

CHAPTER

1

FOUNDATIONS OF EVALUATION

This chapter defines and describes evaluation and sets the frame for this book within the principles of performance improvement. Various kinds of evaluation, as well as some closely related processes, are differentiated from each other. The basic challenges that evaluators face are laid out, and the reason that stakeholder commitment is so important is examined. The benefits of evaluation to an organization are listed. Finally, definitions are provided for some key terms used throughout the book and in the evaluation field.

In our daily lives, we encounter decision points on an almost continuous basis: Should I do this, or should I do that? Should I go right or left? Should I take the highway or the back streets? Should I buy now or later? Should I take my umbrella today or not? Life in an organizational setting is no different: We face decisions about which programs to

4 Performance Evaluation

sustain, which to change, and which to abandon, to name but a few organizational dilemmas. How do members of an organization go about making sound decisions? With the use of relevant, reliable, and valid data, gathered through a sound evaluation process aligned with desired long-term outcomes.

Unfortunately, these data are not always available, and if they are, many decision makers do not know they exist, or do not have access to them, or do not know how to interpret and use them to make sound decisions that lead to improved program and organizational performance. In fact, Lee Cronbach (1980) and others have argued that decisions often emerge rather than being logically and methodically made.

Effective leaders are capable of making sound decisions based on sound data, and evaluators can do much to influence the leadership decision-making process. Evaluation can provide a systematic framework that aligns stakeholders, evaluation purposes, desired results and consequences, and all evaluation activities, so that the evaluation product is a responsive and clear recipe for improving performance. This in essence allows the decision-making process to become clearer and more straightforward. Evaluation is the mechanism that provides decision makers with feedback, whether through interim reports and meetings or a final report and debriefing.

A BRIEF OVERVIEW OF EVALUATION HISTORY

Michael Scriven (1991) describes evaluation as a practice that dates back to samurai sword evaluation. Another type of evaluation was in evidence as early as 2000 B.C.: Chinese officials held civil service examinations to measure the ability of individuals applying for government positions. And Socrates included verbal evaluations as part of his instructional approach (Fitzpatrick, Sanders, & Worthen, 2004).

In response to dissatisfaction with educational and social programs, a more formal educational evaluation can be traced back to Great Britain during the 1800s, when royal commissions were sent by the government to hear testimony from the various institutions. In the 1930s, Ralph Tyler issued a call to measure goal attainment with standardized criteria (Fitzpatrick et al., 2004). And during the 1960s, Scriven and Cronbach introduced formative (used to guide developmental activities) and summative (used to determine the overall value of a program or solution) evaluation, and Stufflebeam stressed outcomes (program results) over process (program activities and resources) (Liston, 1999).

In 1963, Cronbach published an important work, “Course Improvement Through Evaluation,” challenging educators to measure real learning rather than the passive mastery of facts. Moreover, he proposed the use of qualitative instruments, such as interviews and observations, to study outcomes. In the latter part of the 1960s, well-known evaluation figures such as Edward Suchman, Michael Scriven, Carol Weiss, Blaine Worthen, and James Sanders wrote the earliest texts on program evaluation.

In 1971, Daniel Stufflebeam proposed the CIPP model of evaluation, which he said would be more responsive to the needs of decision makers than earlier approaches to evaluation were. In that same year, Malcolm Provus proposed the discrepancy model of evaluation. In 1972, Scriven proposed goal-free evaluation in an effort to encourage evaluators to find unintended consequences. In 1975, Robert Stake provided responsive evaluation. In 1981, Egon Guba and Yvonna Lincoln proposed naturalistic evaluation on the basis of Stake’s work, feeding the debate between qualitative and quantitative methods (Fitzpatrick et al., 2004).

All of this was occurring in the context of a movement to account for the billions of dollars the U.S. government was spending on social, health, and educational programs (Fitzpatrick et al., 2004; Patton, 1997). In order to address a demand for accountability, those responsible for programs soon began to ask evaluators for advice on program improvement. Thus, the initial purpose of program evaluation was to judge the worthiness of programs for continued funding.

When Sputnik became the catalyst for improving the U.S. position in education, which was lagging compared to other countries, educational entities in particular began to commission evaluations, partly in order to document their achievements. The need for evaluators soon grew, and government responded by funding university programs in educational research and evaluation. In the 1970s and 1980s, evaluation grew as a field, with its applications expanding beyond government and educational settings to management and other areas. Evaluations are now conducted in many different settings using a variety of perspectives and methods.

EVALUATION: PURPOSE AND DEFINITION

While some rightly say that the fundamental purpose of evaluation is the determination of the worth or merit of a program or solution (Scriven, 1967), the ultimate purpose, and value, of determining this worth is in

6 Performance Evaluation

providing the information for making data-driven decisions that lead to improved performance of programs and organizations (Guerra-López, 2007a). The notion that evaluation's most important purpose is not to prove but to improve was originally put forward by Egon Guba when he served on the Phi Delta Kappa National Study Committee on Evaluation around 1971 (Stufflebeam, 2003). This should be the foundation for all evaluation efforts, now and in the future. Every component of an evaluation must be aligned with the organization's objectives and expectations and the decisions that will have to be made as a result of the evaluation findings. These decisions are essentially concerned with how to improve performance at all levels of the organization: internal deliverables, organizational gains, and public impact. At its core, evaluation is a simple concept:

- It compares results with expectations.
- It finds drivers and barriers to expected performance.
- It produces action plans for improving the programs and solutions being evaluated so that expected performance is achieved or maintained and organizational objectives and contributions can be realized (Guerra-López, 2007a).

Some approaches to evaluation do not focus on predetermined results or objectives, but the approach taken in this book is based on the premise of performance improvement. The underlying assumption is that organizations, whether they fully articulate this or not, expect specific results and contributions from programs and other solutions. As discussed in later chapters, this does not prevent the evaluator or performance improvement professional from employing means to help identify unanticipated results and consequences. The worth or merit of programs and solutions is then determined by whether they delivered the desired results, whether these results are worth having in the first place, and whether the benefits of these results outweigh their costs and unintended consequences.

An evaluation that asks and answers the right questions can be used not only to determine results but also to understand those results and to modify the evaluation so that it can better meet the intended objectives within the required criteria. This is useful not only to identify what went wrong or what could be better but also to identify what should be maintained. Through appreciative inquiry (Cooperrider & Srivastva, 1987), evaluation can help organizations identify what is

going right. Appreciative inquiry is a process that searches for the best in organizations in order to find opportunities for performance improvement. Here too the efforts are but a means to the end of improving performance. Although the intentions of most evaluators are just that, the language and approach used are charged with assumptions that things are going wrong. For instance, the term *problem solving* implies from the start that something is wrong. Even if this assumption is not explicit in the general evaluation questions, it makes its way into data collection efforts. Naturally the parameters of what is asked will shape the information evaluators get back and, in turn, their findings and conclusions. If we ask what is wrong, the respondents will tell us. If we ask what went right, again they will tell us. The key point is that evaluation should be as unbiased as possible. Evaluators should ask and answer the right questions, so that the data they get are indeed representative of reality.

In specific terms, before evaluators start to plan, and certainly before they collect data, they must determine why they are conducting an evaluation. Is this their initiative, or were they directed to do this work? What is the motivation for the study? What are they looking to accomplish and contribute as a result of this evaluation? Here are some general reasons for conducting an evaluation:

- To see if a solution to a problem is working, that is, delivering valued ends
- To provide feedback as part of a continual monitoring, revision, and improvement process
- To provide feedback for future funding of initiatives
- To confirm compliance with a mandate
- To satisfy legal requirements
- To determine if value was added for all stakeholders
- To hold power over resources
- To justify decisions that have already been made

Although the last two in this list are particularly driven by political agendas, in reality most reasons can be politicized; thus, it takes an insightful evaluator to recognize the feasibility of conducting an honest evaluation. An experienced evaluator will recognize, most of the

8 Performance Evaluation

time, whether evaluation stakeholders are truly interested in using evaluation findings to improve performance or are more concerned with advancing their political interests. With careful attention to detailed planning, either goal can be made to fit a data-driven and results-oriented action approach to evaluation. But if taken too narrowly—in isolation and without proper context—each has its own narrow set of problems, blind spots, and special data generation and collection issues. Perception of the purpose of the evaluation can shape and limit the data that are observed (or not observed), collected (or not collected), and interpreted (or ignored). Thus, evaluators and stakeholders must begin the planning process with a clear articulation of what decisions must be made with the results of their findings, decisions that are linked to the overall purpose for conducting the evaluation.

PERFORMANCE IMPROVEMENT: A CONCEPTUAL FRAMEWORK

The field of performance improvement is one of continuous transition and development. It has evolved through the experience, reflection, and conceptualization of professional practitioners seeking to improve human performance in the workplace. Its immediate roots stem from instructional design and programmed instruction. Most fundamentally, it stems from B. F. Skinner and his colleagues, whose work centered on the behavior of individuals and their environment (Pershing, 2006).

The outgrowth of performance improvement (also called *human performance technology*) from programmed instruction and instructional systems design was illustrated in part by Thomas Gilbert's behavioral engineering model, which presented various categories of factors that bear on human performance: clear performance expectations, feedback, incentives, instruments, knowledge, capabilities, and internal motives, for example. This landmark model was published in Gilbert's 1978 book, *Human Competence: Engineering Worthy Performance*, and was based in large part on the work Gilbert conducted with Geary Rummler and Dale Brethower at the time. Pershing (2006) declares that Joe Harless's 1970 book, *An Ounce of Analysis Is Worth a Pound of Objectives*, also had a significant impact on the field and was well complemented by Gilbert's work. Together these works served as the basis for many researchers who have contributed to and continue to help develop the performance improvement field.

Currently the International Society for Performance Improvement, the leading professional association in the field, defines *performance improvement* as a systematic approach to improving productivity and competence, using a set of methods and procedures—and a strategy for solving problems—for realizing opportunities related to the performance of people. More specifically, it is a process of selection, analysis, design, development, implementation, and evaluation of programs to most cost-effectively influence human behavior and accomplishment. This series of steps, commonly known as the ADDIE model, is the basic model from which many proposed performance improvement evaluation models stem. Pershing (2006) summarized performance improvement as a systematic combination of three fundamental processes: performance analysis (or needs assessment), cause analysis (the process that identifies the root causes of gaps in performance), and intervention selection (selecting appropriate solutions based on the root causes of the performance gaps). These three processes can be applied to individuals, small groups, and large organizations. The proposition that evaluation of such interventions should also be at the core of these fundamental processes is presented in the final chapter of this book.

This is the context in which evaluation is seen and described in this book—not as an isolated process but rather as one of a series of processes and procedures that, when well aligned, can ensure that programs and organizations efficiently and effectively deliver valuable results.

MAKING EVALUATION HAPPEN: ENSURING STAKEHOLDERS' BUY-IN

One of the most important elements of any evaluation is its stakeholders. Before we define the stakeholders, it is worthwhile to define the term *stake*. A *stake* is essentially a claim, an interest, or a share in some endeavor and how that claim or interest might be affected by anything that is used, done, produced, or delivered. The traditional view of a stake used to be limited to the financial realm (for example, stockholders), but in fact a claim or interest can be financial, legal, or moral (Carroll, 2000). Thus, a stakeholder is any individual or group with a stake in an endeavor and can either affect or be affected by the decisions and actions of the organization.

Stakeholders can be broadly categorized as internal (owners, employees, and management) and external (customers, customers' customers, the community, suppliers, competitors, the government, and the

media, to name a few), and both categories can then be subdivided into various groups.

Not every individual within each stakeholder group has to participate directly in an evaluation; what is important is that those who participate are seen as representative by their group members. The greater the sense of stakeholder involvement and influence there is, the less likely it is that the evaluator will encounter resistance to the evaluation process and the findings.

While ideally evaluators will select stakeholders who will help define useful evaluation expectations, questions, and criteria, in fact, they realistically will be faced with stakeholders who have their own special interests or represent a powerful lobby. Although it is not particularly unusual for human beings to have their own special interests, evaluators should neutralize as much as possible the risk that the evaluation will become a manipulation tool for the special interests of one—or some—at the expense of others.

A vital challenge in working with stakeholders to help all be successful is to keep them focused on results and consequences rather than on politics of means. Single-issue politics from both within and outside organizations have a tremendous impact on defining objectives and selecting means. It is essential that evaluators learn enough about the specific political climate of a given evaluation to understand how it will affect the evaluation and the implementation of its recommendations. If evaluation recommendations are not implemented or are implemented improperly, performance probably will not improve, and the evaluation may have been conducted in vain.

THE EVALUATOR: A JOB OR A ROLE?

The term *evaluator* describes not only one profession or occupation, but also a given role at a particular time. Individuals conducting evaluation often wear many hats. They may be internal employees, members of the management team, faculty members, or consultants who have acquired interest and expertise in measurement and evaluation through education, training, or experience. In some cases, individuals arrive at this point by default and face an unexpected request to conduct an evaluation. They could be trainers who are charged with demonstrating the value of their training programs and departments. They may even be individuals who because of their status as a subject matter expert in some solution or program are also faced with demonstrating the value of their efforts.

Their common function is, or should be, an interim goal to document the results and impact achieved by a given solution: a program, a project, a tool, or the use of a resource. The final goal should be to use this information to make sound decisions and help the organization take appropriate action to improve performance at all levels.

Evaluators should be competent in some basic areas. Sanders (1979) proposed that at a minimum, evaluators should be able to

- Accurately describe the object (the evaluand) and context of that which is being evaluated
- Conceptualize the purpose and framework of the evaluation
- Derive useful evaluation questions, data requirements, and appropriate data sources
- Select the means for collecting and analyzing data
- Determine the value of the evaluand
- Effectively communicate results and recommendations to the audience
- Manage the evaluation project
- Maintain ethical standards
- Adjust to external factors influencing the evaluation
- Evaluate the evaluation

THE RELATIONSHIP TO OTHER INVESTIGATIVE PROCESSES

The results and consequences we want to accomplish are the primary drivers for deriving the useful questions of an organizational study. Another driver is the types of decisions that have to be made; in large part, they will determine what data have to be gathered and for what purpose. For instance, if decisions have to be made about what programs, interventions, and solutions should be continued, revised, or discontinued, then the data collection approach may take an evaluative perspective. That is, the data collected will be used to compare predetermined objectives with what was actually achieved. If the need is to make decisions about what results the organization should be targeting and, in turn, what types of programs, interventions, and solutions will help it get there, the data collection approach will take on a needs assessment

12 Performance Evaluation

perspective. Notice that in both cases, results—and gaps in results—are the primary drivers.

Table 1.1 illustrates some sample questions from both perspectives that could apply to any organization in any sector. Both approaches to data collection should be systematic and designed to answer specific questions that can be used to improve performance.

Assessors and evaluators may share data collection techniques, but the types of questions they seek to answer differ. In this sense, the roles of assessor and evaluator differ in purpose or function rather than in importance and methods.

Needs assessors help create the future by providing hard and soft data for identification of performance-based, vision-aligned missions

TABLE 1.1. Unique Perspectives of Needs Assessment and Evaluation

Needs Assessment Questions	Evaluation Questions
What value-added results should we be targeting?	How much closer did we get to reaching our vision and mission?
What value-added results are we now getting?	Did we add to or subtract value from our external clients and our shared society?
Who or what is the primary client of the results and their consequences?	Which objectives in our mission did we achieve?
How do we get from current results and consequences to desired ones?	How are we doing in comparison to last quarter? Last year?
What interim results must be accomplished and when?	Which internal results targets were reached? Not reached?
What are our options?	Which implemented programs, projects, or solutions were effective?

and building-block objectives, as well as the gaps between current and desired results. In addition, they help identify the best solutions for closing these gaps and thereby ultimately reaching the organizational vision. It should be noted that asking people what they need is not a needs assessment; this simply creates a “wants list” or “wish list” without rigorous applicability (Kaufman, 2000). *Evaluators* help to determine whether they are heading toward reaching the future they set out to create during the needs assessment process. One of the primary ways they do this is by determining the effectiveness and efficiency of the implemented programs and solutions, as well as the causal factors associated with any gaps between expected and accomplished results. Measurably improving organizational and individual performance depends heavily on these two roles and processes.

Needs Assessment Questions	Evaluation Questions
What are the most effective and efficient ways for reaching our desired or required results?	How efficient are these implemented programs, projects, or solutions?
What will it cost us to reach those results?	In which of these should we continue to invest?
What will it cost us to ignore those results?	What results do we have to justify our continued programs?
How far do we have to go to reach those results?	What should we discontinue?
Which results take priority over others?	Which projects, programs, or solutions could be successful with some modifications? Is it worth it?
Where do we have the most—and least—leverage?	Did we add or subtract value from our internal clients and employees?

Source: Guerra (2003b).

Although both assessors and evaluators collect data with regard to the results of a process or activity, evaluators collect data to determine whether results match the results expected from solutions that have already been implemented—for example, new programs, new technologies, new processes, or any other means selected to help achieve objectives. Assessors, in contrast, seek to anticipate the expected return on investment of potential interventions before they are implemented by collecting data about both current results (what is) and potential results (what should be). With these data in hand, decision makers are able to choose among competing alternatives.

So how does scientific research come into the picture? Before answering this question, let us first explore the meaning of *science*. Science is based on a series of assumptions about the world—assumptions that can be true today but false tomorrow. Scientists are always testing these assumptions, ready to change them when the findings support such a change. To this end, scientists collect data about reality and consult with other sources to ensure the reliability of the data. Results are considered basic data, later subject to repeatable observations in order to confirm findings and scientific reports. Thus, we want to make decisions and take action based on what is currently known through scientific inquiry.

Research is essentially another systematic process of inquiry, with the purpose of finding, interpreting, and updating facts, events, behavior, and theories. In this sense, research skills are a basic requirement in today's world and can be applied in just about any context, whether needs assessment, evaluation, or scientific inquiry. In fact, the heart of the data collection plan is very much the same for all of these. Following are the common elements among these three inquiry processes. These are stated generically but can be made specific to investigative contexts.

1. Important decisions that must be taken by stakeholders are identified. They lead to element 2:
2. Guiding questions, purposes, or hypotheses that the inquiry process must answer or test, which are related to element 3:
3. Key variables or results that are the central focus of the questions or hypotheses.
4. When results are not directly observable, measurable and observable indicators must be identified.

5. These indicators become the data to collect and inform the data source used.
6. The types of data sought inform the data collection tools appropriate for use.
7. The types of data sought determine the types of data analysis tools qualified to summarize and test these data.
8. The process concludes with findings, interpretations, and reporting that are supported by the data collected.

The key in the methodology is the alignment of all the elements: from adding value to all internal and external stakeholders to linking with resources and methods to deliver worthy results.

Certainly in much basic research, generalizability of findings is critical, and thus there is a strong push for controlled environments and the isolation of effects. However, the complexity of real-world evaluation does not easily lend itself to the control of variables. Perhaps evaluation overlaps more closely with applied research, where the goal of the study is the solution of real organizational problems rather than the advancement of the theoretical body of knowledge. However, both evaluation and applied research benefit from the knowledge obtained through basic research. Table 1.2 provides a side-by-side comparison of basic research, applied research, and evaluation. Although the dimensions of each process are described generally, modification of any of them is possible, thereby blurring these distinctions. For example, an evaluation report could be used as part of the literature review of a basic research study, thereby influencing what research questions are studied and how.

WHEN DOES EVALUATION OCCUR?

Having measurable performance objectives in the correct format does not guarantee that the objectives address the right things. Decades ago, people realized that focusing only on objectives could cause an evaluator to miss important data on process and environment. In the 1960s, the realization that evaluation could play a role in the development of educational programs to adjust content and process along the way to the final results gave rise to a famous distinction when Scriven (1967) introduced the terms *formative* and *summative* as well as *goal-free* evaluation. Since then, evaluators have had a term for the type of evaluation activity used to guide developmental activity in programs (*formative*)

TABLE 1.2. Dimensions of Investigative Processes

Dimensions	Basic Research	Applied Research	Evaluation
Goal	Advancement of knowledge and the theoretical understanding of relevant variables	Application of scientific knowledge to the solution of a specific, defined problem	Identification of relevant information to improve specific objects and organizations
Approach	Exploratory and often driven by the researcher's curiosity and interests	Generally descriptive rather than exploratory and conducted by educational or other institutions	Generally guided by the need to make important organizational decisions
Use	Conducted without a currently practical end in mind	Done to solve specific, practical questions	Done to solve specific, practical questions
Basis	As its name suggests, it could provide the basis for further, often more applied research	Often done on the basis of basic research or on previous valid research findings	Should be done on the basis for a needs assessment, while also considering past basic and applied research findings

and another term for evaluation that is used to comment on overall final value (*summative*).

While determining the overall value added—or potentially subtracted—by programs and organizations should be one of its key functions, formative evaluation is also quite important to the overall contributions of programs and other solutions. Moreover, formative evaluation can be designed in such a way that it continuously monitors the alignment of a program with its subsystems and suprasystems to facilitate the achievement of its ultimate value.

Formative evaluation should start at the same time as the identification, design, development, and implementation of the program or solution of interest. Some general questions include the following ones:

- Are we targeting the right objectives?
 - Are they based on assessed needs (gaps in results)?
- Are the criteria measurable and soundly based?
- Are we using the right criteria to judge the effectiveness and efficiency of our solution?
- Did we identify the appropriate program or solution?
 - Did we base our selection on an analysis of alternatives?
 - Did we weigh the pros and cons?
 - Did we weigh the costs and consequences?
- Is our design useful and relevant?
 - Is it aligned with the front-end analysis findings (causes for gaps in results)?
 - Is it appropriate for the ends we want to reach?
- Is the development of the program or solution aligned with its intended design?
 - Is our pilot designed to capture the right data required for improvements?
 - Does our prototype meet the requirements of our users?
- Is the program or solution being implemented appropriately?

18 Performance Evaluation

Incidentally, implementation questions may also be appropriate during summative evaluation approach, where we not only look at the results and consequences but also at the factors that may have led to those results and consequences. Obviously, if the intent is to ensure the effectiveness of the solution, we want to know if we are implementing it effectively before and during implementation, not just after the fact:

- Were those affected by the program or solution included in the problem identification, solution selection, and every other stage?
- Were fears and unfounded ideas about the implications of the program or solution confronted, clarified, or disproved, as appropriate?
- Is the program or solution being implemented according to initial plans?
- Is the implementation of the program or solution flexible and responsive to the current situation (for example, challenges not previously foreseen)?

Evaluating each stage and using evaluation data to improve it will allow evaluators and stakeholders to stay on track in order to reach the short- and long-term objectives of the program or solution.

GENERAL EVALUATION ORIENTATIONS

Two common distinctions in evaluation are formative and summative. *Formative evaluation* typically occurs during the developmental stage of a program and can be used to improve the program before it is formally launched. The formative approach can also be used to improve all stages of performance improvement, from assessment to implementation, and the evaluation itself.

Summative evaluation occurs after the implementation of a program or solution and usually requires some appropriate amount of time to have transpired so that the object of evaluation has the opportunity to have the full impact required on performance at various levels of the organization. It is worth noting that summative evaluation can also be used to improve programs and solutions. Stufflebeam and Webster (1980) hold that an objectives-based view of program evaluation is the most common type of evaluation. Once the results that have been

accomplished have been determined, the evaluator is well advised to identify causal factors contributing to those results. These data should provide insights as to what the drivers and barriers to the success of the program are, thereby providing the basis for recommendations for improving performance.

Another distinction often made among evaluation orientations is that of *process evaluation* versus *results evaluation*. These terms are used to describe the same processes that formative and summative approaches, respectively, take. Depending on how these are interpreted and implemented, they can also differ somewhat from their counterparts described above. For instance, the Canadian Evaluation Society uses the term *process evaluation* (also referred to as *efficiency evaluation*) to describe the monitoring of the implementation of programs. Obviously, there should be a well-planned logic model with specified results and processes, but modifications are made if a discrepancy between the program design and the actual implementation is found. For example, one might want to determine if the program is being delivered as intended, if it is being delivered to the targeted clients or participants, or if it is being delivered with the intended effort or in the intended quantity.

Process evaluation is critical in helping evaluators address the variations in program delivery. The greater the variation in program delivery, the greater the requirement is for useful data gathered through a process evaluation approach. For instance, there may be differences in staff, clients, environments, or time, to name a few variables.

Stufflebeam and Webster (1980) have argued that objectives-based program evaluation is the most prevalent type used in the name of educational evaluation. Scriven (1972) proposed goal-free evaluation to urge evaluators to also examine the process and context of the program in order to find unintended outcomes.

Results evaluation, also referred to as *effectiveness evaluation*, is used to determine whether the immediate outcomes of a program meet predetermined objectives specified by program planners; *impact evaluation* tends to refer to an evaluation that looks at not only immediate outcomes but also the long-term outcomes of a program and their interdependency. A results evaluation approach is important because it allows us to ensure and document that we are on track by gathering data that show quality accomplishments. It also helps us stay accountable and our programs to stay cost-effective by making program benefits and costs tangible.

Other evaluation approaches are associated with effectiveness evaluation. *Cost-benefit evaluation* is the translation of costs and benefits into monetary terms, which is used to compare the relative net benefits of doing one thing versus another. However, monetary terms are not always applicable, and they are seldom sufficient to appreciate costs and benefits. *Cost-effectiveness evaluation* considers alternative forms of program delivery according to both their costs and their effects with regard to producing some result or set of results. Of course, a stable measure of result should be defined. The least costly program is not necessarily the best one. And in the context of technology solutions, an additional orientation to evaluation is *usability testing*, which focuses on whether people are using the product and how well they are using it to meet required objectives.

CHALLENGES THAT EVALUATORS FACE

A common excuse for avoiding evaluation is insufficient resources to conduct one. In fact, it can often take more resources to maintain programs blindly and indefinitely than it does to conduct a rigorous and focused evaluation. The decision about whether to conduct an evaluation in the first place requires thinking about not only its cost but also the benefits it can render. Both cost and benefit categories contain monetary and nonmonetary items, and they should be honestly and carefully considered before making decisions about conducting or not conducting an evaluation.

One of the most serious challenges faced by evaluators—and probably researchers in general—is getting people to use the findings and recommendations. One study (Henderson, Davies, & Willis, 2006) cited lack of key stakeholder and consumer involvement as a factor that reflects the adoption of evidence for changes in practice. Lack of leadership support was also identified as a factor. When these two factors are combined, there is no support at all for creating and promoting change. The default stance is maintaining the status quo, even if “changing” is the logical proposition.

Limited expertise can also become a barrier. When no one, or few people, in the organization understands the benefits of evaluation or the process itself, finding a champion is difficult. Even if evaluation efforts are undertaken, these are frequently undermined by poor evaluation direction, design, findings, and recommendations. The consequences of conducting a poor evaluation can be worse than not doing one at all.

Once the expectation for improvements has been created, failure to see such improvements can severely harm the morale and trust of organizational members.

Fear and cynicism are supported not only by poor evaluations but also by past efforts to use evaluation as a means of control and intimidation. Findings—or even the mere “threat” of evaluation—have been used to point the finger at the inadequacies of programs, organizations, and human competence. In fact, even when evaluation has provided useful information for improving programs, it is not uncommon for people to disbelieve the evidence.

Another challenge is the low awareness of the utility and benefits of evaluation. People are not often short of ideas about what to do; the challenge begins with helping them articulate how they will know whether the things they have done or implemented have delivered valuable results. Our culture is one of action, so there is often a false sense of accomplishment in just doing something. Verification and documentation of desired results are often neglected and not viewed as an integral part of what we do unless there is a funding source or a mandate to do so.

Perhaps the biggest challenge—and the most important one—is helping those around us understand that every organization, program, department, function, employee, and resource must be ultimately aligned with positive results and consequences for society (Kaufman, 1992, 2000, 2006a). If what is being used, done, produced, and delivered is not adding benefit to society, it is probably doing quite the contrary. Evaluation and needs assessment are uniquely positioned tools for helping stakeholders make sound decisions about what direction to set, how best to get there, how close they have come to getting there, and what improvements must take place in order to ensure the attainment of organizational and societal ends. In fact, even the business community is embracing this reality through movements like corporate social responsibility. Milton Friedman’s old paradigm about the “business of business is business” is being disputed even by the heads of top management consulting firms like McKinsey, who argues that “social issues are not so much tangential to the business of business as fundamental to it” (Davis, 2005, p. 1).

There are enormous societal needs to fill in all areas: education, physical and mental health, economic development, crime, and discrimination, to name a few. Kaufman has set out his ideal vision of “the world we want to create for tomorrow’s child,” which identifies the basic

22 Performance Evaluation

indicators of societal ends, and thus of needs, and has been used as the basis for strategic planning, needs assessment, and evaluation. Kaufman (2000) defines needs as gaps between what should be accomplished and what is currently accomplished:

There will be no losses of life nor elimination or reduction of levels of survival, self-sufficiency, or quality of life from any source including (but not limited to) the following:

- *War, riot, terrorism, or unlawful civil unrest*
- *Unintended human-caused changes to the environment including permanent destruction of the environment and/or rendering it non-renewable*
- *Murder, rape, or crimes of violence, robbery, or destruction to property*
- *Substance abuse*
- *Permanent or continuing disabilities*
- *Disease*
- *Starvation and/or malnutrition*
- *Destructive behavior (including child, partner, spouse, self, elder, others)*
- *Accidents, including transportation, home, and business/ workplace*
- *Discrimination based on irrelevant variables including color, race, age, creed, gender, religion, wealth, national origin, or location*

Consequences: Poverty will not exist, and every woman and man will earn at least as much as it costs them to live unless they are progressing toward being self-sufficient and self-reliant. No adult will be under the care, custody or control of another person, agency, or substance: all adult citizens will be self-sufficient and self-reliant as minimally indicated by their consumption being equal to or less than their production (p. 95).

Societal ends are not defined by a single organization, and it is not expected that any one organization will accomplish them on its own. These strategic-level objectives represent the shared ambitions of the organizations, individuals, and other partners that stand for our shared communities and society.

ENSURING COMMITMENT

It is vital to get the ownership of evaluation and performance by those who define, deliver, and receive organizational objectives, products, and services. Evaluators and stakeholders must define the required contributions each will make so that they can create a solid partnership for success. Trust, understanding, and agreement on a common destination—results to be achieved—are all key to a successful enterprise. Without the commitment and participation of all of the stakeholders, the success of the evaluation will be less than it could be.

Evaluation data can sometimes be unnerving for stakeholders. Imagine the sense of loss of control when faced with evaluation: on the one hand, they want to know what issues must be resolved and how, and on the other hand, they may resort to any passive-aggressive tactic to keep the evaluator from finding out anything because they are apprehensive that the evaluation will confirm their worst fears.

Consider this situation. A manufacturer implemented a pilot program and rollout to its dealers of a state-of-the-art inventory management and automatic replenishment system. It was designed to minimize inventory (freeing up cash) while maximizing availability to customers (increased sales). The problem was that it quickly and unequivocally shed light on the very poor state of affairs at most dealers, highlighting their expensive inventories and thus discouraging the managers of those inventories from buying in or wanting to participate. Although they had very difficult jobs, they had a vested interest in maintaining the status quo because they did not want others, in particular, their bosses, to find out about their problems. A responsive evaluator would attempt to obtain buy-in from these managers with the common purpose of improving things—in a sense, becoming part of the solution rather than being the problem.

In easing these fears, which are based in part on past experiences with evaluation, the evaluator might want to consider staying away from the term *evaluation* and focusing more on describing the process, which is finding out what is working well, what should be modified and why, and then identifying actions to take that will support continuous improvement. If there is good news, it should be trustworthy. If there is bad news, it is best provided in an environment of trust and the common purpose of continual improvement. Evaluators should never withhold disappointing evidence, but must simultaneously ensure that the successes and shortfalls are based on solid evidence. Trust, common purpose, and shared destiny are keys to getting and maintaining commitment.

Creating the partnership for evaluation and improved performance hinges not only on seriously involving stakeholders but also on listening to them. Although it might be tempting to move ahead with plans and evaluations without the stakeholders' involvement and commitment, doing so risks that later they will see these worthwhile efforts as deceptive or worse. Peter Drucker (1993) had good advice when he suggested that evaluators get "transfer of ownership" of their stakeholders; people see it as their own rather than as belonging to someone else. And the best way to get such a transfer is to involve the partners in setting the objectives and sharing with them the results of any successes and shortfalls. With ways to build trust, evaluation study will be easier, recommendations and findings will have more impact, and the evaluation will stand a better chance of leading to meaningful change.

Evaluation provides the opportunity to have an open and honest relationship with the stakeholders based on performance data, not just biased opinions and perceptions. Involving stakeholders is the best way to ensure that the evaluation meets their expectations and adds demonstrable value.

BENEFITS OF EVALUATION

Conducting an evaluation requires resources, but the benefits outweigh those costs in most situations. Here are some of the many benefits to include in an evaluation proposal or business case:

- Evaluation can provide relevant, reliable, and valid data to help make justifiable decisions about
 - how to improve programs and other solutions,
 - what programs and solutions to continue or discontinue,
 - how to get closer to organizational goals, and
 - whether current goals are worth pursuing.
- Evaluation plans and frameworks provide the basis for design, development, and implementation project management plans.
- Evaluation can identify any adjustments that have to be made during and after development and implementation, so that resources are maximized.

- Evaluation provides the means to document successes so that
 - the merit of decisions, department, staff, and solutions is recognized by all;
 - budget requirements and jobs are justified;
 - the quality of this work is respected by organizational partners;
 - the value of opinions and data is taken into account throughout the organization; and
 - evaluators gain credibility and competence, are granted autonomy and power along with accountability, and are seen as true strategic partners in the organization.
- Evaluation reports can be used to disseminate and market the organization's successes to internal and external partners, such as current and prospective customers.

BASIC DEFINITIONS

Some basic definitions will help convey the concepts in this book:

Performance: The accomplishments of behavior rather than the behavior itself

Performance improvement: A systemic and systematic process for assessing and analyzing performance gaps; planning improvements in performance; designing and developing efficient, effective, and ethically justifiable interventions to close performance gaps; implementing the interventions; and evaluating all levels of results

Ends: Results sought at various organizational levels

Means: The behaviors, activities, processes, procedures, projects, and programs used to achieve results

Levels of results: Society, organization, program, department or team, individual (adapted from Kaufman, 2002)

Needs: Gaps in results (not processes or resources) at any level

Needs assessment: The process of identifying gaps in results and placing them in priority order for resolution

Stakeholders: Anyone who has an interest in the evaluation process and recommendations

Goal: A stated result that identifies a desired end

Objective: A precisely stated goal that identifies who is responsible for achieving it, what accomplishment will be delivered, under what conditions, and with what measurable criteria or metric it will be deemed as reached

Value added: The ultimate result and contribution made by an organization to society—essentially, its societal impact

KEY POINTS

- Evaluation is a systematic way to make decisions based on reliable data.
- Evaluation compares results with expectations, finds drivers and barriers to expected performance, and produces action plans for improving the programs and solutions being evaluated. It can be formative (done during the design of a program or solution) or summative (done in the context of an existing program or solution)
- The approach to evaluation presented throughout this book is based on performance improvement principles.
- A useful evaluation begins by ensuring the commitment and active participation from key stakeholders.
- Evaluators face many challenges, including involving stakeholders, convincing organizations to use the evaluation results, and convincing people that programs require evaluation. If they are not committed to the process, stakeholders can (and will) keep evaluations from being successful.

REFLECTION QUESTIONS

1. How does evaluation affect decision making?
2. How did evaluation evolve into a field?
3. When is a needs assessment more appropriate than an evaluation?
4. How can evaluators ensure commitment from their stakeholders? What happens if they neglect this step?
5. What are the most compelling benefits for conducting an evaluation?