 illustrates the place of program evaluation as a feedback loop for a human service program. Assessing needs, measuring the implementation of programs to meet those needs, evaluating the achievement of carefully formed goals and objectives, and comparing the level of outcome with the costs involved relative to similar programs serve to provide information from which to develop program improvements (Schermerhorn, Hunt, & Hunt, 2008; Wholey, 1991) or to make wise choices among possible programs (Levin & McEwan, 2001).



Feedback can be provided for different purposes. First, as noted earlier, evaluations can strengthen the plans for services and their delivery in order to improve the outcomes of programs or to increase the efficiency of programs; these formative evaluations (Scriven, 1967) are designed to help form the programs themselves. Alternately, summative evaluations can help us decide whether a program should be started, continued, or chosen from two or more alternatives (Scriven, 1967). There is a finality to summative evaluations; once the value and worth of the program have been assessed, the program might be discontinued. In actuality, very few single evaluations determine whether or not a program will be continued (Cook, Levitón, & Shadish, 1985). Many sources of information about programs are available to administrators, legislators, and community leaders. Because program evaluation is only one of these sources, evaluators are not surprised when clear evaluation findings do not lead to specific decisions based on the evaluation. Chelimsky (1997) uses the terms *evaluation for development* and *evaluation for accountability* to refer to the formative and summative purposes, respectively.

Once a program has been carefully refined using feedback from formative evaluations and found to be effective using a summative evaluation, frequent feedback is still needed to maintain the quality of the program. This third form of evaluation is called "monitoring”; its purpose is quality assurance (Schwartz & Mayne, 2005; Wholey, 1999). This can be as simple as checking the speedometer from time to time while on a highway and as complicated as examining how international aid is used (Svensson, 1997). Monitoring critical process and outcome variables to verify that an effective program stays effective is a crucial activity after programs have been successfully implemented. Furthermore, monitoring can be expected to identify problems when the community or economic conditions change. As efforts are made to resolve those problems, monitoring feedback takes on a formative function as well. If problems cannot be resolved, monitoring can come to serve a summative purpose also.

Some evaluations are also carried out to learn about programs; Chelimsky (1997) uses the term *evaluation for knowledge* to describe these. This purpose is seldom part of the work of evaluators working on specific programs. Their work might be used by others to deepen our understanding of, for example, rehabilitation programs, but this occurs only after evaluations are completed and shared. Evaluators recognize the need to develop better understandings of programs and intervention strategies; however, this text does not focus on such types of evaluation because they involve more ambitious undertakings than most evaluators deal with regularly.

THE ROLES OF EVALUATORS

Principles, Competencies, and Credentialing

Along with the above points about components, types, and issues in program evaluations, there are additional ways to think about what makes for good evaluations and good evaluators. Although on one level, describing such expectations might seem straightforward, in fact, there is no simple answer to either of these two similar but not identical points. (Although good evaluators should be expected to produce good evaluations, assessing people and reports involve quite different processes.) Still there are a number of relevant details about the current state of the field in these areas, with both similarities and differences in approaches across the world. The main categories are Standards or Principles, Competencies, and Credentialing, and the following brief summary describes the state in fall 2017. For updates since this edition was published, see the eResources.

The American Evaluation Association (2017) has adopted “Guiding Principles for Evaluators” that cover five basic areas such as Integrity/Honesty, while the Canadian Evaluation Society (CES) has adopted standards that cover five perspectives such as Utility and Accuracy (Yarbrough, Shulha, Hopson, & Caruthers, 2011). Like many organizational principles, many of the points are more aspirational than precise. For example, one of the specific AEA principles is that evaluators should “seek to maintain and improve their competencies in order to provide the highest level of performance in their evaluations” (American Evaluation Association, 2017, p. 313), which does not indicate how much improvement is enough. Still, the intent of the principles and standards is obvious, and blatant failures to follow them can be noted.

In some areas, the general agreement in the field has not yet crystallized into uniform policies. Proposals for establishing a list of competencies are more than a decade old (King, Stevahn, Ghere, & Minnema, 2001; Stevahn, King, Ghere & Minnema, 2005) and current research in the field refers to these proposals favorably (e.g., Galport & Azzam, 2017). The details fit well with both the principles and common practice, noting crucial elements such as showing ethical approaches like respecting clients and considering public welfare, demonstrating knowledge about appropriate research methods, and being responsive to the specific context of a given program (Stevahn et al., 2005). The Canadian Evaluation Society (2010) formally adopted a set of Competencies, while the American Evaluation Association has not done so at this time, although many evaluators continue to demonstrate their voluntary commitment to those values.

The CES also offers the designation of “Credentialed Evaluator” to CES members who apply for it and meet the criteria (Kuji-Shikatani, 2015). The possibility of a credentialing process in the United States and the related possibility of an accreditation process for programs that train evaluators continue to be raised and discussed (Altschuld & Engle, 2015; King & Stevahn, 2015; Kuji-Shikatani, 2015; LaVelle & Donaldson, 2015; McDavid & Huse, 2015; Seidling, 2015; Shackman, 2015). The trade-offs between more uniform quality on the one hand and flexibility to address substantially different needs and contexts on the other have not yet been resolved.

Comparison of Internal and External Evaluators

There are two primary ways evaluators relate to organizations needing evaluation services: (i) evaluators can work for the organization itself and evaluate a variety of programs in that setting, or (2) they can work for a research firm, a university, or a government agency to evaluate a specific program. In this book, the terms *internal evaluator* and *external evaluator* refer to these two types of evaluators, respectively. Although there are great similarities in what all evaluators do, the affiliation of an evaluator has implications for the manner in which evaluations are done.

*Factors Related to Competence*

In terms of knowledge about a program, internal evaluators have an advantage since they have better access to program directors and to the administrators of the organization. A person who is physically present on a routine basis is likely to see programs in action, to know staff members, and to learn about the reputation of programs from other people in the organizations. Such information is unavailable to an external evaluator. The more that is known about the actual operation of programs, the easier it is to ask relevant questions during the planning and interpretation of evaluations.

The technical expertise of an evaluator is important. An internal evaluator often works with a small group of two or three; some evaluators work alone. An external evaluator, however, can often draw upon the resources of a greater number of people, some of whom are very skilled in sampling, qualitative analyses, or statistics. Some independent evaluators, however, essentially work alone and do not have access to such resources.

A different facet of the question of technical expertise is suggested by the need to perform evaluations of programs in different areas of an organization. Moving from topic to topic is stimulating; however, there is also a risk that inexperience with the services being evaluated could limit an internal evaluator’s insight into aspects of some programs. By selecting external evaluators with experience in the type of program being evaluated, organizations might avoid errors due to inexperience.

Personal Quauties

In addition to technical competence, an evaluator’s personal qualities are also important. Evaluators can do more effective work if they are objective and fair, trusted by administrators, and concerned that program improvements be encouraged. Being well known and having been found worthy of trust (Taut & Alkin, 2003), internal evaluators usually find program directors and staff more willing to devote time to an evaluation, to admit problems, and to share confidences than they would be with an evaluator not affiliated with the organization. Being trusted increases the likelihood that a competent evaluator can fill an organizational educator role. Since credibility can be easily lost, evaluators avoid acting in ways that might suggest that they have allowed personal biases to influence their conclusions.

An internal evaluator can also be expected to want to improve the organization sponsoring the program and paying the evaluator’s salary. External evaluators, being less dependent on the organization sponsoring the program, might not have the same level of commitment to working for program improvements. On the other hand, being dependent on the organization could affect the objectivity of an evaluator. Knowing program directors and staff members might make it difficult for an internal evaluator to remain objective (Ray, 2006; Scriven, 1997a); it is not easy to criticize a friend. An external evaluator is less likely to experience conflicting pressures when an evaluation reveals problems in a program. Internal evaluators might find it easier to remain objective if they remember that their value to an organization depends on their work being credible. Developing a reputation for tackling sensitive issues is somewhat easier when evaluators emphasize that most deficiencies in programs are due to system problems rather than personal inadequacies. Deming (1986), an early champion of the quality revolution in industry, insisted that 85% of problems in workplaces are due to limitations in organizational procedures or work setting designs and 15% due to staff problems. Since many people mistakenly believe that evaluations focus on finding bad apples, evaluators should often refer to Deming’s point.

Factors Related to the Purpose oe an Evaluation

Earlier in this chapter the purposes of evaluation were discussed—formative, summative, and quality assurance. Both internal evaluators and external evaluators can perform any type of evaluations; however, internal evaluators may have an advantage in performing formative and quality assurance evaluations. This seems true because such evaluations cannot prompt a decision to end a program. In fact, the internal evaluator’s rapport with managers and staff may encourage nondefensive communications, which are essential if avenues to improve programs are to be found. If, by contrast, a summative evaluation is wanted because support for a program might be withdrawn, it may be better if an external evaluator is used instead of an internal evaluator. For example, if a small elementary school district knows that it must close one of three schools, it may be wise to have an external evaluator perform an evaluation of the schools and select the one to close. No matter which one is selected, some residents will object. It could be hard for an administrator to serve in the school district if a sizable portion of the residents were dissatisfied with the decision; however, the external evaluator need never set foot in the school district again.

As should be clear, internal and external evaluators have many complementary strengths and weaknesses. Depending on the specifics of a given evaluation, one or the other may be preferable. A growing possibility in recent years is to develop a partnership between internal and external evaluators that has the potential to make good use of the strengths from both (Le Menestrel, Walahoski, & Mieke, 2014).

Evaluation and Service

An evaluator’s job falls somewhere between the role of the social scientist concerned about theory, the design of research, and the analysis of data (but for the most part uninvolved with the delivery of services) and the role of the practitioner dealing with people in need (but seldom deeply interested or extensively trained in the methods of data collection and analysis). Evaluators are able to use the language and tools of the research scientist; however, they must also be sensitive to the concerns and style of the service delivery staff. In addition, evaluators are called upon to communicate with organization administrators who have different priorities, such as balancing a budget and weighing proposals for expansions of services. Since the role of program evaluator is still fairly new, it is likely that the evaluator will at times seem out of step with everyone.

Participating in such a new field has advantages and disadvantages. The advantages include the intellectual stimulation provided by exposure to people serving various roles in human service settings and the satisfaction of seeing research methods used in ways that can benefit people. The disadvantages include being viewed as intrusive and unnecessary by some service delivery personnel. Instead of concentrating on problems, sensitive evaluators make sure program managers and staff have opportunities to talk about what goes well and what is rewarding in their work (Preskill & Coghlan, 2003). Yet to demonstrate the value and worth of the program, effective evaluators must ask challenging questions and tactfully insist that answers be supported by data.

Of course, at times even the most skilled evaluator will be in conflict with an organization. One of the early leaders in evaluation was concerned that the perceived threat of evaluation would make it impossible to collect data of sufficient validity to draw useful conclusions (Campbell, 1983). By contrast, other observers (Lincoln, 1990; Vermillion & Pfeiffer, 1993) predict that staff members themselves will more frequently use evaluation methods to develop improvements in programs. When evaluation systems are developed in ways which clearly distinguish between information collected for (1) the purpose of program improvement and (2) the purpose of individual performance appraisals for salary increases, conflicts among evaluation, service delivery, and administration may be minimized. That happy day has not yet come, but when evaluation is built into program management and all agree that program improvement is both possible and desirable, the potential for conflict will be reduced (Donaldson, Gooler, & Scriven, 2002).

Activities Often Confused with Program Evaluation

Sometimes it is easier to understand a concept when one understands what the concept does not include. Program evaluation is often confused with basic research, individual assessment, and compliance audits. Although these activities are valuable, when program evaluation work is confused with one of these other activities, the evaluation becomes all the more difficult to carry out.

Basic research concerns questions of theoretical interest, without regard to the information needs of people or organizations. By contrast, program evaluators gather information to help people improve their effectiveness, to assist administrators to make program-level decisions, and to enable interested parties to examine program effectiveness. Evaluators, of course, are very interested in theories of why a service might help its participants. Understanding theories helps in planning programs and selecting variables to observe. However, contributing to the development of theories can only be a delightful side benefit of a program evaluation. Evaluation findings should be relevant to the immediate or short-term needs of managers and must be timely. If program staff members believe that evaluators are collecting information primarily to serve research interests, cooperation may well be lost.

Human service staff members often confuse program evaluation with the assessments made by educational psychologists, personnel workers, and counseling psychologists who administer intelligence, aptitude, interest, achievement, and personality tests for the purpose of evaluating a person’s need for service or measuring qualifications for a job or a promotion. These activities are not part of the work of a program evaluator. In program evaluation, information about the level of job performance, educational achievement, or health may well be gathered; however, the purpose is not to diagnose people, determine eligibility for benefits, or choose whom to hire or promote. Instead, the purpose is to learn how well a program is helping people improve on those variables.

Last, the methods and objectives of program evaluation differ from those used by program auditors examining government-sponsored programs to verify that they operate in compliance with laws and regulations. When Congress supports a program, it is important that the funds be spent as Congress intends. If the appropriation was for the enrichment of elementary schools, spending the funds for high school laboratories constitutes fraud. If 10,000 students were to be served, documentation of services for close to that many children ought to be available. Program evaluators are concerned that the program serve the right number of children, but in addition, evaluators would be particularly interested in how the services have affected the children.

The work of program auditors is closely related to accounting, whereas program evaluators tend to identify with education and other social sciences. Auditing and evaluation have moved toward each other in recent years because neither can give a complete picture of a program (Hatry, Wholey, & Newcomer, 2004). The change of the name of the GAO from the “General Accounting Office” to the “Government Accountability Office” illustrates this change. Nevertheless, if a program evaluator seeking to help a program make improvements in its services is viewed as a program auditor seeking to verify compliance with laws, it is easy to imagine that the evaluator’s purpose could be misunderstood.

Evaluation and Related Activities of Organizations

Some agencies may combine evaluation with another activity of the agency. Five functions that sometimes include evaluation are research, education and staff development, auditing, planning, and human resources.

Research

Some human service organizations routinely sponsor research directly with operating funds or indirectly through grants. For example, a police department was awarded a gran study the effect of providing emotional support to victims of crimes. The social scientists responsible for such research would be excellent colleagues for an evaluator. Univers affiliated hospitals carry on research; researchers working with psychiatric problems or emotional issues related to medical care deal with research problems sufficiently simila those faced by evaluators to serve as colleagues.

Education and State Development

Some organizations have joined the functions of education and evaluation. There is a long-standing precedent for this marriage in educational psychology. New study materials ; curricula are evaluated before being widely used by school districts. Universities have combined faculty development and program evaluation into one office. Businesses h developed major educational programs to train employees to qualify for more responsible positions; such programs need monitoring.

Auditing

As mentioned earlier in this chapter, some states have combined the functions of program evaluation and program oversight often carried out by state auditors. There are import differences in the assumptions and roles of auditors and evaluators, but there is precedent for coordinating these efforts in the inspector general offices in the federal and si governments.

Pianning

Planning is an activity in which people with skills in program evaluation are needed because many service providers are not data oriented, feel uncomfortable constructing survt and may be unable to analyze the information obtained. Evaluators can contribute to the task of estimating the level of community acceptance. Sometimes evaluators can h

CASE STUDY 1

Open Arms

In May 2014, REACH Evaluation presented its program evaluation report of the Open Arms Children’s Health Dental Service, a program at Home of the Innocents. The report is a good example of a comprehensive evaluation. Open Arms involved on-site dental services for special needs children supported with funds from the Social Innovation Fund through the Foundation for a Healthy Kentucky and in partnership with the University of Louisville School of Dentistry’s Pediatric Dental Program.

Meeting Needs

The extensive three-page section “Documentation of Need” included a summary of research that “linked poor oral health to illness, chronic disease and early mortality” (p. 9), research on the wide span of health needs among foster children, and the comparative point that Kentucky “has the highest percentage of edentate persons (those who have lost all their natural teeth due to tooth decay or gum disease) ages 18 to 64” (p. 9) along with multiple additional points of evidence regarding the need that Open Arms was intended to address.

Implementation

The issue of what was actually done in the program was addressed at several points in the report. For example, copies of medical notes for dental visits for the same child before and after the program was initiated show enormous differences in detail—about two pages of very specific notes after compared to one line on a card before. In the “Early Lessons Learned” section, the progression of staffing strategies during the program was described, detailing how initial plans were adapted.

Stakeholders

The evaluators had contact with clients, client family members, and permanent staff as well as dental students and dental residents who received training as a part of the program. As with many evaluation reports, the category of “stakeholders” was not listed explicitly, but the perspectives of multiple stakeholders were clearly considered as a part of the evaluation. For example, although programs generally focus on clients, and clients’ perspectives were included, so were accounts of some parents of clients, who are also important stakeholders.

Side Effects

As with stakeholders, the report did not explicitly note that it found side effects. But with the stated goal of improving children’s dental care, the report’s note that having dental services provided on site saved staff time in several ways was an implicit description of an unanticipated side effect. Specifically, reduced time was needed to transport children to dental services because the services were on site, and they found greater efficiency when scheduling visits because the person planning the times for the dental services had access to information about children’s other activities, and was thus able to avoid potential conflicts more easily. No negative side effects were noted.

Improvement Focus

Suggestions by staff and clients for ways to improve the program were requested and listed. In addition, the evaluators showed an implicit focus on improvement by suggesting strategies for addressing one of the central challenges for the program—how to continue it past the initial period of outside funding.

Outcomes

As the primary goals of the program were fairly straightforward—serving 880 children a year, serving special needs children, and serving 80 dental students and 6 pediatric residents, the report appropriately presented direct counts of the relevant groups as the primary outcomes. For two 12-month periods (overlapping, because the available data covered only 21 months) 920 and 1,411 children were served. The 1,475 special needs individuals served during the 21-month period were described with appropriate categories, such as “Medically Fragile” or “Foster Care.” And the planned number of students and residents was met. The results of inferential statistical analyses showed a significant improvement in treatment acceptance over time, while also noting that the changes were relatively modest, partly because the scores were very high to begin with (a ceiling effect). They also found very much greater improvement in four groups of children (such as Medically Fragile and Refugee) than in four other groups (such as Outpatient/Nonmedically Fragile). Several summaries of qualitative data addressed issues not analyzed quantitatively, such as the improvement in the medical notes indicated above.

SUMMARY AND PREVIEW

This chapter outlines the major questions and purposes of program evaluation: assess unmet need, document implementation, consider as many stakeholders as possible, look for positive as well as negative side effects, consider how to improve the program when possible, measure outcomes, and examine what can be learned about why the outcomes occur. Such activities are undertaken to help plan and refine programs, to assess their worth, and to make corrections in ongoing operations. Some ways in which evaluators fit into the activities of agencies are described.

Chapter 2 deals with steps in planning an evaluation. Working with program staff and sponsors will often reveal that some people fear program evaluations; when an evaluation must have the cooperation of staff, it is necessary to allay these fears before beginning.

STUDY QUESTIONS

1. One of the interesting aspects of program evaluation is its relevance to a variety of activities and agencies. Gather some examples of activities related to program evaluation from newspapers or news magazines. You might find material on medical care, education, or public policy in any area. The material might concern advocacy of new proposals or criticisms of current activities. Consider what information might illuminate the debates.
2. Some people have suggested that while people want to believe that something is being done about social problems, most people don't really care whether programs have measurable impacts on the problems. In other words, social programs are symbols, not real efforts. Consider the validity of such comments in the context of the reports you found in response to the first question.
3. Illustrate how program evaluation activities could be applied in an organization with which you are familiar.
4. List the advantages and disadvantages of making program evaluation a part of human service delivery systems. Save your lists; when you are finished with the course, consider how you might change them.
5. Compare your ideas on evaluating the Sexual Assault Prevention Program mentioned in the introductory paragraph of this chapter with those of other students. What reasons lie behind different choices of outcomes considered relevant?

ADDITIONAL RESOURCE

W. K. Kellogg Foundation. (2010). *Evaluation handbook.* Retrieved from <https://www.wkkf.org/resource-directory/resource/2010/w-k-kellogg-foundation-evaluation-handbook>.

This work provides a helpful overview and succinct details of the program evaluation process. In addition, it highlights some related points such as the way that values and assumptions affect evaluation work. Best of all, like many resources on the web, it is free.