1

Evaluation's Basic Purpose, Uses, and Conceptual Distinctions

Orienting Questions

1. How does evaluation serve society? 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? Give some examples from your experience with evaluation.
4. What are the major differences between formative and summative evaluations?
5. What is an example of an issue an evaluator might address in a needs assessment, a process evaluation, and an outcome evaluation?
6. Under what circumstances might an external evaluator be preferable to an internal 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 versatile employees; combating disease and mental illness; fighting discrimination; reducing crime, drug abuse, and child and spouse abuse. More recently, pursuing and balancing environmental and economic goals and working to insure peace and economic growth in developing countries have become prominent concerns. 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, and national agencies have launched a veritable flotilla of programs aimed at identifying and eliminating the underlying causes of these problems. Specific programs judged to have been ineffective have been "mothballed" or sunk outright, usually 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, policy makers and managers have been faced with tough choices, being forced to cancel some programs or program components to provide sufficient funds to launch new ones or continue others.

To make such choices intelligently, policy makers need good information about the relative effectiveness of each program. Which programs are working well? Which poorly? What are the programs' relative costs and benefits? Similarly, each program manager needs to know how well different parts of the program are working. Are some parts contributing more than others? What can be done to improve those parts of the program that are not contributing what 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.

A Brief Definition of Evaluation

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, recently noted there are nearly sixty 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 (cited in Patton, 2000) 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" (p. 7). This chapter will introduce the reader to the array of variations in application, but, at this point, we would like to 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, 1999; 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 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 **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.

Differences in Evaluation and Research

It may be important here to distinguish between evaluation and research. While some methods of evaluation emerged from social science research traditions,

there are important distinctions between evaluation and research.[1] 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. While the results of an evaluation study may contribute to knowledge development (Mark, Henry, Julnes, 1999), that is a secondary concern in evaluation. Evaluation's primary purpose is to help those who hold a stake in whatever is being evaluated (stakeholders), often consisting of many different groups, 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 he is conducting would be regarded as a failure if it produced no data on the usefulness 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—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 describe a particular thing. Sometimes, describing that thing involves examining causal relationships; often, it does not. Whether the evaluation focuses on a causal issue depends on the 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 and his 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.

Another difference concerns *generalizability of results.* Given evaluation's purpose of describing 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 and have less desire to generalize to other settings than a researcher would. (Note that the setting or context may be large, national programs with many sites, or small, a program in one school.) In contrast, because the purpose of research is to add to general knowledge, the methods are designed to maximize generalizability to many different settings. If one's findings are to add to knowledge in a field, ideally, the results should transcend the particulars of time and setting.

Research and evaluation differ further in the *criteria* or standards used to judge their adequacy. Two important criteria for judging the adequacy of research are internal validity, or causality, and external validity, or generalizability to other

[1]Research itself varies across a wide spectrum, from basic research (which we use here to highlight the distinction of research and evaluation) to applied research, which sometimes resembles evaluation in being applied to solve educational, social, and private sector problems or issues. For a more extended discussion of the differences and similarities of research and evaluation, see Worthen and Sanders, 1973.

settings and other times. These criteria, however, are not sufficient, or appropriate, for judging the quality of an evaluation. 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 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 were developed by the Joint Committee on Standards for Evaluation to help both users of evaluation and evaluators themselves understand what evaluations should do. (See Chapter 18 for more on the Standards.)

Finally, the *preparation* of researchers and evaluators differs significantly. Researchers are trained in depth in a single discipline, their field of inquiry. This approach is appropriate because the researcher's work, in almost all cases, will remain within a single discipline or field. Evaluators, by contrast, are responding to the needs of clients and stakeholders with many different information needs and operating in many different settings. As such, evaluators' education must be interdisciplinary. Only through interdisciplinary training can evaluators 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 that they can choose those most appropriate for the particular program and needs of stakeholders. Finally, evaluators differ from researchers in that they must establish personal working relationships with clients. As a result, they require preparation in interpersonal and communication skills (Fitzpatrick, 1994).

Sanders (1979) identified several general areas of competence important for evaluators. These included the ability to describe the object and context of an evaluation; to conceptualize appropriate purposes and frameworks for the evaluation; to identify and select appropriate evaluation questions, information needs, and sources of information; to select means for collecting and analyzing information; to determine the value of the object of an evaluation; to communicate plans and results effectively to audiences; to manage the evaluation; to maintain ethical standards; to adjust for external factors influencing the evaluation; and to evaluate the evaluation (metaevaluation).

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, the generalizability of their results, and the criteria used to judge their work. 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 discussion above highlights these differences to help those 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.

We will discuss shortly the matter of how one's definition of *evaluation* is the product of what one believes the purpose of evaluation to be. First, however, we need to distinguish between systematic, formal evaluation studies—the focus of this book—and the much more informal, even casual evaluation that is a part of our everyday life.

Informal versus Formal Evaluation

Evaluation is not a new concept. If one focuses on the aspect of "examining and judging, to determine value," then the practice of evaluation doubtlessly long preceded its definition, tracing its roots back to 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, or 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 they relinquished their crossbows, perfected and improved on the Welsh longbow, and the English armies 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 everyday. Practitioners, managers, and policy makers 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 policy maker 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 by both 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 evaluate formally everything it does. Often informal evaluation is the only practical 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 (2001) 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 overapplication 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 (1999) echo this concept when they describe evaluation as a form of assisted sensemaking. 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.

Distinguishing between Evaluation's Purposes and Evaluators' Roles and Activities

We mentioned earlier that how one defines evaluation stems from what one perceives evaluation's basic purpose to be. We treat that topic in more depth in this section as we attempt to separate the basic *purpose* of evaluation from the *roles* a

professional evaluator can play in different evaluations and the *activities* undertaken to complete an evaluation successfully.

Purposes of Evaluation

Just as evaluators are not all agreed on one final, authoritative definition of evaluation, they are by no means unanimous in what they believe evaluation's purpose to be. Consistent with our earlier definition of evaluation, we believe that the basic purpose of evaluation is to render judgments about the value of whatever is being evaluated. Many different uses may be made of those value judgments, as we shall discuss shortly, but in every instance the central purpose of the evaluative act is the same: to determine the merit or worth of some thing (in program evaluation, of the program or some part of it).

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 noted that evaluation plays many roles but argued that it has a single goal: to determine the worth or merit of whatever is evaluated. He distinguished between the goal of evaluation, providing answers to significant evaluative questions that are posed, and evaluation roles, the ways in which those answers are used. According to Scriven, evaluation's goal usually relates to value questions, requires judgments of worth or merit, and is conceptually distinct from its roles. Scriven made the distinction this way:

> In terms of goals, we may say that evaluation attempts to answer certain *types of questions* about certain *entities*. The entities are the various . . . instruments (processes, personnel, procedures, programs, etc.). The types of question include questions of the form: *How well* does this instrument perform (with respect to such-and-such criteria)? Does it perform *better* than this other instrument? What *merits*, or drawbacks does this instrument have . . . ? Is the use of this instrument *worth* what it's costing?
>
> . . . But the roles which evaluation has in a particular . . . context may be enormously various; it may form part of a . . . training activity, of the process of curriculum development, of a field experiment, . . . of . . . an executive training program, a prison, or a classroom (pp. 40–41).

In the decades since this original distinction between evaluation's basic purpose (goal) and its diverse uses (roles), Scriven (1980, 1991a, 1991c) has greatly elaborated his view without abandoning it. While he has more recently added that "evaluation is concerned with significance, not just merit and worth" (1994, p. 380), he continues to present powerful philosophical arguments that evaluation of any object (e.g., a marketing plan, a school curriculum, or a residential treatment facility for drug abusers) is undertaken to identify and apply defensible criteria to determine its worth, merit, or quality.

This view of evaluation's basic purpose has been most widely adopted by prominent evaluators working in the field of education, ultimately being incorporated into the Program Evaluation Standards developed by the Joint Committee on Standards for Educational Evaluation (1994). Yet, while this view is broadly held, other articulate colleagues have argued that evaluation has several purposes. For example, Talmage (1982) notes that "three purposes appear most frequently in definitions of evaluation: (1) to render judgments on the worth of a program; (2) to assist decision makers responsible for deciding policy; and (3) to serve a political function" (p. 594). Talmage also notes that, while these purposes are not mutually exclusive, they are clearly different. Rallis and Rossman (2000) have argued that the fundamental purpose of evaluation is learning, helping practitioners and others better understand and interpret their observations.

Some recent discussions of the purposes of evaluation move beyond these more immediate purposes to evaluation's ultimate impact on society. Weiss (1998b) and Henry (2000) have argued that the purpose of evaluation is to bring about social betterment. Mark, Henry, and Julnes (1999) define achieving social betterment as "the alleviation of social problems, meeting of human needs" (p. 190). Chelimsky (1997) takes a global perspective, extending evaluation's context in the new century to worldwide challenges rather than domestic ones: 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. House and Howe (1999) argue that the goal of evaluation is to foster deliberative democracy. This goal, which they recognize as idealistic, calls on the evaluator to work to help less powerful stakeholders gain a voice and to stimulate dialogue among stakeholders in a democratic fashion.

Mark, Henry, and Julnes (1999) have articulated four different purposes for evaluation: assessment of merit and worth, oversight and compliance, program and organizational improvement, and knowledge development. They note that oversight and compliance is often viewed as achieving the purpose of assessing merit and worth, but because such activities generally focus only on whether the designated services are delivered to the appropriate clients, Mark and his coauthors do not see them as effectively contributing to decisions about overall merit and worth. Similarly, they separate program and organizational improvement from merit and worth because, while such activities can focus on the merit and worth of subsets of programs, such evaluations do not lead to overall judgments of merit and worth. They note, as do we, that knowledge development can be a useful outcome or corollary to evaluation. We would emphasize, however, that it is not the primary purpose.

We will expand on these differing views of evaluation later in the book. At this point, we want to present them to introduce the reader to differing views on purposes. These views are useful in shedding light for the reader new to evaluation on the types of things evaluation might do and what evaluation means. Determining merit and worth is a quite abstract concept. The views of these different authors, we would argue, help illustrate what determining merit and worth means and what it can involve. For this text, we will continue to define the primary purpose of evaluation as determining merit and worth because it emphasizes the

valuing component of evaluation that we see as critical and because we believe many, if not most, of these distinctions can be subsumed within determining merit and worth.

Roles and Activities of Professional Evaluators

Scriven (1967) discusses the roles of evaluation in terms of how evaluation is used, but 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 use (Patton, 1996; 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 16, we will elaborate the types of uses of evaluation and ways to maximize 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 16).

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 a critical friend. As noted, they view the primary purpose of evaluation as learning. They then 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 (2001) describes the evaluator in the role of a teacher, helping practitioners develop critical judgment. Patton (1996) envisions evaluators in many different roles including facilitator, collaborator, teacher, management consultant, OD specialist, and social-change agent. These roles reflect his approach to working with organizations to bring about developmental change. Preskill & Torres (1999) 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 (see Chapter 2 on recent trends). 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 (2001) and Chen (1990) emphasize the important role 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 emergency 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.

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 underemployment in the region. In conducting an evaluation on the employability of new immigrant groups to a state, the evaluator may act to stimulate dialogue among immigrants, policy makers, and non-immigrant 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 may 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 direct work with students or clients, and some evaluation tasks thrown into the mix. Each of these will assume some of the roles described above 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 potentially 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 about how school budgets are allocated
2. To judge the quality of school curricula in specific content areas
3. To accredit schools that meet minimum accreditation standards
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 implement an urban development program
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 cost–benefit contribution of a new sports stadium for a metropolitan area

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, Total Quality Management (TQM), or Continuous Quality Improvement (CQI) turn out, on closer inspection, to possess many characteristics of program evaluation. In Chapter 20 we treat this topic more fully. Suffice it to say here that many uses are made of evaluation concepts in business and industry.

Uses of Evaluation Are Generally Applicable. As should be obvious by now, uses of evaluation are clearly portable, if one wishes to use evaluation in the same way in another arena. 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 may 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.

A Word about the Objects of Formal Evaluation Studies. As is evident from the previous discussion, formal evaluation studies have been conducted to answer questions about a wide variety of entities, which we have referred to as evaluation objects. The evaluation object is whatever is being evaluated. Like many disciplines, evaluation has developed its own technical terminology. For example, the word *evaluand* is sometimes used to refer to the evaluation object, unless it is a person, who is then an *evaluee* (Scriven, 1991a).

While we do not mind precise language, we see no need to use new terminology when familiar terms will do. Thus, except as they may appear in quoted material, we will not use *evaluand* or *evaluee* further, preferring to refer to both as *objects* of the evaluation.

In some instances, so many evaluations are conducted of the same type of evaluation object that it prompts suggestions for evaluation techniques found to be particularly helpful in evaluating something of that particular type. An example would be Kirkpatrick's (1983) 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 judgments, 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 *a 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 based on research showing adolescents' biological clocks cause them to be more groggy in the early morning hours and 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 calendar and visit other schools using similar calendars to provide information for planning the schedule. The planning staff will give the information to the Late Schedule Advisory Group, which will make final recommendations for the new schedule.

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 useful. They plan to use the results to revise this portion of the training program. A formative evaluation might observe 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 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. More recently, 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, 1991, p. 19). In the examples below 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 the 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 teacher turnover, satisfaction, and performance.

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, 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 health officials in agencies that could adopt the program), funding sources (taxpayers or a funding agency), and supervisors and other officials, as well as program personnel. The audiences for summative evaluations are often policy makers 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, 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 state departments of education on whether schools achieve certain standards. 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" test 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.

Failure to use formative evaluation is myopic, for formative data collected early can help rechannel time, money, and all types of human and material resources into more productive directions. Evaluation conducted only when a project nears completion may simply come too late to be of much help. Apparently, many instructional designers and trainers understand this point. Zemke (1985) surveyed readers of *Training* magazine and found that over 60 percent reported that they used formative evaluation in their training activities. In a later survey of corporate training officials, Tessmer and Wedman (1992) found that nearly half of their respondents reported that they use formative evaluation.

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 overall purpose of evaluation, determining merit and worth. While organizations may

engage in more summative evaluations, Scriven (1996) has noted that professional evaluators are more frequently involved in the formative role and often obtain more satisfaction from it. As a result, he has often come to the defense of summative evaluations for purposes of balance.

Although formative evaluations more often occur in early stages of a program's development and summative evaluations more often occur in their later stages, as these two terms imply, 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.

Two important factors that influence the usefulness of formative evaluation are control and timing. If suggestions for improvement are to be implemented, then it is important that the formative study collect data on variables over which program administrators have some control. Also, information that reaches those administrators too late for use in improving the program is patently useless. Summative evaluations must attend to the timing of budgetary and legislative decisions that may affect program adoption, continuation, and expansion.

An effort to distinguish between formative and summative evaluation on several dimensions appears in Figure 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. Scriven (1986) himself

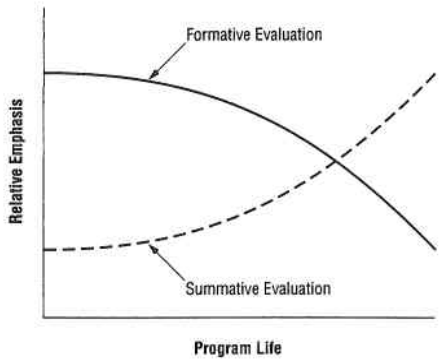

FIGURE 1.1 ***Relationship between Formative and Summative Evaluation across Life of a Program***

FIGURE 1.2 *Differences between Formative and Summative Evaluation*

	Formative Evaluation	*Summative Evaluation*
Purpose	To determine value or quality	To determine value or quality
Use	To improve the program	To make decisions about the program's future or adoption
Audience	Program managers and staff	Administrators, policy makers, and/or potential consumer or funding agency
By Whom	Primarily internal evaluators supported by external evaluators	Generally external evaluators, supported by internal evaluators in unique cases
Major Characteristics	Provides feedback so program personnel can improve it	Provides information to enable decision makers to decide whether to continue it, or consumers to adopt it
Design Constraints	What information is needed? When?	What evidence is needed for major decisions?
Purpose of Data Collection	Diagnostic	Judgmental
Frequency of Data Collection	Frequent	Infrequent
Sample Size	Often small	Usually large
Questions Asked	What is working? What needs to be improved? How can it be improved?	What results occur? With whom? Under what conditions? With what training? At what cost?

suggested one reason why they sometimes blur, noting that, when programs have many components, summative evaluations that result in replacing weak components have played a formative role in improving the program in its entirety.

Needs Assessment, Process, and Outcome Evaluations

The distinctions between formative and summative evaluation are concerned primarily with the kinds of decisions or judgments to be made with the evaluation results. The distinction between the relative emphasis on formative or summative evaluation is an important one to make at the beginning of a study because it informs the evaluator about the context, intention, and potential use of the study and has implications for the most appropriate audiences for the study. However, the terms do not dictate the nature of the questions the study will address. Chen

(1996) has proposed a typology to permit consideration of process and outcome along with the formative and summative dimension. We will elaborate that typology here, adding needs assessment to the mix.

Some evaluators make use of the terms *needs assessment, process,* and *outcome* to refer to the types of questions the evaluation study will address or the focus of the evaluation. These terms also help make the reader aware of the full array of issues evaluators examine. **Needs assessment** questions are concerned with establishing (a) whether a problem or need exists and describing that problem, and (b) making recommendations for ways to reduce the problem, i.e., the potential effectiveness of various interventions. **Process,** or **monitoring** studies, typically describe how the program is delivered. Such studies may focus on whether the program is being delivered according to some delineated plan or model or may be more open-ended, simply describing the nature of delivery and the successes and problems encountered. Process studies can examine a variety of different issues including characteristics of the clients or students served, qualifications of the deliverers of the program, characteristics of the delivery environment (equipment, printed materials, physical plant, and other elements of the context of delivery), and the actual nature of the activities themselves. **Outcome** studies are concerned with describing, exploring, or determining changes that occur in program recipients, secondary audiences (families of recipients, coworkers, etc.), or communities as a result of a program. These outcomes can range from immediate impacts (for example, satisfaction of learners) to final goals and unintended outcomes.

Note these terms do *not* have implications for how the information will be used. The terms *formative* and *summative* help us distinguish the purposes of the evaluation. Needs assessment, process, and outcome evaluations refer to the nature of the issues or questions that will be examined. In the past, people have occasionally misused the terms *"formative"* to be synonymous with *"process evaluation"* and *"summative"* to be synonymous with *"outcome evaluation."* However, Scriven (1996) himself notes that "formative evaluations are not a species of process evaluation. . . . Conversely, summative evaluation may be largely or entirely process evaluation" (p. 152).

Figure 1.3 illustrates a typology of evaluation terms building on the typology proposed by Chen (1996); we add needs assessment to Chen's typology and label this dimension "evaluation focus." (Chen views this dimension as reflecting the stage of the program, but, while process studies typically precede outcome studies, the choice of focus depends, not on the stage of the program, but on the information needs of the stakeholders.) As Figure 1.3 illustrates, an evaluation can be characterized by the action the evaluation will serve (improvement or otherwise) as well as by the nature of the issues it will address. To illustrate, a needs assessment study can be summative (Should we adopt this new program or not?) or formative (How should we modify this program to deliver it in our school or agency?). A process study often serves formative purposes, providing information to program providers or managers about how to change activities to improve the quality of the program, but a process study may serve summative purposes when we find that the program is too complex or expensive to deliver or that program

FIGURE 1.3 *A Typology of Evaluation Studies*

Focus of Questions	*Judgment*	
	What to revise/change Formative	*What to begin, continue, expand Summative*
Needs Assessment	How should we adapt the model we are considering?	Should we begin a program? Is there sufficient need?
Process	Is more training of staff needed to deliver the program appropriately?	Are sufficient numbers of the target audience participating in the program to merit continuation?
Outcome	How can we revise our curricula to better achieve desired outcomes?	Is this program achieving its goals to a sufficient degree that its funding should be continued?

recipients (students, trainees, clients) do not enroll as expected. In such cases, a process study that began as a formative evaluation for program improvement may lead to a summative decision to discontinue the program. Accountability studies often make use of process data to make summative decisions. An outcome study can, and often does, serve formative or summative purposes. Formative purposes may be best served by examining more immediate outcomes because program deliverers have greater control over the actions leading to these outcomes. For example, describing whether students are achieving immediate learning objectives is more useful to teachers in deciding how to revise their curricula than examining students' subsequent employment records or post-secondary performance. Policy makers making summative decisions, however, are often more concerned with the program's success at achieving "final" outcomes, e.g., employment, health, safety, because their responsibility is with these outcomes. Their decisions regarding funding concern whether programs achieve these ultimate outcomes. The fact that a study examines program outcomes, or effects, however, tells us nothing about whether the study serves formative or summative purposes.

The formative and summative distinction comes first, then, to help focus our attention on the judgment to be made or the action to be taken. In beginning an evaluation, evaluators are first concerned with determining this focus and, then, determining the extent to which the stakeholder can assist in making such judgments. (For example, a school board or state legislator is generally an inappropriate audience for a formative evaluation because they are typically too removed from immediate program activity. If the intention of the evaluation is formative and these are the primary audience, the evaluator should suggest that he work more closely with those involved in the day-to-day delivery of the program.) Only after the focus is determined will the evaluator proceed to examining whether the focus of the evaluation is needs assessment, process, or outcome and to developing the particular evaluation questions the study will address.

Internal and External Evaluations

The adjectives *internal* and *external* distinguish between evaluations conducted by program employees and those conducted by outsiders. An experimental year-round education program in the San Francisco public schools might be evaluated by a member of the school district staff (internal) or by a site-visit team appointed by the California State Board of Education (external). A large health maintenance organization (HMO) with facilities in six cities may have a member of each facility's staff evaluate the utility of their training of local residents to serve in paraprofessional roles (internal), or the HMO may hire a consulting firm or university research group to look at that paraprofessional training program (external).

Seems pretty simple, right? Often it is, but assume that the HMO sends a team out from their headquarters to evaluate the program in the six separate facilities. Is that an internal or external evaluation? Actually, the correct answer is "both," for such an evaluation is clearly external from the perspective of those in the individual facility, yet it clearly is an internal evaluation from the perspective of the headquarters administrators who assigned their staff to evaluate those parts of the parent HMO operation.

There are obvious advantages and disadvantages connected with both internal and external evaluation roles. Figure 1.4 summarizes some of these. Internal evaluators are likely to know more about the program, its history, its staff, its clients, and its struggles than any outsider. They also know more about the organization and its culture and styles of decision-making. They are familiar with the kinds of information and arguments that are persuasive, and know who is likely to take action and who is likely to be persuasive to others. These very advantages, however, are also disadvantages. They may be so close to the program that they cannot see it clearly. (Note, though, that each evaluator, internal and external, will bring his or her own history and "biases" to the evaluation, but the internal evaluators' closeness may prevent them from seeing solutions or changes that those newer to the situation might see more readily.) While successful internal

FIGURE 1.4 *Advantages of Internal and External Evaluators*

Internal	*External*
More familiar with organization & program history	Can bring greater credibility, perceived objectivity
Knows decision-making style of organization	Typically brings more breadth and depth of technical expertise
Is present to remind others of results now and in future	Has knowledge of how other similar organizations and programs work
Can communicate technical results more frequently and clearly	

evaluators may overcome the hurdle of perspective, it can be much more difficult for them to overcome the barrier of position. If internal evaluators are not provided with sufficient decision-making power, autonomy, and protection, their evaluation will be hindered.

The strengths of external evaluators lie in their distance from the program and, if the right evaluators are hired, their expertise. External evaluators are perceived as more credible by the public and, often, by policy makers. In fact, external evaluators typically do have greater administrative and financial independence. Nevertheless, the "objectivity" of the external evaluator can be overdone. (Note the role of the external Arthur Andersen firm in the 2002 Enron bankruptcy and scandal. The lure of obtaining or keeping a large contract can prompt external parties to "bend the rules" to keep the contract.) For programs with high visibility or cost or those surrounded by much controversy, an external evaluator can provide a needed degree of autonomy. External evaluators, if the search and hiring process are conducted appropriately, can also bring the specialized skills needed for a particular project. In all but very large organizations, internal evaluators must be "jacks-of-all-trades" to permit them to address the ongoing evaluation needs of the organization. When seeking an external evaluator, however, an organization can pinpoint and seek the types of skills and expertise needed for that time-limited project.

Possible Role Combinations

The dimensions of formative and summative evaluation can be combined with the dimensions of internal and external evaluation to form the two-by-two matrix shown in Figure 1.5. The most common roles in evaluation might be indicated by cells 1 and 4 in the matrix. Formative evaluations are often conducted by internal evaluators, and there are clear merits in such an approach. Their knowledge of the program, its history, staff, and clients is of great value, and credibility is not nearly the problem it would be in a summative evaluation. Program personnel are often the primary audience, and the evaluator's ongoing relationship with them can enhance the use of results in a good learning organization. Summative evaluations are probably most often (and probably best) conducted

	Internal	External
Formative	1 Internal Formative	2 External Formative
Summative	3 Internal Summative	4 External Summative

FIGURE 1.5 ***Combination of Evaluation Roles***

by external evaluators. It is difficult, for example, to know how much credibility to accord a Ford Motor Company evaluation that concludes that a particular Ford automobile is far better than its competitors in the same price range. The credibility accorded to an internal summative program evaluation (cell 3) may be no better. In most organizations, summative evaluation is generally best conducted by an external evaluator or agency, but there are two circumstances in which we would alter that statement quite dramatically. First, in some instances, there is simply no possibility of the program's obtaining such external help because of financial constraints or absence of competent personnel willing to do the job. In these cases, the summative evaluation is weakened by the lack of outside perspective, but it might be possible to retain a semblance of objectivity and credibility by choosing the internal summative evaluator from among those who are some distance removed from the actual development of the program or product being evaluated.

For example, assume that an elementary school in a large (in geography, not budget) rural district in Saskatchewan needs to have a summative evaluation of an innovative French language and culture program they have been running. No funds are available to bring an evaluator in from outside the district, and, because much of the program is oral, it would be hard to bundle it up and send it off for review. Everyone in the school is either a zealous enthusiast or a bitter opponent of the program, so there is no way to get an unbiased internal evaluation. In this context, it is far better to obtain a "quasi-external" summative evaluation than do none at all. By "quasi-external," we mean that one should conduct the evaluation so as to maximize its "externality." Why not ask a qualified staff member of another school in the district to evaluate the program in return for helping with a later task in that school? While still internal to the district, this evaluator would be external to the school, hence quasi-external. If the evaluation were commissioned with a strong request for the quasi-outsider to "tell-it-like-it-is," with no punches pulled and no weaknesses overlooked, there is good reason to suspect that many of the advantages of a true external summative evaluation would occur. If one still worried that being in the same district tainted the outcomes, perhaps someone from an adjacent district, or a school not too far beyond the province's boundary, would make it a true external evaluation. Whatever definitional cutoffs one chooses to use, it is important to remember that there is a continuum from external to internal; it is a matter of degree, not black or white.

The second circumstance when we might soften our cautions about the biases that can occur in internal evaluations is when organizations have structured their internal evaluation unit (and its evaluators) to enhance their ability to be forthright about their findings. Such structuring can take many forms, but the key is that the internal evaluators are insulated and shielded from the consequences of displeasure of those whose program is evaluated.

Fortunately, a number of large agencies have structured their internal evaluation function to give it maximum independence (and avoid evaluators being placed in the untenable posture of evaluating programs developed by the

boss or close associates). The larger the organization, the more insulated its evaluation staff can be and the fewer problems or pressures one might expect to be caused by hierarchical or close social relationships. Indeed, the unit (and its function) may even lose much of its internal flavor and appear more like a built-in external evaluation unit (if that non sequitur is permitted), free to pursue evaluations throughout the organization as need demands. Sonnichsen (2000) writes of the high impact that internal evaluation can have if the organization has established the conditions that permit the internal evaluator to operate effectively. The factors that he cites as being associated with evaluation offices that have a strong impact on the organization include operating as an independent entity, reporting to a top official, giving high rank to the head of the office, having the authority to self-initiate evaluations, making recommendations and monitoring their implementation, and disseminating results widely throughout the organization. He envisions the promise of internal evaluation, writing, "The practice of internal evaluation can serve as the basis for organizational learning, detecting and solving problems, acting as a self-correcting mechanism by stimulating debate and reflection among organizational actors, and seeking alternative solutions to persistent problems" (Sonnichsen, 2000, p. 78).

Evaluation's Importance—and Its Limitations

Given its many formative and summative uses, it may seem almost axiomatic to assert that evaluation is not only valuable but essential in any effective system or society. Scriven (1991b) has said it well:

> The process of disciplined evaluation permeates all areas of thought and practice. . . . It is found in scholarly book reviews, in engineering's quality control procedures, in the Socratic dialogues, in serious social and moral criticism, in mathematics, and in the opinions handed down by appellate courts. . . . It is the process whose duty is the systematic and objective determination of merit, worth, or value. Without such a process, there is no way to distinguish the worthwhile from the worthless (p. 4).

Scriven also argues the importance of evaluation in *pragmatic* terms ("bad products and services cost lives and health, destroy the quality of life, and waste the resources of those who cannot afford waste"), *ethical* terms ("evaluation is a key tool in the service of justice"), *social* and *business* terms ("evaluation directs effort where it is most needed, and endorses the 'new and better way' when it is better than the traditional way—and the traditional way where it's better than the new high-tech way"), *intellectual* terms ("it refines the tools of thought"), and *personal* terms ("it provides the only basis for justifiable self-esteem") (p. 43). Perhaps for these reasons, evaluation has increasingly been used as an instrument to pursue goals of organizations and agencies at local, regional, national, and international levels.

Potential Limitations of Evaluation

The usefulness of evaluation has led some people to look to it as a panacea for all the ills of society, but evaluation alone cannot solve all the problems of society. One of the biggest mistakes of evaluators is to promise results that cannot possibly be attained. Even ardent supporters of evaluation are forced to admit that many evaluation studies fail to lead to significant improvements in the programs they evaluate. Why? Partly it's a question of grave inadequacies in the conceptualization and conduct of many evaluations. It's also a question of understanding too little about other factors that affect the use of evaluation information, even from studies that are well conceptualized and well conducted. In addition, both evaluators and their clients may have been limited by an unfortunate tendency to view evaluation as a series of discrete studies rather than a continuing system of self-renewal. A few poorly planned, badly executed, or inappropriately ignored evaluations should not surprise us; such failings occur in every field of human endeavor. This book is intended to help evaluators, and those who use their results, to improve the practice and utility of evaluation.

A parallel problem exists when those served by evaluation naively assume that its magic wand need only be waved over an enterprise to correct all its malfunctions and inadequacies. For example, developing and measuring standards in education or in nonprofit agencies, as is the current trend, can certainly provide useful information for judging the quality of programs, but these performance monitoring programs are only the first step. Formative evaluations, specific to the context of the program, are then needed to bring about improvement. Though evaluation can be enormously useful, it is generally counterproductive for evaluators or those who depend on their work to propose evaluation as the ultimate solution to every problem or, indeed, as any sort of solution, because evaluation, in and of itself, won't effect a solution, though it might suggest one. Evaluation serves to identify strengths and weaknesses, highlight the good, and expose the faulty, but it cannot singlehandedly correct problems, for that is the role of management and other stakeholders, using evaluation findings as one tool that will help them in that process. Evaluation has a role to play in enlightening its consumers, and it may be used for many other roles, but it is only one of many influences on improving the policies, practices, and decisions in the institutions that are important to us.

Major Concepts and Theories

1. Evaluation is the identification, clarification, and application of defensible criteria to determine an evaluation object's value, its merit or worth, in regard to those criteria. The specification and use of explicit criteria distinguish formal evaluation from the informal evaluations most of us make daily.

2. Evaluation differs from research in its purpose, its concern with generalizability, its involvement of stakeholders, and the breadth of training those practicing it require.

3. The basic purpose of evaluation is to render judgments about the value of the object under evaluation. Other purposes include providing information for program improvement, working to better society, encouraging meaningful dialogue among many diverse stakeholders, and providing oversight and compliance for programs.

4. Evaluators play many roles including scientific expert, facilitator, planner, collaborator, aid to decision makers and critical friend.

5. Evaluations can be formative or summative. Formative evaluations are designed for program improvement and the audience is, most typically, stakeholders close to the program. Summative evaluations serve decisions about program adoption, continuation, or expansion. Audiences for these evaluations must have the ability to make such "go–no go" decisions.

6. Evaluators may be internal or external to the organization. Internal evaluators know the organizational environment and can facilitate communication and use of results. External evaluators can provide more credibility in high-profile evaluations and bring a fresh perspective and different skills to the evaluation.

Discussion Questions

1. Consider a program in your organization. If it were to be evaluated, what might be the purpose of the evaluation? The goal? The role of evaluators in conducting the evaluation?
2. What kind of evaluation do you think is most useful, formative or summative? What kind of evaluation would be most useful to you in your work? To your school board or elected officials?
3. Which do you prefer, an external or internal evaluator? Why?
4. Describe a situation in which an internal evaluator would be more appropriate than an external evaluator. What is the rationale for your choice? Now describe a situation in which an external evaluator would be more appropriate.

Application Exercises

1. List the types of evaluation studies that have been conducted in an institution or agency of your acquaintance, noting in each instance whether the evaluator was internal or external to that institution. Determine whether each study was formative or summative and focused on needs assessment, process, or outcome questions. Finally, consider whether the study would have been strengthened by having someone with the opposite (internal/external) relationship to the institution conduct the study.
2. Think back to any formal evaluation study you have seen conducted (or if you have never seen one conducted, find a written evaluation report of one). Identify three things that make it different from informal evaluations. Then list ten informal evaluations you have performed so far today. (Oh, yes you have!)
3. Discuss the potential and limitations of program evaluation. Identify some things evaluation can and cannot do for programs in your field.
4. Within your own organization (if you are a university student, you might choose your university), identify several evaluation objects that you believe would be appropriate for study. For each, identify (a) the use the evaluation study would serve, and (b) the basic focus of the evaluation.

Suggested Readings

Mark, M. M., Henry, G. T., & Julnes, G. (1999). Toward an integrative framework for evaluation practice. *American Journal of Evaluation, 20,* 177–198.

Rallis, S. F., & Rossman, G. B. (2000). Dialogue for learning: Evaluator as critical friend. In R. K. Hopson (Ed.), *How and why language matters in evaluation.* New Directions for Evaluation, No. 86, 81–92. San Francisco: Jossey-Bass.

Sonnichsen, R. C. (2000). *High impact internal evaluation.* Thousand Oaks, CA: Sage.

Stake, R. E. (2000). A modest commitment to the promotion of democracy. In K. E. Ryan & L. DeStefano (Eds.), *Evaluation as a democratic process: Promoting inclusion, dialogue, and deliberation.* New Directions for Evaluation, No. 85, 97–106. San Francisco: Jossey-Bass.

Stufflebeam, D. L. (2001). *Evaluation models.* New Directions for Evaluation, No. 89. San Francisco: Jossey-Bass.