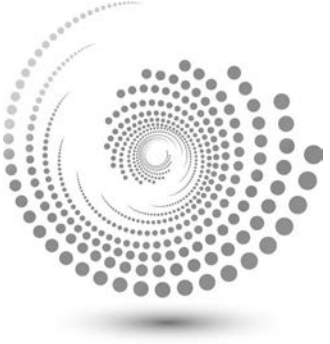
CHAPTER 1



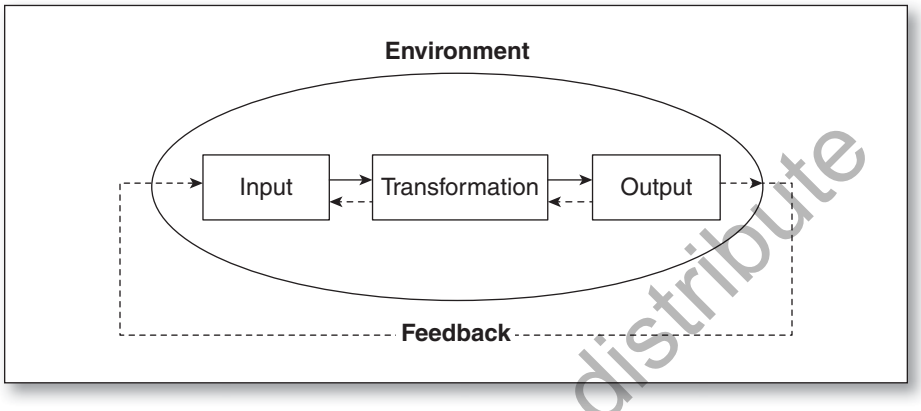
Fundamentals of Program Evaluation

The programs that evaluators can expect to assess have different names such as *treatment* program, *action* program, or *intervention* program. These programs come from different substantive areas, such as health promotion and care, education, criminal justice, welfare, job training, community development, and poverty relief. Nevertheless, they all have in common organized efforts to enhance human well-being—whether by preventing disease, reducing poverty, reducing crime, or teaching knowledge and skills. For convenience, programs and policies of any type are usually referred in this book as "intervention programs" or simply "programs." An intervention program intends to change individuals' or groups' knowledge, attitudes, or behaviors in a community or society. Sometimes, an intervention program aims at changing the entire population of a community; this kind of program is called a *population-based intervention program*.

THE NATURE OF INTERVENTION PROGRAMS AND EVALUATION: A SYSTEMS VIEW

The terminology of systems theory (see, e.g., Bertalanffy, 1968; Ryan & Bohman, 1998) provides a useful means of illustrating how an intervention program works as an open system, as well as how program evaluation serves the program. In a general sense, as an open system an intervention program consists of five components (input, transformation, outputs, environment, and feedback), as illustrated in Figure 1.1.

Figure 1.1 A Systems View of a Program



Inputs. Inputs are resources the program takes in from the environment. They may include funding, technology, equipment, facilities, personnel, and clients. Inputs form and sustain a program, but they cannot work effectively without systematic organization. Usually, a program requires an implementing organization that can secure and manage its inputs.

Transformation. A program converts inputs into outputs through transformation. This process, which begins with the initial implementation of the treatment/intervention prescribed by a program, can be described as the stage during which implementers provide services to clients. For example, the implementation of a new curriculum in a school may mean the process of teachers teaching students new subject material in accordance with existing instructional rules and administrative guidelines. Transformation also includes those sequential events necessary to achieve desirable outputs. For example, to increase students' math and reading scores, an education program may need to first boost students' motivation to learn.

Outputs. These are the results of transformation. One crucial output is the attainment of the program's goals, which justifies the existence of the program. For example, an output of a treatment program directed at individuals who engage in spousal abuse is the end of the abuse.

Environment. The environment consists of any factors that, despite lying outside a program's boundaries, can nevertheless either foster or constrain that program's implementation. Such factors may include social norms, political structures, the economy, funding agencies, interest groups, and concerned

citizens. Because an intervention program is an open system, it depends on the environment for its inputs: clients, personnel, money, and so on. Furthermore, the continuation of a program often depends on how the general environment reacts to program outputs. Are the outputs valuable? Are they acceptable? For example, if the staff of a day care program is suspected of abusing children, the environment would find that output unacceptable. Parents would immediately remove their children from the program, law enforcement might press criminal charges, and the community might boycott the day care center. Finally, the effectiveness of an open system, such as an intervention program, is influenced by external factors such as cultural norms and economic, social, and political conditions. A contrasting system may be illustrative: In a biological system, the use of a medicine to cure an illness is unlikely to be directly influenced by external factors such as race, culture, social norms, or poverty.

Feedback. So that decision makers can maintain success and correct any problems, an open system requires information about inputs and outputs, transformation, and the environment's responses to these components. This feedback is the basis of program evaluation. Decision makers need information to gauge whether inputs are adequate and organized, interventions are implemented appropriately, target groups are being reached, and clients are receiving quality services. Feedback is also critical to evaluating whether outputs are in alignment with the program's goals and are meeting the expectations of stakeholders. Stakeholders are people who have a vested interest in a program and are likely be affected by evaluation results; they include funding agencies, decision makers, clients, program managers, and staff. Without feedback, a system is bound to deteriorate and eventually die. Insightful program evaluation helps to both sustain a program and prevent it from failing. The action of feedback within the system is indicated by the dotted lines in Figure 1.1.

To survive and thrive within an open system, a program must perform at least two major functions. First, internally, it must ensure the smooth transformation of inputs into desirable outcomes. For example, an education program would experience negative side effects if faced with disruptions like high staff turnover, excessive student absenteeism, or insufficient textbooks. Second, externally, a program must continuously interact with its environment in order to obtain the resources and support necessary for its survival. That same education program would become quite vulnerable if support from parents and school administrators disappeared.

Thus, because programs are subject to the influence of their environment, every program is an open system. The characteristics of an open system can also be identified in any given policy, which is a concept closely related to that of a program. Although policies may seem grander than programs—in terms of

the envisioned magnitude of an intervention, the number of people affected, and the legislative process—the principles and issues this book addresses are relevant to both. Throughout the rest of the book, the word *program* may be understood to mean program *or* policy.

Based upon the above discussion, this book defines *program evaluation* as the process of systematically gathering empirical data and contextual information about an intervention program—specifically answers to what, who, how, whether, and why questions that will assist in assessing a program's planning, implementation, and/or effectiveness. This definition suggests many potential questions for evaluators to ask during an evaluation: The "what" questions include those such as, what are the intervention, outcomes, and other major components? The "who" questions might be, who are the implementers and who are the target clients? The "how" questions might include, how is the program implemented? The "whether" questions might ask whether the program plan is sound, the implementation adequate, and the intervention effective. And the "why" questions could be, why does the program work or not work? One of the essential tasks for evaluators is to figure out which questions are important and interesting to stakeholders and which evaluation approaches are available for evaluators to use in answering the questions. These topics will be systematically discussed in Chapter 2. The purpose of program evaluation is to make the program accountable to its funding agencies, decision makers, or other stakeholders and to enable program management and implementers to improve the program's delivery of acceptable outcomes.

CLASSIC EVALUATION CONCEPTS, THEORIES, AND METHODOLOGIES: CONTRIBUTIONS AND BEYOND

Program evaluation is a young applied science; it began developing as a discipline only in the 1960s. Its basic concepts, theories, and methodologies have been developed by a number of pioneers (Alkin, 2013; Shadish, Cook, & Leviton, 1991). Their ideas, which are foundational knowledge for evaluators, guide the design and conduct of evaluations. These concepts are commonly introduced to readers in two ways. The conventional way is to introduce classic concepts, theories, and methodologies exactly as proposed by these pioneers. Most major evaluation textbooks use this popular approach.

This book, however, not only introduces these classic concepts, theories, and methodologies but also demonstrates how to use them as a foundation for formulating additional evaluation approaches. Readers can not only learn from evaluation pioneers' contributions but also expand or extend their work, informed by lessons learned from experience or new developments in program evaluation. However, there is a potential drawback to taking this path. It

requires discussing the strengths and limitations of the work of the field's pioneers. Such critiques may be regarded as intended to diminish or discredit this earlier work. It is important to note that the author has greatly benefited from the classic works in the field's literature and is very grateful for the contributions of those who developed program evaluation as a discipline. Moreover, the author believes that these pioneers would be delighted to see future evaluators follow in their footsteps and use their accomplishments as a basis for exploring new territory. In fact, the seminal authors in the field would be very upset if they saw future evaluators still working with the same ideas, without making progress. It is in this spirit that the author critiques the literature of the field, hoping to inspire future evaluators to further advance program evaluation.

Indeed, the extension or expansion of understanding is essential for advancing program evaluation. Readers will be stimulated to become independent thinkers and feel challenged to creatively apply evaluation knowledge in their work. Students and practitioners who read this book will gain insights from the discussions of different options, formulate their own views of the relative worth of these options, and perform better work as they go forward in their careers.

EVALUATION TYPOLOGIES

Stakeholders need two kinds of feedback from evaluation. The first kind is information they can use to improve a program. Evaluations can function as improvement-oriented assessments that help stakeholders understand whether a program is running smoothly, whether there are problems that need to be fixed, and how to make the program more efficient or more effective. The second kind of feedback evaluations can provide is an accountability-oriented assessment of whether or not a program has worked. This information is essential for program managers and staff to fulfill their obligation to be accountable to various stakeholders.

Different styles of evaluation have been developed to serve these two types of feedback. This section will first discuss Scriven's (1967) classic distinction between formative and summative evaluation and then introduce a broader evaluation typology.

The Distinction Between Formative and Summative Evaluation

Scriven (1967) made a crucial contribution to evaluation by introducing the distinction between formative and summative evaluation. According to Scriven, formative evaluation fosters improvement of ongoing activities. Summative evaluation, on the other hand, is used to assess whether results have met the stated goals.

Summative evaluation informs the go or no-go decision, that is, whether to continue or repeat a program or not. Scriven initially developed this distinction from his experience of curriculum assessment. He viewed the role of formative evaluation in relation to the ongoing improvement of the curriculum, while the role of summative evaluation serves administrators by assessing the entire finished curriculum. Scriven (1991a) provided more elaborated descriptions of the distinction. He defined *formative evaluation* as "evaluation designed, done, and intended to support the process of improvement, and normally commissioned or done, and delivered to someone who can make improvement" (p. 20). In the same article, he defined *summative evaluation* as "the rest of evaluation; in terms of intentions, it is evaluation done for, or by, any observers or decision makers (by contrast with developers) who need valuative conclusions for any other reasons besides development." The distinct purposes of these two kinds of evaluation have played an important role in the way that evaluators communicate evaluation results to stakeholders.

Scriven (1991a) indicated that the best illustration of the distinction between formative and summative evaluation is the analogy given by Robert Stake: "When the cook tastes the soup, that's formative evaluation; when the guest tastes it, that's summative evaluation" (Scriven, p. 19). The cook tastes the soup while it is cooking in case, for example, it needs more salt. Hence, formative evaluation happens in the early stages of a program so the program can be improved as needed. On the other hand, the guest tastes the soup after it has finished cooking and is served. The cook could use the guest's opinion to determine whether to serve the soup to other guests in the future. Hence, summative evaluation happens in the last stage of a program and emphasizes the program's outcome.

Scriven (1967) placed a high priority on summative evaluation. He argued that decision makers can use summative evaluation to eliminate ineffective programs and avoid wasting money. However, Cronbach (1982) disagreed with Scriven's view, arguing that program evaluation is most useful when it provides information that can be used to strengthen a program. He also implied that few evaluation results are used for making go or no-go decisions. Which type of evaluation has a higher priority is an important issue for evaluators, and the importance of this issue will be revisited later in this chapter.

Analysis of the Formative and Summative Distinction

The distinction between formative and summative evaluation provides an important framework evaluators can use to communicate ideas and develop approaches, and these concepts will continue to play an important role. However, Scriven (1991a) proposed that formative and summative evaluations are the two main evaluation types. In reality, there are other important evaluation types that are not

covered in this distinction. To avoid confusion and to lay a foundation for advancing the discipline, it is important to highlight these other evaluation types as well.

In Scriven's conceptualization, evaluation serves to improve a program only during earlier stages of the program (formative evaluation), while evaluation renders a final verdict at the outcome stage (summative evaluation). However, this conceptualization may not sufficiently cover many important evaluation activities (Chen, 1996). For example, evaluations at the early stage of the program do not need to be used to improve the program. Evaluators could administer summative evaluations during earlier phases of the program. Similarly, evaluations conducted at the outcome stage do not have to be summative. Evaluators could administer a formative evaluation at the outcome stage to gain information that would inform and improve future efforts.

Since Scriven regarded Robert Stake's soup-tasting analogy as the best way to illustrate the formative/summative distinction, let's use this analogy to illustrate that all evaluations do not fit this description. According to Stake's analogy, when "the cook tastes the soup," that act represents formative evaluation. This concept of formative evaluation has some limitations. The cook does not always taste the soup for the purpose of improvement. The cook may taste the soup to determine whether the soup is good enough to serve to the guests at all, especially if it is a new recipe. Upon testing the soup, she/he may feel it is good enough to serve to the guests; alternatively, she/he may decide that the soup is awful and not worth improving and simply chuck the soup and scratch it off the menu. In this case, the cook has not tasted the soup for the purpose of improvement but to reach a conclusion about including the soup or excluding it from the menu.

To give another illustration, a Chinese cook, who is a friend of mine, once tried to prepare a new and difficult dish, called Peking duck, for his restaurant. Tasting his product, he found that the skin of the duck was not as crispy as it was supposed to be, nor the meat as flavorful. Convinced that Peking duck was beyond his capability as a chef, he decided not to prepare the dish again. Again, the cook tasted the product to conduct a summative assessment rather than a formative one. The formative/summative distinction does not cover this kind of evaluation.

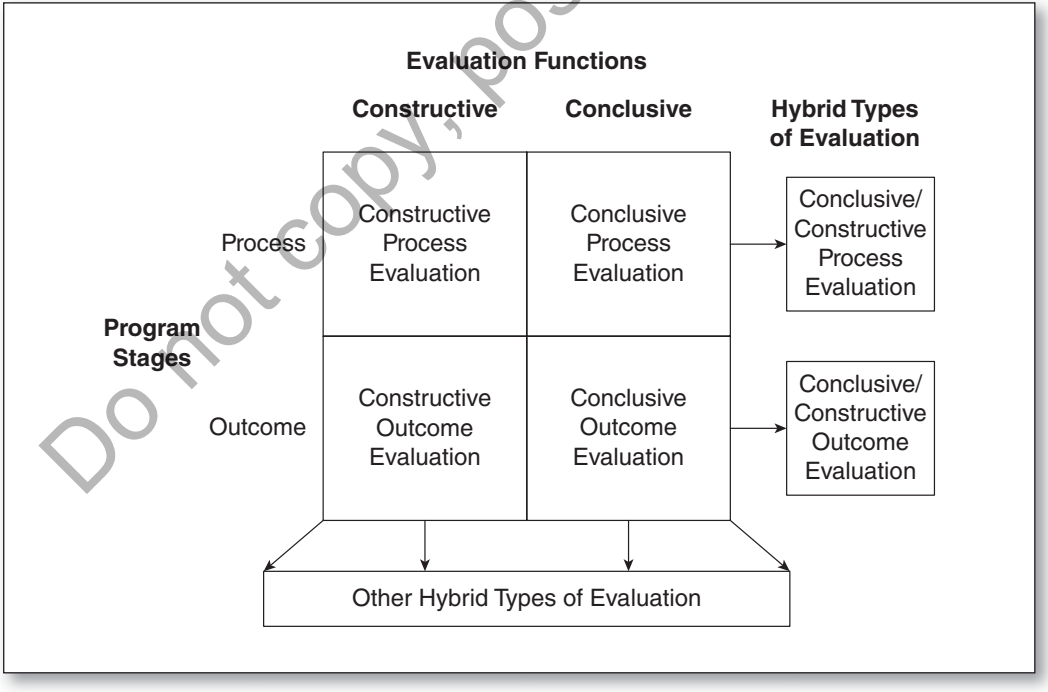
Returning to Stake's analogy, when "the guest tastes the soup," this is regarded as a summative evaluation since the guest provides a conclusive opinion of the soup. This concept of summative evaluation also has limitations. For example, the opinion of the guests is not always used solely to determine the soup's final merit. Indeed, a cook might well elicit opinions from the guests for the purpose of improving the soup in the future. In this case, this type of evaluation is also not covered by the formative/summative distinction.

Stake's analogy, though compelling, excludes many evaluation activities. Thus, we need a broader conceptual typology so as to more comprehensively communicate or guide evaluation activities.

A Fundamental Evaluation Typology

To include more evaluation types in the language used to communicate and guide evaluation activities, this chapter proposes to extend Scriven's formative and summative distinction. The typology developed here is a reformulation of an early work by Chen (1996). This typology has two dimensions: the program stages and evaluation functions. In terms of program stages, evaluation can focus on program process (such as program implementation) and/or on program outcome (such as the impact of the program on its clients). In terms of evaluation functions, evaluation can serve a constructive function (providing information for improving a program) and/or a conclusive function (judging the overall merit or worth of a program). A fundamental typology of evaluation can thus be developed by placing program stages and evaluation functions in a matrix, as shown in Figure 1.2.

Figure 1.2 Fundamental Evaluation Typology



SOURCE: Adapted from Chen (1996).

This typology consists of both basic evaluation types and hybrid evaluation types. The rest of this section will discuss the basic types first and then the hybrid types.

Basic Evaluation Types

The basic types of evaluation include constructive process evaluation, conclusive process evaluation, constructive outcome evaluation, and conclusive outcome evaluation.

Constructive Process Evaluation

Constructive process evaluation provides information about the relative strengths/weaknesses of the program's structure or implementation processes, with the purpose of program improvement. Constructive process evaluation usually does not provide an overall assessment of the success or failure of program implementation. For example, a constructive process evaluation of a family-planning program may indicate that more married couples can be persuaded to utilize birth control in an underdeveloped country if the service providers or counselors are local people, rather than outside health workers. This information does not provide a conclusive judgment of the merits of program implementation, but it is useful for improving the program. Decision makers and program designers can use the information to strengthen the program by training more local people to become service providers or counselors.

Conclusive Process Evaluation

This type of evaluation, which is frequently used, is conducted to judge the merits of the implementation process. Unlike constructive process evaluation, conclusive process evaluation attempts to judge whether the implementation of a program is a success or a failure, appropriate or inappropriate. A good example of conclusive process evaluation is an assessment of whether program services are being provided to the target population. If an educational program intended to serve disadvantaged children is found to serve middle-class children, the program would be consider an implementation failure. Another good example of conclusive process evaluation is manufacturing quality control, when a product is rejected if it fails to meet certain criteria. Vivid examples of conclusive process evaluation are the investigative reports seen on popular TV programs, such as *60 Minutes* and *20/20*. In these programs, reporters use hidden cameras to document

whether services delivered by such places as psychiatric hospitals, nursing homes, child care centers, restaurants, and auto repair shops are appropriate.

Constructive Outcome Evaluation

This type of evaluation identifies the relative strengths and/or weaknesses of program elements in terms of how they may affect program outcomes. This information can be useful for improving the degree to which a program is achieving its goals, but it does not provide an overall judgment of program effectiveness. For example, evaluators may facilitate a discussion among stakeholders to develop a set of measurable goals or to reach consensus about program goals. Again, such activity is useful for improving the program's chance of success, but it stops short of judging the overall effectiveness of the program. This type of evaluation will be discussed in detail in Chapter 9. In another example, a service agency may have two types of social workers, case managers whose work is highly labor-intensive and care managers whose work is less labor-intensive. An evaluator can apply constructive outcome evaluation to determine which kind of social worker is more cost-effective for the agency.

Conclusive Outcome Evaluation

The purpose of a conclusive outcome evaluation is to provide an overall judgment of a program in terms of its merit or worth. Scriven's summative evaluation is synonymous with this category. A typical example of conclusive outcome evaluation is validity-focused outcome evaluation that determines whether changes in outcomes can be causally attributed to the program's intervention. This kind of evaluation is discussed in detail in Chapter 10.

The typology outlined above eliminates some of the difficulties found in the soup-tasting analogy. Formerly, when the cook tasted the soup for conclusive judgment purposes, this activity did not fit into the formative/summative distinction. However, it can now be classified as conclusive process evaluation. Similarly, when the guest tastes the soup for improvement purposes, this action can now be classified as constructive outcome evaluation.

Furthermore, the typology clarifies the myth that process evaluation is always a kinder, gentler type of evaluation in which evaluators do not make tough conclusive judgments about the program. Constructive process evaluation may be kinder and gentler, but conclusive process evaluation is not necessarily so. For example, TV investigative reports that expose the wrongdoing in a psychiatric hospital, auto shop, restaurant, or day care center have resulted in changes in service delivery, the firing of managers and employees, and even the closing of

the agencies or businesses in question. In such cases, process evaluations were tougher than many outcome evaluations in terms of critical assessment and impact. Moreover, the basic typology disrupts the notion that outcome evaluation must always be carried out with a "macho" attitude so that it threatens program providers while failing to offer any information about the program. A conclusive outcome evaluation may provide information whether a program has been successful or not, but the constructive outcome evaluation can provide useful information for enhancing the effectiveness of a program without threatening its existence. For example, the survival of a program is not threatened by a constructive outcome evaluation that indicates that program effectiveness could be improved by modifying some intervention elements or procedures.

Hybrid Evaluation Types

Another important contribution of this fundamental evaluation typology is to point out that evaluators can move beyond the basic evaluation types to conduct hybrid evaluations. As illustrated in Figure 1.2, a hybrid evaluation can combine evaluation functions, program stages, or both (Chen, 1996). This section intends to introduce two types of hybrid evaluation that, across evaluation, functions at a program stage.

Conclusive/Constructive Process Evaluation

Conclusive/constructive process evaluation serves both accountability and program improvement functions. A good example is evaluation carried out by the Occupational Safety and Health Administration (OSHA). OSHA inspectors may evaluate a factory to determine whether the factory passes a checklist of safety and health rules and regulations. The checklist is so specific, however, that these inspections can also be used for improvement. If a company fails the inspection, the inspector provides information concerning areas that need correction to satisfy safety standards. Other regulatory agencies, such as the Environmental Protection Agency (EPA), perform a similar type of evaluation. In these kinds of evaluation, the overall quality of implementation is represented by a checklist of crucial elements. These elements provide exact clues for how to comply with governmental regulations.

A similar principle can be applied to assess the implementation of an intervention. As will be discussed in Chapter 7, a conclusive/constructive process evaluation can look into both overall quality and discrete program elements so as to provide information about the overall quality of implementation as well as specific areas for its future improvement.

Conclusive/Constructive Outcome Evaluation

Another hybrid evaluation type is the conclusive/constructive outcome evaluation. An excellent example of this kind of evaluation is real-world outcome evaluation, which will be discussed in great detail in Chapter 11. Another excellent example is theory-driven outcome evaluation. This type of evaluation elaborates causal mechanisms underlying a program so that it examines not only whether the program has an impact but why. It also informs stakeholders as to which mechanisms influence program success or failure for program improvement purposes. Theory-driven outcome evaluation will be discussed in Chapters 12 and 14 of the book.

Applications of the Fundamental Evaluation Typology

The fundamental evaluation typology discussed here prevents evaluators from hewing rigidly to just two types of evaluation, that is, formative evaluation in the early stages of the program and summative evaluation toward the end. The fundamental evaluation typology provides evaluators and stakeholders many options for devising basic or hybrid types of evaluation at implementation and outcome stages so as to best meet stakeholders' needs. However, the fundamental evaluation typology does not cover the planning stage. Thus, Chapter 2 will expand the fundamental evaluation typology into a comprehensive evaluation typology that covers a full program cycle from program planning to implementation to outcome. Then the rest of the book will provide concrete examples of these evaluation approaches and illustrate their applications across the entire life cycle of programs.

INTERNAL VERSUS EXTERNAL EVALUATORS

Evaluators are usually classified into two categories: internal and external evaluators. Internal evaluators are employed by an organization and are responsible for evaluating the organization's own programs. External evaluators are not employees of the organization but are experts hired from outside to evaluate the program. One of the major differences between the two is independence. Internal evaluators are part of the organization. They are familiar with the organizational culture and the programs to be evaluated. Like other employees, they share a stake in the success of the organization. External evaluators are not constrained by organizational management and relationships with staff members and are less invested in the program's success. The general conditions that tend to favor either internal evaluation or external evaluation are summarized as follows:

Internal Evaluation

- Cost is a great concern.
- Internal capacity/resources are available.
- The evaluator's familiarity with the program is important.
- The program is straightforward.
- Evaluation is for the purpose of monitoring or is constructive in nature.

External Evaluation

- The cost of hiring an external evaluator is manageable.
- Independence and objectivity are essential.
- A program is large or complicated.
- The evaluation will focus on conclusive assessment or conclusive/constructive assessment.
- Comprehensive assessment or fresh insight is needed.

POLITICS, SOCIAL JUSTICE, EVALUATION STANDARDS, AND ETHICS

One important distinction that separates program evaluation from research is that evaluations are carried out under political processes. The purpose of an evaluation is to evaluate an intervention program. However, the program is created by political processes. What kinds of programs are to be funded? Which programs need evaluation in a community? These decisions are made through bargaining and negotiation by key players such as politicians and advocacy groups. After a program is funded and evaluators are hired to evaluate it, the focus of the evaluation and the questions to be asked are determined, or largely influenced, by stakeholders. Cronbach and colleagues (1980) argued that a theory of evaluation must be as much a theory of political interaction as it is a theory of how to determine facts. Weiss (1998), too, indicated that evaluators must understand the political nature of evaluations and be aware of the obstacles and opportunities that can impinge upon evaluation efforts.

Since evaluation provides feedback to a program, evaluators may have high hopes that decision makers will use the findings as a basis for action. However, since program evaluation is part of political processes, evaluation findings are just one of many inputs that decision makers use. Decision making is more often based on factors such as political support and community service needs than evaluation findings. Since evaluations take place within a political and an organizational context, Chelimsky (1987) stated that evaluators are shifting their view of the role evaluations play, from reforming society to the more realistic aim of bringing the

best possible information to bear on a wide variety of policy questions. Also because evaluation takes place in a political environment, evaluators' communication skills are critical. Evaluators' qualifications should include research skills but should emphasize group facilitation skills, political adroitness, managerial ability, and cultural sensitivity to multiple stakeholders.

In evaluation, stakeholders are those persons, groups, or organizations who have a vested interest in the evaluation results. Stakeholders often are not a homogenous group but rather multiple groups with different interests, priorities, and degrees of power or influence. The number of stakeholder groups evaluators must communicate with often depends on the magnitude of an intervention program. In a small community-based program, key stakeholders may include the program director, staff, and clients. Stakeholder groups of a large federal program, on the other hand, could include federal agencies, state agencies, community-based organizations, university researchers, clients, program directors, program administrators, implementers, community advocates, computer experts, and so on.

Evaluators are usually hired by decision makers, and one of the major purposes of program evaluation is to provide information to decision makers that they will use to allocate funds or determine program activities. This contractual arrangement has a potential to bias evaluators toward the groups in power, that is, the decision makers who hire them or the stakeholders with whom the decision makers are most concerned. Critics such as House (1980) argued that evaluation should address social justice and specifically the needs and interests of the poor and powerless. However, Scriven (1997) and Chelimsky (1997) were concerned that when evaluators take on the role of program advocates, their evaluations' credibility will be tarnished.

Social justice is a difficult issue in evaluation. Participatory evaluation has the potential to alleviate some of the tension between serving social justice and decision makers. Including representatives of the various stakeholder groups in evaluation has been proposed as a way to address some social justice issues. Generally, stakeholders participate in an evaluation for two purposes: practical and transformative (Greene, Lincoln, Mathison, Mertens, & Ryan, 1998). Practical participatory evaluation is meant to enhance evaluation relevance, ownership, and utilization. Transformative participatory evaluation seeks to empower community groups to democratize social change. Either way, participatory evaluation can provide evaluators with an opportunity to engage with different stakeholder groups and balance diverse views, increase buy-in from all stakeholder groups, and enhance their willingness to use evaluation results.

Another way of enhancing evaluators' credibility is to promote professional ethics. Like other professionals, evaluators must adhere to professional ethics and standards. The American Evaluation Association (2004) adopted the following ethical principles for evaluators to follow:

- *Systematic inquiry.* Evaluators conduct systematic, data-based inquiries.
- *Competence.* Evaluators provide competent performance to stakeholders.
- *Integrity/honesty.* Evaluators ensure honesty and integrity of the entire evaluation process.
- *Respect for people.* Evaluators respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders.
- *Responsibilities for general and public welfare.* Evaluators articulate and take into account the diversity and values that may be related to the general and public welfare. ("The Principles")

In addition, to ensure the credibility of evaluation, the Joint Committee on Standards for Education (Yarbrough, Shulha, Hopson, & Caruthers, 2011) has specified the following five core standards for evaluators to follow:

1. *Utility standards*. The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.

2. *Feasibility standards*. The feasibility standards are intended to increase evaluation effectiveness and efficiency.

3. *Propriety standards*. The propriety standards support what is proper, fair, legal, right, and just in evaluations.

4. *Accuracy standards*. The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.

5. *Evaluation accountability standards*. The evaluation accountability standards encourage adequate documentation of evaluations and a meta-evaluative perspective focused on improvement of and accountability for evaluation processes and products.

EVALUATION STEPS

The Centers for Disease Control and Prevention (CDC) published the *CDC Framework of Program Evaluation for Public Health* (CDC, 1999) to help evaluators understand how to conduct evaluation based on evaluation standards. The document specified six steps that are useful guides to the evaluation of public health and social betterment programs:

Step 1: Engage Stakeholders deals with engaging individuals and organizations with an interest in the program in the evaluation process.

Step 2: Describe the Program involves defining the problem, formulating program goals and objectives, and developing a logic model showing how the program is supposed to work.

Step 3: Focus the Evaluation Design determines the type of evaluation to implement, identifies the sources needed to implement the evaluation, and develops evaluation questions.

Step 4: Gather Credible Evidence identifies how to answer the evaluation questions and develop an evaluation plan that will include, among other things, indicators, data sources and methods for collecting data, and the timeline.

Step 5: Justify Conclusions involves collecting, analyzing, and interpreting the evaluation data.

Step 6: Ensure Use and Share Lessons Learned identifies effective methods for sharing and using the evaluation results.

EVALUATION DESIGN AND ITS COMPONENTS

When proposing an evaluation to stakeholders or organizations such as funding agencies, evaluators must describe the evaluation's purposes and methodology. An evaluation design needs to include at least five components:

1. *Purposes of and Background Information about the Intervention Program.* The first thing that evaluators need to do when assessing an intervention program is to gain a solid knowledge of the background of the program and document this understanding. Background information includes the purposes of the intervention program, the target population, the organizations responsible for implementing the program, key stakeholders of the program, implementation procedures, reasons for conducting the evaluation, the evaluation's timeline, the resources that will be used, and who will utilize the evaluation results. Evaluators usually gather information by reviewing existing documents such as program reports and the grant application proposal, as well as by interviewing key stakeholders of the program. The background information serves as a preliminary basis for communication by evaluators and stakeholders about the program and evaluation.

2. *A Logic Model or Program Theory for Describing the Program.* A sound evaluation requires a systematic and coherent description of the intervention program, which will serve as a basis for communication between evaluators and stakeholders and for the evaluation design. In reality, a systematic and coherent program description is often not available. It is unwise for evaluators to conduct a program evaluation without a mutual agreement with stakeholders about what the program looks like. In this situation, how could an evaluation provide useful information to stakeholders? Or, even worse, stakeholders later could easily claim that an evaluation failed to accomplish what they expected from it, if the evaluation results do not convey good news. Program description is an important step in evaluation.

If a program does not have a systematic and coherent program description, evaluators must facilitate stakeholders in developing one. This book discusses two options for describing a program: logic models and program theory. Logic models are used to identify the major components of a program in terms of a set of categories such as inputs, activities, outputs, and outcomes. However, if evaluators and stakeholders are interested in looking into issues such as contextual factors and causal mechanisms, this book encourages the use of program theory. Both logic models and program theory will be discussed in Chapter 3.

3. *Assertion of a Program's Stage of Development.* As will be discussed in the next chapter, an intervention program's life cycle can be generally classified as being in one of four phases: planning, initial implementation, mature implementation, and outcome. Program designers, during the planning phase, work with partners to identify or develop an intervention and organize resources and activities for supporting the intervention. After the planning phase, the program goes into the initial implementation phase. The major tasks here are training implementers, checking clients' acceptance, and ensuring appropriate implementation. After the initial implementation, the program progresses to the mature implementation stage. The major tasks here include ensuring or maintaining the quality of implementation. During the outcome phase, the program is expected to have desirable impacts on clients. The different stages of a program require different evaluation approaches. For example, constructive evaluation is most useful to a program during the initial implementation stage when it can help with service delivery, but it is not appropriate for a formal assessment of a program's merits at the outcome stage.

Evaluators and stakeholders have to agree on which stage a program is in to select an appropriate evaluation type(s) and approach. Chapter 2 will provide detailed discussions of the nature of program stages and how they relate to different evaluation types and approaches.

4. *Evaluation Types, Approaches, and Methodology.* This component is the core of evaluation design. Using information regarding the evaluation's purposes and the logic model/program theory, evaluators and stakeholders need to determine what type of evaluation, whether one of the basic evaluation types—constructive process, conclusive process, constructive outcome, or conclusive outcome—or a hybrid type, is suitable for correctly evaluating the program. Once program stage and evaluation type are determined, evaluators can move on to select or design an evaluation approach or approaches for evaluating a program. Chapter 2 will provide a comprehensive typology for guiding evaluators in selection of evaluation types and approaches.

Determining the most appropriate evaluation approach is challenging and time-consuming. However, it ensures that all involved share a mutual understanding of why a particular evaluation type has been selected. Without it, stakeholders are likely to find that the results of the evaluation address issues that are not of concern to them and/or are not useful to them. Stakeholders are often not trained on evaluation techniques. They often do not express what they expect and need from an evaluation as clearly and precisely as evaluators could hope. Evaluators usually must double- or even triple-check with stakeholders to make sure everyone shares the same understanding and agrees on the evaluation's purposes up front.

5. *Budget and Timeline.* Regardless of stakeholders' and evaluators' visions of an ideal evaluation plan, the final evaluation design is bound to be shaped by the money and time allocated. For example, if stakeholders are interested in a rigorous assessment of an intervention program's outcomes but can provide only a small evaluation budget, the research method used in the evaluation is not likely to be a randomized controlled trial over a few years, which would likely cost over a few million dollars. Similarly, if the timeline is short, evaluators will likely use research methods such as rapid assessments rather than conduct a thorough evaluation.

When facilitating stakeholders in making an informed decision, it is highly preferable for evaluators to propose a few options and explain the information each option is likely to provide, as well as the price tag of each.

MAJOR CHALLENGES OF EVALUATION: LESSONS LEARNED FROM PAST PRACTICE

Program evaluation has been practiced over several decades. Lessons learned from experience indicate that program evaluation faces a set of unique challenges that are not faced by other disciplines.

Judge a Program Not Only by Its Results but Also by Its Context

One important characteristic distinguishing program evaluation is its need, rarely shared by other disciplines, to use a holistic approach to assessment. The holistic approach includes contextual or transformation information when assessing the merit of a program. By comparison, product evaluation is more streamlined, perhaps focusing solely on the intrinsic value of its object. Products like televisions can be assessed according to their picture, sound, durability, price, and so on. In many situations, however, the value of a program may be contextual as well as intrinsic or inherent. That is, to adequately assess the merit of a program, both its intrinsic value and the context in which that value is assigned must be considered together. For example, say an educational program has, according to strictly performance-based evaluation, attained its goals (which are its intrinsic values). But in what context was the performance achieved? Perhaps the goal of higher student scores on standardized tests was attained by just "teaching students the tests." Does the program's performance still deserve loud applause? Probably not.

Similarly, what about a case in which program success is due to the participation of a group of highly talented, well-paid teachers with ample resources and strong administrative support, but the evaluated program is intended for use in ordinary public schools? This "successful" program may not even be relevant, from the viewpoint of the public schools, and is not likely to solve any of their problems. Therefore, *how* a program achieved its goals is just as important as *whether* it achieved them. For example, an outcome evaluation of one family-planning program in a developing country limited its focus to the relationship between program inputs and outputs; it appeared possible, on this basis, to claim success for the program. A large drop in the fertility rate was indeed observed following the intervention. Transformation information, however, showed that such a claim was misleading. Although the drop in fertility was real, it had little to do with the intervention. A larger factor was that, following implementation, a local governor of the country, seeking to impress his prime minister with the success of the program, ordered soldiers to seize men on the streets and take them to be sterilized. An evaluator with a less holistic approach might have declared that the goals of the program were attained, whereas other people's personal knowledge led them to condemn the program as inhumane. Lacking a holistic orientation, program evaluation may reach very misleading conclusions.

Evaluations Must Address Both Scientific and Stakeholder Credibility

Program evaluation is both a science and an art. Evaluators need to be capable of addressing both scientific and stakeholder credibility in an evaluation. The scientific credibility of program evaluation reflects the extent to which that evaluation was governed by scientific principles. Typically, in scientific research, scientific credibility is all that matters. The more closely research is guided by scientific principles, the greater its credibility. However, as an applied science, program evaluation also exhibits varying degrees of *stakeholder* credibility. The stakeholder credibility of a program evaluation reflects the extent to which stakeholders believe the evaluation's design gives serious consideration to their views, concerns, and needs.

The ideal evaluation achieves both high scientific and high stakeholder credibility, and the two do not automatically go hand in hand. An evaluation can have high scientific credibility but little stakeholder credibility, as when evaluators follow all the scientific principles but set the focus and criteria of evaluation without considering stakeholders' views and concerns. Their evaluation will likely be dismissed by stakeholders, despite its scientific credibility, because it fails to reflect the stakeholders' intentions and needs. For example, there are good reasons for African-Americans to be skeptical of scientific experiments that lack community input, due to incidents such as the Tuskegee syphilis experiment (Jones, 1981/1993). Researchers in the experiment withheld effective treatment from African-American men suffering from syphilis so that the long-term effects of the disease could be documented. Conversely, an evaluation overwhelmed by the influence of stakeholders, such as program managers and implementers, may neglect its scientific credibility, resulting in suspect information.

One of the major challenges in evaluation is how to address the tension between scientific credibility and stakeholder credibility. Evaluation theorists, such as Scriven (1997), argued that objectivity is essential in evaluation because without it, evaluation has no credibility. On the other hand, Stake (1975) and Guba and Lincoln (1981) argued that evaluations must respond to stakeholders' views and needs in order to be useful. Both sides make good points, but objectivity and responsiveness are conflicting values. How would evaluators address this tension?

One strategy is to prioritize, choosing one type of credibility to focus on. However, this prioritization strategy does not satisfactorily address the conflict between the two values. A better strategy, proposed by and used in this book, is perhaps to strike a balance between the two. For example, evaluators might

pursue stakeholder credibility in the earliest phases of evaluation design but turn their attention toward scientific credibility later in the process. Initially, evaluators experience a great deal of interaction and communication with a program's stakeholders for the specific purpose of understanding their views, concerns, and needs. Evaluators then incorporate the understanding they have acquired into the research focus, questions, and design, along with the necessary scientific principles. From this point on, to establish scientific credibility, the evaluators require autonomy to design and conduct evaluations without interference from stakeholders. Stakeholders are usually receptive to this strategy, especially when evaluators explain the procedure to them at the beginning of the process. While stakeholders do not object to a program being evaluated, or dispute the evaluator's need to follow scientific procedures, they do expect the evaluation to be fair, relevant, and useful (Chen, 2001).

As will be discussed in the rest of the book, the tension between scientific and stakeholder credibility arises in many situations. Such tension makes evaluation challenging, but resolving it is essential for advancing program evaluation.

Evaluations Must Provide Information That Helps Stakeholders Do Better

Earlier in this chapter, we learned that Scriven placed a higher priority on conclusive assessment than on program improvement, while Cronbach preferred otherwise. This is an important, but complicated, issue for evaluators. Many evaluators quickly learn that stakeholders are eager to figure out what to do next in order to make a program work better. Stakeholders find evaluations useful if they both offer conclusions about how well programs have worked *and* provide information that assists the stakeholders in figuring out what must be done next to maintain—or even surpass—program goals. Thus, the assessment of a program's performance or merit is only one part of program evaluation (or, alone, provides a very limited type of evaluation). To be most useful, program evaluation needs to equip stakeholders with knowledge of the program elements that are working well and those that are not. Program evaluation in general should facilitate stakeholders' search for appropriate actions to take in addressing problems and improving programs. There are important reasons why evaluations must move beyond narrow merit assessment into the determination of needed improvements. In the business world, information on product improvement is provided by engineering and market

research; likewise, in the world of intervention programs, the agency or organization overseeing an effort relies on program evaluation to help it continually guarantee or improve the quality of services provided.

Consider that intervention programs typically operate in the public sector. In the private sector, the existence or continuation of a product is usually determined by market mechanisms. That is, through competition for consumers, a good product survives, and a bad product is forced from the market. However, the great majority of intervention programs do not encounter any market competition (Chen, 1990). Drug abusers in a community may find, for example, that only one treatment program is available to them. In the absence of an alternative, the treatment program is likely to continue whether or not its outcomes justify its existence. Furthermore, well-known programs with good intentions, such as Head Start, would not be discontinued based on an evaluation saying the programs were ineffectual; decision makers rarely use program evaluation results alone to decide whether a program will go on.

Under these circumstances, an evaluation that simply assesses the merit of a program's past performance and cannot provide stakeholders with insights to help them take the next step is of limited value (Cronbach, 1982). In fact, many stakeholders look to a broad form of program evaluation to point out apparent problems, as well as strengths upon which to build. In general, to be responsive and useful to stakeholders, program evaluation should meet both assessment needs *and* improvement needs rather than confine itself solely to conclusive assessment. Stakeholders need to know whether the program is reaching the target group, the treatment/intervention is being implemented as directed, the staff is providing adequate services, the clients are making a commitment to the program, and the environment seems to be helping the delivery of services. Any part of this information can be difficult for stakeholders to collect; thus, program evaluators must have the necessary training and skills to gather and synthesize it all systematically.

In a broad sense, therefore, merit assessment is a means, rather than the end, of program evaluation. Our vision of program evaluation should extend beyond the design of supremely rigorous and sophisticated assessments. It is important to grasp that evaluation's ultimate task is *to produce useful information that can enhance the knowledge and technology we employ to solve social problems and improve the quality of our lives.*

Furthermore, as discussed in the last section, constructive evaluation for program improvement and conclusive evaluation for merit assessment are not mutually exclusive categories. Evaluation does not have to focus on either program improvement or merit assessment. The introduction of hybrid evaluation types in this book provides options by which evaluation can address both issues.

ADDRESSING THE CHALLENGES: THEORY-DRIVEN EVALUATION AND THE INTEGRATED EVALUATION PERSPECTIVE

To better address these challenges, this book applies the frameworks provided by the theory-driven evaluation approach and the integrated evaluation perspective.

Theory-Driven Evaluation Approach

The theory-driven evaluation approach requires evaluators to understand assumptions made by stakeholders (called program theory) when they develop and implement an intervention program. Based on stakeholders' program theory, evaluators design an evaluation that systematically examines how these assumptions operate in the real world. By doing so, they ensure that the evaluation addresses issues in which the stakeholders are interested. The usefulness of the theory-driven evaluation approach has been discussed intensively in the evaluation literature (e.g., Chen, 1990, 2005, 2012a, 2012b; Chen & Rossi, 1980, 1983a; Chen & Turner, 2012; Coryn, Noakes, Westine, & Schröter, 2011; Donaldson, 2007; Funnell & Rogers, 2011; Nkwake, 2013; Rossi, Lipsey, & Freeman, 2004; Weiss, 1998). The concept and application of program theory will be intricately discussed in Chapter 3.

It is important to know that theory-driven evaluation provides a sharp contrast to traditional method-driven evaluation. Method-driven evaluation views evaluation as mainly an atheoretical activity. Evaluation is carried out by following research steps of a chosen research method such as randomized experiments, survey, case study, focus group, and so on. Within this tradition, evaluation does not need any theory. If evaluators are familiar with the research steps of a particular method, then they can apply the same research steps and principles across different types of programs in different settings. To some degree, method-driven evaluation simplifies evaluation tasks. However, because the focus of method-driven evaluation is mainly on methodological issues, it often does not capably address stakeholders' views and needs. The theory-driven evaluation approach argues that while research methods are important elements of an evaluation, evaluation should not be dictated or driven by one particular method.

Because theory-driven evaluation uses program theory as a conceptual framework for assessing program effectiveness, it provides information not only on whether an intervention is effective but also how and why a program

is effective. In other words, it is capable of addressing the challenge discussed in the last section: The success of a program has to be judged not only by its results but also by its context. This approach is also useful for addressing the following challenge: Evaluation must be capable of providing information for stakeholders to do better. The theory-driven evaluation approach will be intensively discussed in Chapters 3, 7, 12, 13, and 14.

Integrated Evaluation Perspective

Program evaluation is challenging because it has to provide evaluative evidence for a program that meets two requirements. The first requirement is that the evaluative evidence must be credible; that is, program evaluation has to generate enough credible evidence to gain a scientific reputation. This requirement is called the scientific requirement. The second requirement is that the evidence must respond to the stakeholders' views, needs, and practices so as to be useful. Stakeholders are consumers of evaluation. Program evaluation has little reason to exist unless it is able to adequately serve stakeholders' needs. This requirement is called the stakeholder requirement.

Ideally, evaluations should meet both requirements, but in reality evaluators often find it difficult to meet both. One the one hand, they must apply rigorous methods to produce credible evidence. On the other hand, evaluators often find it difficult to apply rigorous methods—such as randomized controlled trials (RCTs)—to evaluate real-world programs given insufficient resources and short time lines. In many situations, administrative hindrances and ethnic concerns add barriers to such an application. Furthermore, even should these barriers be removed and a rigorous method applied, stakeholders may feel that the focus of the evaluation is then too narrow or too academic to be relevant or useful to them. The reason for this disconnect is that the stakeholders' views on community problems and how to solve them are quite different from the conventional scientific methods' underlying philosophy—reductionism. Reductionism postulates that a program is stable and can be analytically reduced to a few core elements. If a program can be reduced to core components, such as intervention and outcome, then an adjustment can be implemented and desirable changes will follow. Given this view, the evaluators' main task is to rigorously assess whether the change produces predetermined outcomes.

However, stakeholders' views on and experiences with social problems and addressing them in a community are more dynamic and complicated

than those assumed by reductionism. Their views can be characterized as the following:

1. *An intervention program is implemented as a social system.* In a social system, contextual factors in a community—such as culture, norms, social support, economic conditions, and characteristics of implementers and clients—are likely to influence program outcomes. As discussed at the beginning of this chapter, program interventions are open systems, not closed like a biological system in terms of contextual factors.

2. *Health promotion/social betterment programs require clients, with the help of implementers, to change their values and habits in order to work.* Unfortunately, people are notoriously resistant to changing their values and habits. For example, an education program may require children fond of playing video games to substantially cut down on game playing to make time for studying; these children may vastly prefer playing the latest zombie massacre game to studying. Victims of bullying in schools may be asked to start reporting bullying incidents to school authorities and parents; based on past experience, these victims may believe reporting these incidents is useless or even dangerous. Because an intervention requires changes, its demands may be highly challenging to both clients and implementers. Not only must program designers wrestle with this challenge when designing an effective intervention program but evaluators must also take this reality into consideration when designing a useful evaluation.

Because of the above factors, stakeholders believe that they need to take a much broader approach in solving a community problem. An intervention is not a stand-alone entity but, rather, has to connect to contextual factors and/or change clients' values and habits to work. Their broad view of community problem solving is inconsistent with the traditional scientific methods, which focus on narrow issues such as assessing the causal relationships between an intervention and its outcomes. The inconsistency between stakeholders' views and reductionism's assumptions regarding community problems and interventions is partly why there is such a huge chasm between the academic and practice communities regarding interventions, as will be discussed in Chapter 15.

Stakeholders respect the value and reputation of scientific methods but view the information provided by using them as just one piece of a jigsaw puzzle they need to assemble. They need other pieces to complete the picture. They hope evaluators can figure out ways to provide all, not just one, of those pieces to them. Stakeholders are concerned that, if evaluators focus too much

on the scientific piece, it will blind them or prevent them from simultaneously investigating other means to solve the puzzle. Stakeholders' views on community problem solving are relevant to ideas proposed by systems thinking (e.g., Meadows, 2008). According to systems thinking, a system is made up of diverse and interactive elements and must address environmental turbulence. Problem solving thus requires the modification of groups of variables simultaneously.

The above analysis shows that evaluators face a dilemma in meeting the scientific requirement and the responsiveness requirement at the same time. An evaluation emphasizing the scientific requirement may scarify the responsiveness requirement, and vice versa. The dilemma has significant implications for evaluation practices, but it has not been intensively and systematically discussed in the literature. There are three general strategies evaluators use to address the dilemma:

Prioritizing the Scientific Requirement as the Top Priority in Evaluation: The first strategy is to stress the scientific requirement by arguing that evaluation's utility relies on whether it can produce credible evidence. Following this general strategy, evaluators must apply rigorous methods as best as they can. Issues related to the responsiveness requirement are addressed only when they do not compromise the rigor issues. Currently, this strategy is the most popular one used by evaluators (Chen, Donaldson, & Mark, 2011). The strategy appeals particularly to evaluators who are strongly committed to scientific values and evidence-based interventions.

Prioritizing the Responsiveness Requirement as the Top Priority in Evaluation. The second strategy is to put the emphasis on the responsiveness requirement. This strategy requires that evaluators use a participatory evaluation approach and qualitative methods to meet stakeholders' information needs (e.g., Cronback, 1982; Stake, 1975). This method is attractive to evaluators who view traditional scientific methods as too narrow and rigid to accommodate stakeholders' views and to meet their informational needs.

Synthesizing the Scientific and Responsiveness Requirements in Evaluation. The third general strategy is to synthesize the scientific and responsiveness requirements in evaluation. This strategy does not prioritize either requirement as the prime focus and thus avoids maximizing one at the expense of the other. Evaluations following this strategy may not be able to provide highly rigorous evidence but can provide good-enough evidence to balance the scientific and responsiveness requirements.

The first two strategies have merits. They are especially useful when there is a strong mandate for evaluation to be either highly rigorous or highly responsive. However, the author believes that, in many typical intervention programs, stakeholders are more likely to benefit from evaluations that use the synthesizing strategy. This book advocates this strategy and formally calls it the *integrated evaluation perspective*. Specifically, the integrated evaluation perspective urges evaluators to develop evaluation theories and approaches that can synthetically integrate *stakeholders' views and practices,* thus acknowledging the *dynamic* nature of an intervention program in a community, with *scientific principles and methods* for enhancing the usefulness of evaluation.

In spite of its conceptual appeals, the integrated evaluation perspective faces a challenge in developing specific evaluation theories and approaches to guide the work. It does not have advantages such as the scientific prioritization strategy. For example, advocates of the scientific prioritization strategy can borrow scientific methods and models developed by more matured disciplines and apply them to evaluation. The integrated evaluation perspective, however, does not have this ability because other disciplines do not face the kind of inconsistency between scientific and responsiveness requirements experienced in evaluation. They thus do not need to deal with synthesizing issues. For example, in biomedical research, both researchers and physicians consistently demand rigorous evidence for a medicine's efficacy. Accordingly, biomedical research cannot offer evaluators clues or solutions on synthesizing the conflict between scientific and responsiveness requirements. The integrated evaluation perspective, therefore, requires evaluators to develop innovative, *indigenous* theories and approaches to synthesize the requirements unique to the discipline.

This book contributes to the integrated evaluation perspective by introducing many innovative, indigenous theories and approaches evaluators can use in balancing the scientific and responsiveness requirements. At the same time, this book does not neglect traditional theories and approaches promoted by the scientific prioritization or responsiveness prioritization strategies. Instead, the author intends to introduce both traditional and innovative evaluation theories and approaches from these three strategies to enrich evaluators' toolbox so they can apply all theories and approaches as needed.

The nature and applications of the integrated evaluation perspective will be illustrated in detail in Chapters 11, 12, 13, 14, and 15, but its spirit and the principles it employs to develop indigenous concepts, theories, approaches, and methodologies are manifested throughout the book.

PROGRAM COMPLEXITY AND EVALUATION THEORIES

The discussion above of the dynamic and complicated nature of an intervention program in a community raises an interesting issue about program complexity. Evaluation theorists have different perceptions of how complex (e.g., in content, context, transformation, and stability) intervention programs are in general. Some may view these elements as quite stable or as fixed goals to achieve, whereas others view them as highly complex or fluid. How theorists view the complexities of a program can influence the theories or approaches they propose to use (Chen, 2012a).

To allow us to understand the issue, envisioning a continuum of program complexity is helpful, with reductionism at one end and fluid complexity at the other end. As discussed in the last section, reductionism postulates that a program is stable and can be analytically reduced to a few core elements.

On the other hand, fluid complexity, a concept created by the author for facilitating discussions, represents the view that an intervention program needs to constantly change its diverse and interactive elements to address ongoing environmental turbulences. Under the fluid complexity view, evaluators must speedily collect and analyze any available information on changes and promptly report the findings to decision makers to quickly adjust and readjust courses of action. For example, Christopher Columbus's expedition team not only had to constantly revise its plans for addressing ongoing external threats but also had to completely change its mission. Upon replacing the original mission of finding a route to India with the new mission of discovering a new world, the expedition was judged an enormous success.

Reductionism has its strengths and limitations. Reductionism has merits in its easy coexistence with known quantitative methodological and statistical models. Evaluators can use these methods and techniques to provide rigorous evidence of an intervention's efficacy or effectiveness. However, in its purest form, reductionism oversimplifies a program and provides an unsustainable solution. Fluid complexity also has its strengths and limitations. It may provide creative or sustainable solutions for complicated problems; however, at least for now, common quantitative methods and statistical models are not capable of effectively analyzing complex, fluid interactions. Furthermore, if a program is extremely complex and dynamic, then it lacks an entity for meaningful evaluation. Consultants are then more suitable than evaluators for offering opinions on how to assess and address problems.

Evaluation theories can be placed somewhere along the reductionist–fluid complexity continuum. Some are closer to reductionism; some, to fluid complexity. For example, the experimentation evaluation approach, which will be

discussed in Chapter 10, is closer to the reductionist end of the continuum. On the other hand, the developmental evaluation approach (Patton, 2011) is closer to the fluid complexity end of the continuum. The integrated evaluation perspective with its related theory-driven evaluation approach proposed in this book lies close to the middle. The perspective attempts to provide a synthesis of the different views proposed by reductionism and fluid complexity. It agrees with fluid complexity that the environment can create uncertainties and pressure stakeholders and evaluators to make changes, but it also proposes that a program can find proactive measures to reduce uncertainty and maintain a level of stability. For example, program managers and staff can build partnerships to buffer political pressure, and particular strategies, such as environmental scanning or problem-solving networks, can be helpful in reducing uncertainties. In applying the integrated evaluation perspective, this book proposes and examines many evaluation approaches and methods that consider variables and factors beyond what reductionism would recognize, but they are not too complex to be analyzed within existing quantitative and qualitative methods.

It is not clear where real-world programs fall along the continuum. However, it is clear that every theorist wishes to see or argue that the distribution congregates where his or her theory lies on the continuum. For example, this author would argue that real-world programs likely fall along the continuum in a normal (bell shape) distribution, with the majority at the middle. If this distribution proves true, the evaluation concepts, approaches, and methods proposed by this book are likely to be applicable to the majority of typical programs. Of course, other theorists would disagree. I encourage readers to form their own opinions and join the discussion.

WHO SHOULD READ THIS BOOK AND HOW THEY SHOULD USE IT

This book introduces practical evaluation approaches and methods to evaluators, but it avoids becoming a "cookbook." The approaches and methods discussed here are supported by a context of underlying principles and theoretical justification. This context, it is hoped, delivers knowledge with the latitude and flexibility program evaluators need to design suitable evaluation models. With this in mind, this book was prepared for two audiences.

Students

The first anticipated audience is students, especially those interested in issues related to the practice of evaluation, including the challenges evaluators

can expect and practical means of dealing with them. The book may liberate such students from the notion that evaluations are mainly methodological activities. Students should not feel like mindless number crunchers. The book will challenge students to seek strategies for broadening basic social science theories learned in the classroom, linking these to action and intervention theories employed in the field by program staff, evaluators, and social reformers. This text is ideal for use as a textbook for the following evaluation courses:

Introductory evaluation course. For an introductory course, the author would recommend covering Chapters 1 to 12.

Advanced evaluation course. Because of its depth and comprehensive scope, the book can also be used as one of the books in an advanced evaluation course. In such a course, instructors are encouraged to cover all the chapters from Chapters 1 to 15.

Evaluation Practitioners

The second audience is evaluation practitioners, especially those who seek new knowledge to strengthen their practical skills or expand the scope of their work. Such practitioners should generally look to the book to broaden their vision of evaluation alternatives so as to increase their skill at designing evaluations that fit a variety of program circumstances and evaluation purposes. Seasoned program evaluators may find both valuable insights into established evaluation strategies and approaches and new, innovative ideas for further enhancing their practice.

INTRODUCING THE REST OF THE CHAPTERS

In Chapter 2, based upon the fundamental typology of evaluation types discussed in this chapter, a road map of evaluation options—the "comprehensive evaluation typology"—is presented. The typology can guide evaluators and stakeholders in selecting the approaches and methods best suited to meet a program's circumstances and the stakeholders' needs at different program stages (program planning, initial implementation, mature implementation, and outcome), as discussed in Chapters 4 through 12. Chapter 3 discusses logic models and program theory, which are the foundation for understanding and

describing a program as discussed throughout the book. Chapters 13 to 15 discuss cutting-edge issues in program evaluation.

As will be discussed in Chapter 2, this book can be applied to start-up programs or established programs. For a start-up program, evaluators may be asked to evaluate one or more program stages, choosing among the planning, initial implementation, mature implementation, and outcome stages. For an established program, evaluators typically are invited to conduct evaluation activities at the mature implementation stage and/or the outcome stage.

QUESTIONS FOR REFLECTION

1. Detail a real-world intervention program and discuss its inputs, transformation, outputs, and environment.
2. Why is the feedback stage necessary to the success of an intervention program?
3. Define formative and summative evaluations. Give examples of each type.
4. Give examples of constructive process, conclusive process, constructive outcome, and conclusive outcome evaluation types.
5. Compare and contrast the dual formative/summative distinction with the fundamental evaluation typology.
6. What are hybrid evaluations? Give examples of this type of evaluation.
7. As the head of an agency or organization, how would you ensure that an internal evaluator provides useful information? How would you ensure that an external evaluator provides useful information?
8. Why are politics so important when planning to conduct an evaluation?
9. List some examples of potential stakeholders in an intervention program and explain why evaluators need to engage them when designing and conducting an evaluation.
10. Explain why the success of a program cannot be judged only by its results. Give examples.
11. Explain why research may be able to focus mainly on scientific credibility, while evaluation must have both scientific and stakeholder credibility.

12. Why do evaluators face a dilemma in addressing scientific and responsiveness requirements? What are three strategies to address the dilemma? Which strategy would you take? Why?

13. What is the integrated evaluation perspective? What are the challenges faced by this perspective?

14. The author argued that the distribution of real-world programs along the program complexity continuum may be like a normal bell-shaped distribution. He also mentioned that other theorists would disagree with him. Why? In your view, what would the distribution look like? Provide examples and justifications supporting your argument.