1 Evaluation’s Basic Purpose, Uses, and Conceptual Distinctions

Orienting Questions

1.

What is evaluation? Why is it important?

2.

What is the difference between formal and informal evaluation?

3.

What are some purposes of evaluation? What roles can the evaluator play?

4.

What are the major differences between formative and summative evaluations?

5.

What questions might an evaluator address in a needs assessment, a process evaluation, and an outcome evaluation?

6.

What are the advantages and disadvantages of an internal evaluator? An external evaluator?

The challenges confronting our society in the twenty-first century are enormous. Few of them are really new. In the United States and many other countries, the public and nonprofit sectors are grappling with complex issues: educating children for the new century; reducing functional illiteracy; strengthening families; training people to enter or return to the workforce; training employees who currently work in an organization; combating disease and mental illness; fighting discrimination; and reducing crime, drug abuse, and child and spouse abuse. More recently, pursuing and balancing environmental and economic goals and working to ensure peace and economic growth in developing countries have become prominent concerns. As this book is written, the United States and many countries around 34the world are facing challenging economic problems that touch every aspect of society. The policies and programs created to address these problems will require evaluation to determine which solutions to pursue and which programs and policies are working and which are not. Each new decade seems to add to the list of challenges, as society and the problems it confronts become increasingly complex.

As society’s concern over these pervasive and perplexing problems has intensified, so have its efforts to resolve them. Collectively, local, regional, national, and international agencies have initiated many programs aimed at eliminating these problems or their underlying causes. In some cases, specific programs judged to have been ineffective have been “mothballed” or sunk outright, often to be replaced by a new program designed to attack the problem in a different—and, hopefully, more effective—manner.

In more recent years, scarce resources and budget deficits have posed still more challenges as administrators and program managers have had to struggle to keep their most promising programs afloat. Increasingly, policymakers and managers have been faced with tough choices, being forced to cancel some programs or program components to provide sufficient funds to start new programs, to continue others, or simply to keep within current budgetary limits.

To make such choices intelligently, policy makers need good information about the relative effectiveness of programs. Which programs are working well? Which are failing? What are the programs’ relative costs and benefits? Similarly, each program manager needs to know how well different parts of programs are working. What can be done to improve those parts of the program that are not working as well as they should? Have all aspects of the program been thought through carefully at the planning stage, or is more planning needed? What is the theory or logic model for the program’s effectiveness? What adaptations would make the program more effective?

Answering such questions is the major task of program evaluation. The major task of this book is to introduce you to evaluation and the vital role it plays in virtually every sector of modern society. However, before we can hope to convince you that good evaluation is an essential part of good programs, we must help you understand at least the basic concepts in each of the following areas:

 • How we—and others—define evaluation

 • How formal and informal evaluation differ

 • The basic purposes—and various uses—of formal evaluation

 • The distinction between basic types of evaluation

 • The distinction between internal and external evaluators

 • Evaluation’s importance and its limitations

Covering all of those areas thoroughly could fill a whole book, not just one chapter of an introductory text. In this chapter, we provide only brief coverage of each of these topics to orient you to concepts and distinctions necessary to understand the content of later chapters.

4 5

Informal versus Formal Evaluation

Evaluation is not a new concept. In fact, people have been evaluating, or examining and judging things, since the beginning of human history. Neanderthals practiced it when determining which types of saplings made the best spears, as did Persian patriarchs in selecting the most suitable suitors for their daughters, and English yeomen who abandoned their own crossbows in favor of the Welsh longbow. They had observed that the longbow could send an arrow through the stoutest armor and was capable of launching three arrows while the crossbow sent only one. Although no formal evaluation reports on bow comparisons have been unearthed in English archives, it is clear that the English evaluated the longbow’s value for their purposes, deciding that its use would strengthen them in their struggles with the French. So the English armies relinquished their crossbows, perfected and improved on the Welsh longbow, and proved invincible during most of the Hundred Years’ War.

By contrast, French archers experimented briefly with the longbow, then went back to the crossbow—and continued to lose battles. Such are the perils of poor evaluation! Unfortunately, the faulty judgment that led the French to persist in using an inferior weapon represents an informal evaluation pattern that has been repeated too often throughout history.

As human beings, we evaluate every day. Practitioners, managers, and policymakers make judgments about students, clients, personnel, programs, and policies. These judgments lead to choices and decisions. They are a natural part of life. A school principal observes a teacher working in the classroom and forms some judgments about that teacher’s effectiveness. A program officer of a foundation visits a substance abuse program and forms a judgment about the program’s quality and effectiveness. A policymaker hears a speech about a new method for delivering health care to uninsured children and draws some conclusions about whether that method would work in his state. Such judgments are made every day in our work. These judgments, however, are based on informal, or unsystematic, evaluations.

Informal evaluations can result in faulty or wise judgments. But, they are characterized by an absence of breadth and depth because they lack systematic procedures and formally collected evidence. As humans, we are limited in making judgments both by the lack of opportunity to observe many different settings, clients, or students and by our own past experience, which both informs and biases our judgments. Informal evaluation does not occur in a vacuum. Experience, instinct, generalization, and reasoning can all influence the outcome of informal evaluations, and any or all of these may be the basis for sound, or faulty, judgments. Did we see the teacher on a good day or a bad one? How did our past experience with similar students, course content, and methods influence our judgment? When we conduct informal evaluations, we are less cognizant of these limitations. However, when formal evaluations are not possible, informal evaluation carried out by knowledgeable, experienced, and fair people can be very useful indeed. It would be unrealistic to think any individual, group, or organization could formally evaluate everything it does. Often informal evaluation is the only 56practical approach. (In choosing an entrée from a dinner menu, only the most compulsive individual would conduct exit interviews with restaurant patrons to gather data to guide that choice.)

Informal and formal evaluation, however, form a continuum. Schwandt (2001a) acknowledges the importance and value of everyday judgments and argues that evaluation is not simply about methods and rules. He sees the evaluator as helping practitioners to “cultivate critical intelligence.” Evaluation, he notes, forms a middle ground “between overreliance on and over-application of method, general principles, and rules to making sense of ordinary life on one hand, and advocating trust in personal inspiration and sheer intuition on the other” (p. 86). Mark, Henry, and Julnes (2000) echo this concept when they describe evaluation as a form of assisted sense-making. Evaluation, they observe, “has been developed to assist and extend natural human abilities to observe, understand, and make judgments about policies, programs, and other objects in evaluation” (p. 179).

Evaluation, then, is a basic form of human behavior. Sometimes it is thorough, structured, and formal. More often it is impressionistic and private. Our focus is on the more formal, structured, and public evaluation. We want to inform readers of various approaches and methods for developing criteria and collecting information about alternatives. For those readers who aspire to become professional evaluators, we will be introducing you to the approaches and methods used in these formal studies. For all readers, practitioners and evaluators, we hope to cultivate that critical intelligence, to make you cognizant of the factors influencing your more informal judgments and decisions.

A Brief Definition of Evaluation and Other Key Terms

In the previous section, the perceptive reader will have noticed that the term “evaluation” has been used rather broadly without definition beyond what was implicit in context. But the rest of this chapter could be rather confusing if we did not stop briefly to define the term more precisely. Intuitively, it may not seem difficult to define evaluation. For example, one typical dictionary definition of evaluation is “to determine or fix the value of: to examine and judge.” Seems quite straightforward, doesn’t it? Yet among professional evaluators, there is no uniformly agreed-upon definition of precisely what the term “evaluation” means. In fact, in considering the role of language in evaluation, Michael Scriven, one of the founders of evaluation, for an essay on the use of language in evaluation recently noted there are nearly 60 different terms for evaluation that apply to one context or another. These include adjudge, appraise, analyze, assess, critique, examine, grade, inspect, judge, rate, rank, review, score, study, test, and so on (cited in Patton, 2000, p. 7). While all these terms may appear confusing, Scriven notes that the variety of uses of the term evaluation “reflects not only the immense importance of the process of evaluation in practical life, but the explosion of a new area of study” (cited in Patton, 2000, p. 7). This chapter will introduce the reader 67to the array of variations in application, but, at this point, we will focus on one definition that encompasses many others.

Early in the development of the field, Scriven (1967) defined evaluation as judging the worth or merit of something. Many recent definitions encompass this original definition of the term (Mark, Henry, & Julnes, 2000; Schwandt, 2008; Scriven, 1991a; Stake, 2000a; Stufflebeam, 2001b). We concur that evaluation is determining the worth or merit of an evaluation object (whatever is evaluated). More broadly, we define evaluation as the identification, clarification, and application of defensible criteria to determine an evaluation object’s value (worth or merit) in relation to those criteria. Note that this definition requires identifying and clarifying defensible criteria. Often, in practice, our judgments of evaluation objects differ because we have failed to identify and clarify the means that we, as individuals, use to judge an object. One educator may value a reading curriculum because of the love it instills for reading; another may disparage the program because it does not move the child along as rapidly as other curricula in helping the student to recognize and interpret letters, words, or meaning. These educators differ in the value they assign to the curricula because their criteria differ. One important role of an evaluator is to help stakeholders articulate their criteria and to stimulate dialogue about them. Our definition, then, emphasizes using those criteria to judge the merit or worth of the product.

Evaluation uses inquiry and judgment methods, including: (1) determining the criteria and standards for judging quality and deciding whether those standards should be relative or absolute, (2) collecting relevant information, and (3) applying the standards to determine value, quality, utility, effectiveness, or significance. It leads to recommendations intended to optimize the evaluation object in relation to its intended purpose(s) or to help stakeholders determine whether the evaluation object is worthy of adoption, continuation, or expansion.

Programs, Policies, and Products

In the United States, we often use the term “program evaluation.” In Europe and some other countries, however, evaluators often use the term “policy evaluation.” This book is concerned with the evaluation of programs, policies, and products. We are not, however, concerned with evaluating personnel or the performance of individual people or employees. That is a different area, one more concerned with management and personnel.1 (See Joint Committee. [1988]) But, at this point, it would be useful to briefly discuss what we mean by programs, policies, and products. “Program” is a term that can be defined in many ways. In its simplest sense, a program is a “standing arrangement that provides for a . . . service” (Cronbach et al., 1980, p. 14). The Joint Committee on Standards for Educational Evaluation (1994) defined program simply as “activities that are provided on a continuing basis” (p. 3). In their new edition of the Standards (2010) the Joint Committee noted that a program is much more than a set of activities. They write:

1 The Joint Committee on Standards for Educational Evaluation has developed some standards for personnel evaluation that may be of interest to readers involved in evaluating the performance of teachers or other employees working in educational settings. These can be found at http://www.eval.org/evaluationdocuments/perseval.html.

7 8

Defined completely, a program is

 • A set of planned systematic activities

 • Using managed resources

 • To achieve specified goals

 • Related to specific needs

 • Of specific, identified, participating human individuals or groups

 • In specific contexts

 • Resulting in documentable outputs, outcomes and impacts

 • Following assumed (explicit or implicit) systems of beliefs (diagnostic, causal, intervention, and implementation theories about how the program works)

 • With specific, investigable costs and benefits. (Joint Committee, 2010, in press)

Note that their newer definition emphasizes programs achieving goals related to particular needs and the fact that programs are based on certain theories or assumptions. We will talk more about this later when we discuss program theory. We will simply summarize by saying that a program is an ongoing, planned intervention that seeks to achieve some particular outcome(s), in response to some perceived educational, social, or commercial problem. It typically includes a complex of people, organization, management, and resources to deliver the intervention or services.

In contrast, the word “policy” generally refers to a broader act of a public organization or a branch of government. Organizations have policies—policies about recruiting and hiring employees, policies about compensation, policies concerning interactions with media and the clients or customers served by the organization. But, government bodies—legislatures, departments, executives, and others—also pass or develop policies. It might be a law or a regulation. Evaluators often conduct studies to judge the effectiveness of those policies just as they conduct studies to evaluate programs. Sometimes, the line between a program and a policy is quite blurred. Like a program, a policy is designed to achieve some outcome or change, but, unlike a program, a policy does not provide a service or activity. Instead, it provides guidelines, regulations, or the like to achieve a change. Those who study public policy define policy even more broadly: “public policy is the sum of government activities, whether acting directly or through agents, as it has an influence on the life of citizens” (Peters, 1999, p. 4). Policy analysts study the effectiveness of public policies just as evaluators study the effectiveness of government programs. Sometimes, their work overlaps. What one person calls a policy, another might call a program. In practice, in the United States, policy analysts tend to be trained in political science and economics, and evaluators tend to be trained in psychology, sociology, education, and public administration. As the field of evaluation expands and clients want more information on government programs, evaluators study the effectiveness of programs and policies.

Finally, a “product” is a more concrete entity than either a policy or a program. It may be a textbook such as the one you are reading. It may be a piece of software. Scriven defines a product very broadly to refer to the output of something. Thus, a product could be a student or a person who received training, the 89work of a student, or a curricula which is “the product of a research and development effort” (1991a, p. 280).

Stakeholders

Another term used frequently in evaluation is “stakeholders.” Stakeholders are various individuals and groups who have a direct interest in and may be affected by the program being evaluated or the evaluation’s results. In the Encyclopedia of Evaluation, Greene (2005) identifies four types of stakeholders:

 (a) People who have authority over the program including funders, policy makers, advisory boards;

 (b) People who have direct responsibility for the program including program developers, administrators, managers, and staff delivering the program;

 (c) People who are the intended beneficiaries of the program, their families, and their communities; and

 (d) People who are damaged or disadvantaged by the program (those who lose funding or are not served because of the program). (pp. 397–398)

Scriven (2007) has grouped stakeholders into groups based on how they are impacted by the program, and he includes more groups, often political groups, than does Greene. Thus, “upstream impactees” refer to taxpayers, political supporters, funders, and those who make policies that affect the program. “Midstream impactees,” also called primary stakeholders by Alkin (1991), are program managers and staff. “Downstream impactees” are those who receive the services or products of the program.

All of these groups hold a stake in the future direction of that program even though they are sometimes unaware of their stake. Evaluators typically involve at least some stakeholders in the planning and conduct of the evaluation. Their participation can help the evaluator to better understand the program and the information needs of those who will use it.

Differences in Evaluation and Research

It is important to distinguish between evaluation and research, because these differences help us to understand the distinctive nature of evaluation. While some methods of evaluation emerged from social science research traditions, there are important distinctions between evaluation and research. One of those distinctions is purpose. Research and evaluation seek different ends. The primary purpose of research is to add to knowledge in a field, to contribute to the growth of theory. A good research study is intended to advance knowledge. While the results of an evaluation study may contribute to knowledge development (Mark, Henry, & Julnes, 2000), that is a secondary concern in evaluation. Evaluation’s primary purpose is to provide useful information to those who hold a stake in whatever is being evaluated (stakeholders), often helping them to make a judgment or decision.

Research seeks conclusions; evaluation leads to judgments. Valuing is the sine qua non of evaluation. A touchstone for discriminating between an evaluator and a researcher is to ask whether the inquiry being conducted would be regarded as a failure if it produced no data on the value of the thing being studied. A researcher answering strictly as a researcher will probably say no.

These differing purposes have implications for the approaches one takes. Research is the quest for laws and the development of theory—statements of relationships among two or more variables. Thus, the purpose of research is typically to explore and establish causal relationships. Evaluation, instead, seeks to examine and describe a particular thing and, ultimately, to consider its value. Sometimes, describing that thing involves examining causal relationships; often, it does not. Whether the evaluation focuses on a causal issue depends on the information needs of the stakeholders.

This highlights another difference in evaluation and research—who sets the agenda. In research, the hypotheses to be investigated are chosen by the researcher based on the researcher’s assessment of the appropriate next steps in developing theory in the discipline or field of knowledge. In evaluation, the questions to be answered are not those of the evaluator, but rather come from many sources, including those of significant stakeholders. An evaluator might suggest questions, but would never determine the focus of the study without consultation with stakeholders. Such actions, in fact, would be unethical in evaluation. Unlike research, good evaluation always involves the inclusion of stakeholders—often a wide variety of stakeholders—in the planning and conduct of the evaluation for many reasons: to ensure that the evaluation addresses the needs of stakeholders, to improve the validity of results, and to enhance use.

Another difference between evaluation and research concerns generalizability of results. Given evaluation’s purpose of making judgments about a particular thing, good evaluation is quite specific to the context in which the evaluation object rests. Stakeholders are making judgments about a particular evaluation object, a program or a policy, and are not as concerned with generalizing to other settings as researchers would be. In fact, the evaluator should be concerned with the particulars of that setting, with noting them and attending to the factors that are relevant to program success or failure in that setting. (Note that the setting or context may be a large, national program with many sites, or a small program in one school.) In contrast, because the purpose of research is to add to general knowledge, the methods are often designed to maximize generalizability to many different settings.

As suggested previously, another difference between research and evaluation concerns the intended use of their results. Later in the book, we will discuss the many different types of use that may occur in evaluation, but, ultimately, evaluation is intended to have some relatively immediate impact. That impact may be on immediate decisions, on decisions in the not-too-distant future, or on perspectives that one or more stakeholder groups or stakeholders have about the object of the evaluation or evaluation itself. Whatever the impact, the evaluation is designed to be used. Good research may or may not be used right away. In fact, research that adds in important ways to some theory may not be immediately noticed, and 1011connections to a theory may not be made until some years after the research is conducted.2 Nevertheless, the research stands alone as good research if it meets the standards for research in that discipline or field. If one’s findings are to add to knowledge in a field, ideally, the results should transcend the particulars of time and setting.

Thus, research and evaluation differ in the standards used to judge their adequacy (Mathison, 2007). Two important criteria for judging the adequacy of research are internal validity, the study’s success at establishing causality, and external validity, the study’s generalizability to other settings and other times. These criteria, however, are not sufficient, or appropriate, for judging the quality of an evaluation. As noted previously, generalizability, or external validity, is less important for an evaluation because the focus is on the specific characteristics of the program or policy being evaluated. Instead, evaluations are typically judged by their accuracy (the extent to which the information obtained is an accurate reflection—a one-to-one correspondence—with reality), utility (the extent to which the results serve the practical information needs of intended users), feasibility (the extent to which the evaluation is realistic, prudent, diplomatic, and frugal), and propriety (the extent to which the evaluation is done legally and ethically, protecting the rights of those involved). These standards and a new standard concerning evaluation accountability were developed by the Joint Committee on Standards for Evaluation to help both users of evaluation and evaluators themselves to understand what evaluations should do (Joint Committee, 2010). (See Chapter 3 for more on the Standards.)

Researchers and evaluators also differ in the knowledge and skills required to perform their work. Researchers are trained in depth in a single discipline—their field of inquiry. This approach is appropriate because a researcher’s work, in almost all cases, will remain within a single discipline or field. The methods he or she uses will remain relatively constant, as compared with the methods that evaluators use, because a researcher’s focus remains on similar problems that lend themselves to certain methods of study. Evaluators, by contrast, are evaluating many different types of programs or policies and are responding to the needs of clients and stakeholders with many different information needs. Therefore, evaluators’ methodological training must be broad and their focus may transcend several disciplines. Their education must help them to become sensitive to the wide range of phenomena to which they must attend if they are to properly assess the worth of a program or policy. Evaluators must be broadly familiar with a wide variety of methods and techniques so they can choose those most appropriate for the particular program and the needs of its stakeholders. In addition, evaluation has developed some of its own specific methods, such as using logic models to understand program theory and metaevaluation. Mathison writes that “evaluation as a practice shamelessly borrows from all disciplines and ways of thinking to get at both facts and values” (2007, p. 20). Her statement illustrates both the methodological breadth required of an evaluator and the fact that evaluators’ methods must serve the purpose of valuing or establishing merit and worth, as well as establishing facts.

2 A notable example concerns Darwin’s work on evolution. Elements of his book, The Origin of the Species, were rejected by scientists some years ago and are only recently being reconsidered as new research suggests that some of these elements were correct. Thus, research conducted more than 100 years ago emerges as useful because new techniques and discoveries prompt scientists to reconsider the findings.

11 12

TABLE 1.1 Differences in Research and Evaluation

Factor

Research

Evaluation

Purpose

Add to knowledge in a field, develop laws and theories

Make judgments, provide information for decision making

Who sets the agenda or focus?

Researchers

Stakeholders and evaluator jointly

Generalizability of results

Important to add to theory

Less important, focus is on particulars of program or policy and context

Intended use of results

Not important

An important standard

Criteria to judge adequacy

Internal and external validity

Accuracy, utility, feasibility, propriety, evaluation accountability

Preparation of those who work in area

Depth in subject matter, fewer methodological tools and approaches

Interdisciplinary, many methodological tools, interpersonal skills

Finally, evaluators differ from researchers in that they must establish personal working relationships with clients. As a result, studies of the competencies required of evaluators often cite the need for training in interpersonal and communication skills (Fitzpatrick, 1994; King, Stevahn, Ghere, & Minnema, 2001; Stufflebeam & Wingate, 2005).

In summary, research and evaluation differ in their purposes and, as a result, in the roles of the evaluator and researcher in their work, their preparation, and the criteria used to judge the work. (See Table 1.1 for a summary of these differences.) These distinctions lead to many differences in the manner in which research and evaluation are conducted.

Of course, evaluation and research sometimes overlap. An evaluation study may add to our knowledge of laws or theories in a discipline. Research can inform our judgments and decisions regarding a program or policy. Yet, fundamental distinctions remain. Our earlier discussion highlights these differences to help those who are new to evaluation to see the ways in which evaluators behave differently than researchers. Evaluations may add to knowledge in a field, contribute to theory development, establish causal relationships, and provide explanations for the relationship between phenomena, but that is not its primary purpose. Its primary purpose is to assist stakeholders in making value judgments and decisions about whatever is being evaluated.

Action Research

A different type of research altogether is action research. Action research, originally conceptualized by Kurt Lewin (1946) and more recently developed by Emily Calhoun (1994, 2002), is research conducted collaboratively by professionals to 1213improve their practice. Such professionals might be social workers, teachers, or accountants who are using research methods and means of thinking to develop their practice. As Elliott (2005) notes, action research always has a developmental aim. Calhoun, who writes of action research in the context of education, gives examples of teachers working together to conceptualize their focus; to collect, analyze, and interpret data on the issue; and to make decisions about how to improve their practice as teachers and/or a program or curriculum they are implementing. The data collection processes may overlap with program evaluation activities, but there are key differences: Action research is conducted by professionals about their own work with a goal of improving their practice. Action research is also considered to be a strategy to change the culture of organizations to one in which professionals work collaboratively to learn, examine, and research their own practices. Thus, action research produces information akin to that in formative evaluations—information to be used for program improvement. The research is conducted by those delivering the program and, in addition to improving the element under study, has major goals concerning professional development and organizational change.

The Purposes of Evaluation

Consistent with our earlier definition of evaluation, we believe that the primary purpose of evaluation is to render judgments about the value of whatever is being evaluated. This view parallels that of Scriven (1967), who was one of the earliest to outline the purpose of formal evaluation. In his seminal paper, “The Methodology of Evaluation,” he argued that evaluation has a single goal or purpose: to determine the worth or merit of whatever is evaluated. In more recent writings, Scriven has continued his emphasis on the primary purpose of evaluation being to judge the merit or worth of an object (Scriven, 1996).

Yet, as evaluation has grown and evolved, other purposes have emerged. A discussion of these purposes sheds light on the practice of evaluation in today’s world. For the reader new to evaluation, these purposes illustrate the many facets of evaluation and its uses. Although we agree with Scriven’s historical emphasis on the purpose of evaluation, to judge the merit or worth of a program, policy, process, or product, we see these other purposes of evaluation at play as well.

Some years ago, Talmage (1982) argued that an important purpose of evaluation was “to assist decision makers responsible for making policy” (p. 594). And, in fact, providing information that will improve the quality of decisions made by policymakers continues to be a major purpose of program evaluation. Indeed, the rationale given for collecting much evaluation data today—by schools, by state and local governments, by the federal government, and by nonprofit organizations—is to help policymakers in these organizations make decisions about whether to continue programs, to initiate new programs, or, in other major ways, to change the funding or structure of a program. In addition to decisions made by policymakers, evaluation is intended to inform the decisions of many others, including program managers (principals, department heads), program staff (teachers, counselors, 1314health care providers, and others delivering the services offered by a program), and program consumers (clients, parents, citizens). A group of teachers may use evaluations of student performance to make decisions on program curricula or materials. Parents make decisions concerning where to send their children to school based on information on school performance. Students choose institutions of higher education based on evaluative information. The evaluative information or data provided may or may not be the most useful for making a particular decision, but, nevertheless, evaluation clearly serves this purpose.

For many years, evaluation has been used for program improvement. As we will discuss later in this chapter, Michael Scriven long ago identified program improvement as one of the roles of evaluation, though he saw that role being achieved through the initial purpose of judging merit and worth. Today, many see organizational and program improvement as a major, direct purpose of evaluation (Mark, Henry, & Julnes, 2000; Patton, 2008a; Preskill & Torres, 1998).

Program managers or those who deliver a program can make changes to improve the program based on the evaluation results. In fact, this is one of the most frequent uses of evaluation. There are many such examples: teachers using the results of student assessments to revise their curricula or pedagogical methods, health care providers using evaluations of patients’ use of medication to revise their means of communicating with patients about dosage and use, and trainers using feedback from trainees to change training to improve its application on the job. These are all ways that evaluation serves the purpose of program improvement.

Today, many evaluators see evaluation being used for program and organizational improvement in new ways. As we will describe in later chapters, Michael Patton often works today in what he calls “developmental evaluation,” working to assist organizations that do not have specific, measurable goals, but, instead, need evaluation to help them with ongoing progress, adaptation, and learning (Patton, 1994, 2005b). Hallie Preskill (Preskill, 2008; Preskill & Torres, 2000) and others (King, 2002; Baker & Bruner, 2006) have written about the role of evaluation in improving overall organizational performance by instilling new ways of thinking. In itself, the process of participating in an evaluation can begin to influence the ways that those who work in the organization approach problems. For example, an evaluation that involves employees in developing a logic model for the program to be evaluated or in examining data to draw some conclusions about program progress may prompt those employees to use such procedures or these ways of approaching a problem in the future and, thus, lead to organizational improvement.

The purpose of program or organizational improvement, of course, overlaps with others. When an evaluation is designed for program improvement, the evaluator must consider the decisions that those managing and delivering the program will make in using the study’s results for program improvement. So the purpose of the evaluation is to provide both decision making and program improvement. We will not split hairs to distinguish between the two purposes, but will simply acknowledge that evaluation can serve both purposes. Our goal is to expand your view of the various purposes for evaluation and to help you consider the purpose in your own situation or organization.

14 15

Some recent discussions of the purposes of evaluation move beyond these more immediate purposes to evaluation’s ultimate impact on society. Some evaluators point out that one important purpose of evaluation is helping give voice to groups who are not often heard in policy making or planning programs. Thus, House and Howe (1999) argue that the goal of evaluation is to foster deliberative democracy. They encourage the evaluator to work to help less powerful stakeholders gain a voice and to stimulate dialogue among stakeholders in a democratic fashion. Others highlight the role of the evaluator in helping bring about greater social justice and equality. Greene, for example, notes that values inevitably influence the practice of evaluation and, therefore, evaluators can never remain neutral. Instead, they should recognize the diversity of values that emerge and arise in an evaluation and work to achieve desirable values of social justice and equity (Greene, 2006).

Carol Weiss (1998b) and Gary Henry (2000) have argued that the purpose of evaluation is to bring about social betterment. Mark, Henry, and Julnes (2000) define achieving social betterment as “the alleviation of social problems, meeting of human needs” (p. 190). And, in fact, evaluation’s purpose of social betterment is at least partly reflected in the Guiding Principles, or ethical code, adopted by the American Evaluation Association. One of those principles concerns the evaluator’s responsibilities for the general and public welfare. Specifically, Principle E5 states the following:

 Evaluators have obligations that encompass the public interest and good. Because the public interest and good are rarely the same as the interests of any particular group (including those of the client or funder) evaluators will usually have to go beyond analysis of particular stakeholder interests and consider the welfare of society as a whole. (American Evaluation Association, 2004)

This principle has been the subject of more discussion among evaluators than other principles, and deservedly so. Nevertheless, it illustrates one important purpose of evaluation. Evaluations are concerned with programs and policies that are intended to improve society. Their results provide information on the choices that policymakers, program managers, and others make in regard to these programs. As a result, evaluators must be concerned with their purposes in achieving the social betterment of society. Writing in 1997 about the coming twenty-first century, Chelimsky and Shadish emphasized the global perspective of evaluation in achieving social betterment, extending evaluation’s context in the new century to worldwide challenges. These include new technologies, demographic imbalances across nations, environmental protection, sustainable development, terrorism, human rights, and other issues that extend beyond one program or even one country (Chelimsky & Shadish, 1997).

Finally, many evaluators continue to acknowledge the purpose of evaluation in extending knowledge (Donaldson, 2007; Mark, Henry, & Julnes, 2000). Although adding to knowledge is the primary purpose of research, evaluation studies can add to our knowledge of social science theories and laws. They provide an opportunity to test theories in real-world settings or to test existing theories or laws with new groups by examining whether those theories hold true in new 1516settings with different groups. Programs or policies are often, though certainly not always, based on some theory or social science principles.3 Evaluations provide the opportunity to test those theories. Evaluations collect many kinds of information that can add to our knowledge: information describing client groups or problems, information on causes or consequences of problems, tests of theories concerning impact. For example, Debra Rog conducted an evaluation of a large intervention program to help homeless families in the early 1990s (Rog, 1994; Rog, Holupka, McCombs-Thornton, Brito, & Hambrick, 1997). At the time, not much was known about homeless families and some of the initial assumptions in planning were incorrect. Rog adapted her evaluation design to learn more about the circumstances of homeless families. Her results helped to better plan the program, but also added to our knowledge about homeless families, their health needs, and their circumstances. In our discussion of the differences between research and evaluation, we emphasized that the primary purpose of research is to add to knowledge in a field and that this is not the primary purpose of evaluation. We continue to maintain that distinction. However, the results of some evaluations can add to our knowledge of social science theories and laws. This is not a primary purpose, but simply one purpose that an evaluation may serve.

In closing, we see that evaluation serves many different purposes. Its primary purpose is to determine merit or worth, but it serves many other valuable purposes as well. These include assisting in decision making; improving programs, organizations, and society as a whole; enhancing democracy by giving voice to those with less power; and adding to our base of knowledge.

Roles and Activities of Professional Evaluators

Evaluators as practitioners play numerous roles and conduct multiple activities in performing evaluation. Just as discussions on the purposes of evaluation help us to better understand what we mean by determining merit and worth, a brief discussion of the roles and activities pursued by evaluators will acquaint the reader with the full scope of activities that professionals in the field pursue.

A major role of the evaluator that many in the field emphasize and discuss is that of encouraging the use of evaluation results (Patton, 2008a; Shadish, 1994). While the means for encouraging use and the anticipated type of use may differ, considering use of results is a major role of the evaluator. In Chapter 17, we will discuss the different types of use that have been identified for evaluation and various means for increasing that use. Henry (2000), however, has cautioned that focusing primarily on use can lead to evaluations focused solely on program and organizational improvement and, ultimately, avoiding final decisions about merit and worth. His concern is appropriate; however, if the audience for the evaluation is one that is making decisions about the program’s merit and worth, this problem may be avoided. (See discussion of formative and summative evaluation in this chapter.) Use is certainly central to evaluation, as demonstrated by the prominent role it plays in the professional standards and codes of evaluation. (See Chapter 3.)

3 The term “evidence-based practice” emerges from the view that programs should be designed around social science research findings when basic research, applied research, or evaluation studies have found that a given program practice or action leads to the desired, intended outcomes.

16 17

Others’ discussions of the role of the evaluator illuminate the ways in which evaluators might interact with stakeholders and other users. Rallis and Rossman (2000) see the role of the evaluator as that of a critical friend. They view the primary purpose of evaluation as learning and argue that, for learning to occur, the evaluator has to be a trusted person, “someone the emperor knows and can listen to. She is more friend than judge, although she is not afraid to offer judgments” (p. 83). Schwandt (2001a) describes the evaluator in the role of a teacher, helping practitioners develop critical judgment. Patton (2008a) envisions evaluators in many different roles including facilitator, collaborator, teacher, management consultant, organizational development (OD) specialist, and social-change agent. These roles reflect his approach to working with organizations to bring about developmental change. Preskill and Torres (1998) stress the role of the evaluator in bringing about organizational learning and instilling a learning environment. Mertens (1999), Chelimsky (1998), and Greene (1997) emphasize the important role of including stakeholders, who often have been ignored by evaluation. House and Howe (1999) argue that a critical role of the evaluator is stimulating dialogue among various groups. The evaluator does not merely report information, or provide it to a limited or designated key stakeholder who may be most likely to use the information, but instead stimulates dialogue, often bringing in disenfranchised groups to encourage democratic decision making.

Evaluators also have a role in program planning. Bickman (2002), Chen (1990), and Donaldson (2007) emphasize the important role that evaluators play in helping articulate program theories or logic models. Wholey (1996) argues that a critical role for evaluators in performance measurement is helping policymakers and managers select the performance dimensions to be measured as well as the tools to use in measuring those dimensions.

Certainly, too, evaluators can play the role of the scientific expert. As Lipsey (2000) notes, practitioners want and often need evaluators with the “expertise to track things down, systematically observe and measure them, and compare, analyze, and interpret with a good faith attempt at objectivity” (p. 222). Evaluation emerged from social science research. While we will describe the growth and emergence of new approaches and paradigms, and the role of evaluators in educating users to our purposes, stakeholders typically contract with evaluators to provide technical or “scientific” expertise and/or an outside “objective” opinion. Evaluators can occasionally play an important role in making program stakeholders aware of research on other similar programs. Sometimes, the people managing or operating programs or the people making legislative or policy decisions on programs are so busy fulfilling their primary responsibilities that they are not aware of other programs or agencies that are doing similar things and the research conducted on these activities. Evaluators, who typically explore existing research on similar programs to identify potential designs and measures, can play the role 1718of scientific expert in making stakeholders aware of research. (See, for example, Fitzpatrick and Bledsoe [2007] for a discussion of Bledsoe’s role in informing stakeholders of existing research on other programs.)

Thus, the evaluator takes on many roles. In noting the tension between advocacy and neutrality, Weiss (1998b) writes that the role(s) evaluators play will depend heavily on the context of the evaluation. The evaluator may serve as a teacher or critical friend in an evaluation designed to improve the early stages of a new reading program. The evaluator may act as a facilitator or collaborator with a community group appointed to explore solutions to problems of unemployment in the region. In conducting an evaluation on the employability of new immigrant groups in a state, the evaluator may act to stimulate dialogue among immigrants, policymakers, and nonimmigrant groups competing for employment. Finally, the evaluator may serve as an outside expert in designing and conducting a study for Congress on the effectiveness of annual testing in improving student learning.

In carrying out these roles, evaluators undertake many activities. These include negotiating with stakeholder groups to define the purpose of evaluation, developing contracts, hiring and overseeing staff, managing budgets, identifying disenfranchised or underrepresented groups, working with advisory panels, collecting and analyzing and interpreting qualitative and quantitative information, communicating frequently with various stakeholders to seek input into the evaluation and to report results, writing reports, considering effective ways to disseminate information, meeting with the press and other representatives to report on progress and results, and recruiting others to evaluate the evaluation (metaevaluation). These, and many other activities, constitute the work of evaluators. Today, in many organizations, that work might be conducted by people who are formally trained and educated as evaluators, attend professional conferences and read widely in the field, and identify their professional role as an evaluator, or by staff who have many other responsibilities—some managerial, some working directly with students or clients—but with some evaluation tasks thrown into the mix. Each of these will assume some of the roles described previously and will conduct many of the tasks listed.

Uses and Objects of Evaluation

At this point, it might be useful to describe some of the ways in which evaluation can be used. An exhaustive list would be prohibitive, filling the rest of this book and more. Here we provide only a few representative examples of uses made of evaluation in selected sectors of society.

Examples of Evaluation Use in Education

 1. To empower teachers to have more say in how school budgets are allocated

 2. To judge the quality of school curricula in specific content areas

 3. To accredit schools that meet or exceed minimum accreditation standards

 18 19

 4. To determine the value of a middle school’s block scheduling

 5. To satisfy an external funding agency’s demands for reports on effectiveness of school programs it supports

 6. To assist parents and students in selecting schools in a district with school choice

 7. To help teachers improve their reading program to encourage more voluntary reading

Examples of Evaluation Use in Other Public and Nonprofit Sectors

 1. To decide whether to expand an urban transit program and where it should be expanded

 2. To establish the value of a job training program

 3. To decide whether to modify a low-cost housing project’s rental policies

 4. To improve a recruitment program for blood donors

 5. To determine the impact of a prison’s early-release program on recidivism

 6. To gauge community reaction to proposed fire-burning restrictions to improve air quality

 7. To determine the effect of an outreach program on the immunization of infants and children

Examples of Evaluation Use in Business and Industry

 1. To improve a commercial product

 2. To judge the effectiveness of a corporate training program on teamwork

 3. To determine the effect of a new flextime policy on productivity, recruitment, and retention

 4. To identify the contributions of specific programs to corporate profits

 5. To determine the public’s perception of a corporation’s environmental image

 6. To recommend ways to improve retention among younger employees

 7. To study the quality of performance appraisal feedback

One additional comment about the use of evaluation in business and industry may be warranted. Evaluators unfamiliar with the private sector are sometimes unaware that personnel evaluation is not the only use made of evaluation in business and industry settings. Perhaps that is because the term “evaluation” has been absent from the descriptors for many corporate activities and programs that, when examined, are decidedly evaluative. Activities labeled as quality assurance, quality control, research and development, Total Quality Management (TQM), or Continuous Quality Improvement (CQI) turn out, on closer inspection, to possess many characteristics of program evaluation.

Uses of Evaluation Are Generally Applicable

As should be obvious by now, evaluation methods are clearly portable from one arena to another. The use of evaluation may remain constant, but the entity it is applied to—that is, the object of the evaluation—may vary widely. Thus, evaluation 1920may be used to improve a commercial product, a community training program, or a school district’s student assessment system. It could be used to build organizational capacity in the Xerox Corporation, the E. F. Lilly Foundation, the Minnesota Department of Education, or the Utah Division of Family Services. Evaluation can be used to empower parents in the San Juan County Migrant Education Program, workers in the U.S. Postal Service, employees of Barclays Bank of England, or residents in east Los Angeles. Evaluation can be used to provide information for decisions about programs in vocational education centers, community mental health clinics, university medical schools, or county cooperative extension offices. Such examples could be multiplied ad infinitum, but these should suffice to make our point.

In some instances, so many evaluations are conducted of the same type of object that it prompts suggestions for techniques found to be particularly helpful in evaluating something of that particular type. An example would be Kirkpatrick’s (1977; 1983; 2006) model for evaluating training efforts. In several areas, concern about how to evaluate broad categories of objects effectively has led to the development of various subareas within the field of evaluation, such as product evaluation, personnel evaluation, program evaluation, policy evaluation, and performance evaluation.

Some Basic Types of Evaluation

Formative and Summative Evaluation

Scriven (1967) first distinguished between the formative and summative roles of evaluation. Since then, the terms have become almost universally accepted in the field. In practice, distinctions between these two types of evaluation may blur somewhat, but the terms serve an important function in highlighting the types of decisions or choices that evaluation can serve. The terms, in fact, contrast two different types of actions that stakeholders might take as a result of evaluation.

An evaluation is considered to be formative if the primary purpose is to provide information for program improvement. Often, such evaluations provide information to judge the merit or worth of one part of a program. Three examples follow:

1. Planning personnel in the central office of Perrymount School District have been asked by the school board to plan a new, and later, school day for the local high schools. This is based on research showing that adolescents’ biological clocks cause them to be more groggy in the early morning hours and on parental concerns about teenagers being released from school as early as 2:30 p.m. A formative evaluation will collect information (surveys, interviews, focus groups) from parents, teachers and school staff, and students regarding their views on the current school schedule calendar and ways to change and improve it. The planning staff will visit other schools using different schedules to observe these schedules and to interview school staff on their perceived effects. The planning staff will then give the information to the Late Schedule Advisory Group, which will make final recommendations for changing the existing schedule. 20 21 2. Staff with supervisory responsibilities at the Akron County Human Resources Department have been trained in a new method for conducting performance appraisals. One of the purposes of the training is to improve the performance appraisal interview so that employees receiving the appraisal feel motivated to improve their performance. The trainers would like to know if the information they are providing on conducting interviews is being used by those supervisors who complete the program. They plan to use the results to revise this portion of the training program. A formative evaluation might include observing supervisors conducting actual, or mock, interviews, as well as interviewing or conducting focus groups with both supervisors who have been trained and employees who have been receiving feedback. Feedback for the formative evaluation might also be collected from participants in the training through a reaction survey delivered either at the conclusion of the training or a few weeks after the training ends, when trainees have had a chance to practice the interview. 3. A mentoring program has been developed and implemented to help new teachers in the classroom. New teachers are assigned a mentor, a senior teacher who will provide them with individualized assistance on issues ranging from discipline to time management. The focus of the program is on helping mentors learn more about the problems new teachers are encountering and helping them find solutions. Because the program is so individualized, the assistant principal responsible for overseeing the program is concerned with learning whether it is being implemented as planned. Are mentors developing a trusting relationship with the new teachers and learning about the problems they encounter? What are the typical problems encountered? The array of problems? For what types of problems are mentors less likely to be able to provide effective assistance? Interviews, logs or diaries, and observations of meetings between new teachers and their mentors will be used to collect data to address these issues. The assistant principal will use the results to consider how to better train and lead the mentors. In contrast to formative evaluations, which focus on program improvement, summative evaluations are concerned with providing information to serve decisions or assist in making judgments about program adoption, continuation, or expansion. They assist with judgments about a program’s overall worth or merit in relation to important criteria. Scriven (1991a) has defined summative evaluation as “evaluation done for, or by, any observers or decision makers (by contrast with developers) who need valuative conclusions for any other reasons besides development” (p. 20). Robert Stake has memorably described the distinction between the two in this way: “When the cook tastes the soup, that’s formative evaluation; when the guest tastes it, that’s summative evaluation” (cited by Scriven, 1991a, p. 19). In the following examples we extend the earlier formative evaluations into summative evaluations.

 1. After the new schedule is developed and implemented, a summative evaluation might be conducted to determine whether the schedule should be continued and expanded to other high schools in the district. The school board might be 2122the primary audience for this information because it is typically in a position to make the judgments concerning continuation and expansion or termination, but others—central office administrators, principals, parents, students, and the public at large—might be interested stakeholders as well. The study might collect information on attendance, grades, and participation in after-school activities. Other unintended side effects might be examined, such as the impact of the schedule on delinquency, opportunities for students to work after school, and other afternoon activities. 2. To determine whether the performance appraisal program should be continued, the director of the Human Resource Department and his staff might ask for an evaluation of the impact of the new performance appraisal on job satisfaction and performance. Surveys of employees and existing records on performance might serve as key methods of data collection. 3. Now that the mentoring program for new teachers has been tinkered with for a couple of years using the results of the formative evaluation, the principal wants to know whether the program should be continued. The summative evaluation will focus on turnover, satisfaction, and performance of new teachers.

Note that the audiences for formative and summative evaluation are very different. In formative evaluation, the audience is generally the people delivering the program or those close to it. In our examples, they were those responsible for developing the new schedule, delivering the training program, or managing the mentoring program. Because formative evaluations are designed to improve programs, it is critical that the primary audience be people who are in a position to make changes in the program and its day-to-day operations. Summative evaluation audiences include potential consumers (students, teachers, employees, managers, or officials in agencies that could adopt the program), funding sources, and supervisors and other officials, as well as program personnel. The audiences for summative evaluations are often policymakers or administrators, but can, in fact, be any audience with the ability to make a “go–no go” decision. Teachers make such decisions with curricula. Consumers (clients, parents, and students) make decisions about whether to participate in a program based on summative information or their judgments about the overall merit or worth of a program.

A Balance between Formative and Summative.

It should be apparent that both formative and summative evaluation are essential because decisions are needed during the developmental stages of a program to improve and strengthen it, and again, when it has stabilized, to judge its final worth or determine its future. Unfortunately, some organizations focus too much of their work on summative evaluations. This trend is noted in the emphases of many funders today on impact or outcome assessment from the beginning of a program or policy. An undue emphasis on summative evaluation can be unfortunate because the development process, without formative evaluation, is incomplete and inefficient. Consider the foolishness of developing a new aircraft design and submitting it to a summative 2223test flight without first testing it in the formative wind tunnel. Program test flights can be expensive, too, especially when we haven’t a clue about the probability of success.

Formative data collected during the early stages of a program can help identify problems in the program model or theory or in the early delivery of the program that can then be modified or corrected. People delivering the program may need more training or resources to effectively implement the model. The model may have to be adapted because the students or clients being served are not exactly as program developers anticipated. Perhaps they have different learning strategies or less knowledge, skills, or motivation than anticipated; therefore, the training program or class curriculum should be expanded or changed. In other cases, students or clients who participate in a program may have more, or different, skills or problems than program planners anticipated. The program, then, must be adapted to address those.4 So, a formative evaluation can be very useful at the beginning of a program to help it succeed in achieving its intended outcomes.

Conversely, some organizations may avoid summative evaluations. Evaluating for improvement is critical, but, ultimately, many products and programs should be judged for their overall merit and worth. Henry (2000) has noted that evaluation’s emphasis on encouraging use of results can lead us to serving incremental, often formative, decisions and may steer us away from the primary purpose of evaluation—determining merit and worth.

Although formative evaluations more often occur in the early stages of a program’s development and summative evaluations more often occur in its later stages, it would be an error to think they are limited to those time frames. Well-established programs can benefit from formative evaluations. Some new programs are so problematic that summative decisions are made to discontinue. However, the relative emphasis on formative and summative evaluation changes throughout the life of a program, as suggested in Figure 1.1, although this generalized concept obviously may not precisely fit the evolution of any particular program.

An effort to distinguish between formative and summative evaluation on several dimensions appears in Table 1.2. As with most conceptual distinctions, formative and summative evaluation are often not as easy to distinguish in the real world as they seem in these pages. Scriven (1991a) has acknowledged that the two are often profoundly intertwined. For example, if a program continues beyond a summative evaluation study, the results of that study may be used for both summative and, later, formative evaluation purposes. In practice, the line between formative and summative is often rather fuzzy.

4 See the interview with Stewart Donaldson about his evaluation of a work-training program (Fitzpatrick & Donaldson, 2002) in which he discusses his evaluation of a program that had been successful in Michigan, but was not adapted to the circumstances of California sites, which differed in the reasons why people were struggling with returning to the workforce. The program was designed anticipating that clients would have problems that these clients did not have.

|  |
| --- |
| **What is Evaluation?**  |

|  |
| --- |
| http://myeclassonline.com/ec/courses/AUO_files/AU_spacer.gif |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| The authoritative, classic, and simple definition of evaluation comes from Michael Scriven: judging the worth or merit of something (p. 5).\* In almost all cases, evaluation involves setting standards for judging quality, collecting information, and applying the standards to assess worth (p. 5). The text elaborates on these issues.\*All page references, unless otherwise specified, are from the course text: Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation* (3rd ed.). Boston, MA: Pearson.

|  |
| --- |
| **Differences between Research and Evaluation**  |

|  |
| --- |
| http://myeclassonline.com/ec/courses/AUO_files/AU_spacer.gif |

|  |
| --- |
| **The Focus and Aims of Each** ***Critical question: What outcomes did participants achieve in this program over and above what would have occurred, had the program not intervened?***Program evaluation focuses on decisions. Research focuses on answering questions about phenomena to discover new knowledge and test theories/hypotheses. The aim of research is truth. The aim of evaluation is action. There are several ways that good program evaluation is similar to good research: 1. It is a systematic process of looking for ways that an action might affect an outcome. In research, we call the action the Independent Variable (IV) and the effect or outcome the Dependent Variable (DV). In evaluation, the program is the IV and the behaviors, attitudes, skills, performance, etc. that we are targeting for change are the DVs.
2. We need good operational definitions of what is being done in the program (IV) and what the specific effects or outcomes are that we will be observing (DVs).
3. We need to use techniques and instruments for measurement of the DVs that are reliable and valid.
4. We need to try to evaluate the probable effect of our program (IV) on the DVs in a way that separates it from other activities and factors that might be affecting the DVs in order to evaluate the specific contributions that come from our program.

**Purposes, Roles, and Uses**Students new to program evaluation are inclined to expect **the** definition of evaluation, of the role of the evaluator, and of the uses and objects of evaluation.  Many will expect an authoritative definition followed by a statement that begins, “The evaluator has three (or four or five or whatever) main roles in the evaluation context. The use of evaluation and what is evaluated might be expected to have a similar dimension. As you study the course textbook and related materials, you will discover what it means to be a young field, as is program evaluation. There are many very bright people who do some really good work, but sometimes very bright people have a variety of conceptions about a range of issues, including the purpose of evaluation, the role of evaluators, and the uses and objects of evaluation. They all say something a little different, but, paradoxically, they are all correct. Therefore, as you read the section that follows on the differences between research and evaluation, keep in mind that you are a participant-observer in a young, vibrant, and important field: program evaluation.  |

|  |
| --- |
| **Formative vs. Summative Evaluation**  |

|  |
| --- |
| http://myeclassonline.com/ec/courses/AUO_files/AU_spacer.gif |

|  |
| --- |
| **Formative evaluation** Formative evaluation is a method of judging the worth of a program while the program activities are forming or happening. Formative evaluation focuses on the process (Bhola, 1990). Here are some sample objectives and steps for a formative evaluation:**Objectives** * To assess a training course or workshop as it progresses
* To find out the extent of program implementation
* To determine improvements and adjustments needed to attain the training objectives

**Steps** 1. Observe trainee behavior.
2. Have informal talks about the training activity with the trainees.
3. Give short tests to trainees.
4. Hold group discussions with the trainees to gain feedback.
5. Answer the following questions:
	* Did you identify training needs correctly?
	* Have you noticed other areas which need attention?
	* Are there indications that the training objectives will be met?
	* Do the objectives need to be revised?
	* Are the training topics being taught?
	* Have additional training topics come up which need to be taught?
	* Are the training methods appropriate, or do you need to adjust them?

**Outcome/impact evaluation focuses on the program’s impact on the participants.** **Ideally, outcome evaluations occur annually.*****Seek to evaluate if the intervention is achieving its intended effect?***(e.g., An outcome evaluation might look at whether completing a college preparatory program encourages more students to enroll in four-year universities.) Outcome evaluations are often conducted by an outside evaluator and result in lessons learned that can be generalized beyond program participants.  |
| Summative Evaluation  |

|  |
| --- |
| http://myeclassonline.com/ec/courses/AUO_files/AU_spacer.gif |

|  |
| --- |
| Summative evaluation is a method of judging the worth of a program at the end of the program activities. The focus is on the outcome (Bhola, 1990). Here are some sample objectives and steps for a summative evaluation: **Objectives** * To find out the extent to which training objectives were achieved and the original training plan implemented.
* To help you decide whether a training activity or any of its parts should be revised, continued, or terminated.

**Steps** 1. Have the trainees conduct demonstration lessons.
2. Have others observe trainees' behavior.
3. Give comprehensive tests to the trainees.
4. Gather impressions of trainers and trainees through questionnaires and interviews.
5. Answer the following questions:
	* Did you meet the training objectives?
	* Will you need to improve and modify some areas?
	* Should you conduct the training activity again?
	* How can you help the trainees attain further training?

**Needs Assessment, Process, and Outcomes**While formative and summative assessment dimensions examine judgments and decisions, there is another dimension that includes the kinds of questions asked, or the focus of the evaluation. There are three areas to be considered:1. Needs assessment
2. Process studies
3. Outcome studies

The authors of your text describe these issues in detail and provide the rationale for why it is important to examine question types along with judgments and decisions.  |

Program evaluation was not an idea that someone dreamed up simply to give Argosy students a challenge. This area has an interesting and significant history, even though it is a relatively young field, because it is transdisciplinary—a number of different disciplines come together and work on evaluation issues. Therefore, at Western Michigan University, a center for program evaluation, persons from a variety of disciplines work together on the theory, practice, and utilization of evaluation.As you read the chapter which reviews the origins and current trends in evaluation, you will see that program evaluation has had a hand in monitoring and changing institutions—primarily education and medicine—and government for over a century. As that happened, well-known and respected leaders in education, psychology, and philosophy have left their marks. Reading about these trends will provide a perspective on the area which has a unique past in the history and philosophy of science.

|  |
| --- |
| Origins and Current Trends in Evaluation |

|  |
| --- |
| http://myeclassonline.com/ec/courses/AUO_files/AU_spacer.gif |

|  |
| --- |
| Program evaluation was not an idea that someone dreamed up simply to give Argosy students a challenge. This area has an interesting and significant history, even though it is a relatively young field, because it is transdisciplinary—a number of different disciplines come together and work on evaluation issues. Therefore, at Western Michigan University, a center for program evaluation, persons from a variety of disciplines work together on the theory, practice, and utilization of evaluation.As you read the chapter which reviews the origins and current trends in evaluation, you will see that program evaluation has had a hand in monitoring and changing institutions—primarily education and medicine—and government for over a century. As that happened, well-known and respected leaders in education, psychology, and philosophy have left their marks. Reading about these trends will provide a perspective on the area which has a unique past in the history and philosophy of science.  |

 |