FUNDAMENTALS FOR PRACTICING PROGRAM EVALUATION

THE NATURE AND CHARACTERISTICS OF PROGRAMS

This book is all about program evaluation—the evaluation of “programs.” But what is meant by that term? How are the nature and characteristics of programs related to evaluation? The programs that evaluators can expect to evaluate have different names, such as treatment program, action program, or intervention program, and come from different substantive areas, such as health promotion, education, criminal justice, and welfare. Nevertheless, they all share the common feature of being organized efforts to enhance human well-being—whether by preventing disease, reducing poverty, or teaching skills. For convenience, programs and policies of any type are usually referred to in this book as “intervention programs,” or simply “programs.” Program evaluation is defined as the application of evaluation approaches, techniques, and knowledge to systematically assess and improve the planning, implementation, and effectiveness of programs.

The terminology of systems theory provides a means of illustrating, rather easily, the nature and characteristics of a given program. A program must perform two functions in order to succeed and survive. First, internally, it must ensure the smooth transformation of inputs into desirable outputs. For example, any education program would be in big trouble if faced with disruptions like
high staff turnover, excessive student absenteeism, or insufficient textbooks. Second, and externally, a program needs to continuously interact with its environment in order to obtain the resources and support necessary for its survival. That same education program would also become quite vulnerable if support from parents and school supervisors evaporated. Thus, because programs are subject to the influence of their environments, every program comprises an “open system.” Drawing further upon systems theory, any intervention program can be conceptualized as having the following five components (input, transformation, output, feedback, and environment), illustrated in Figure 1.1.

**Inputs.** Among the components of an intervention program are “inputs” from the environment. Inputs are resources taken in from the environment. These can include finances, 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.** The component called “transformation” represents the processes by which a program converts inputs into outputs. Transformation begins with the initial implementation of the treatment/intervention prescribed by a program. Transformation is, then, 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 matter in accordance with existing instructional rules and administrative
guidelines; this would represent transformation. Transformation also includes those sequential events necessary to achieve desirable outputs. For example, in order to increase students’ math and reading scores, an education program may need to first boost students’ motivation to learn.

**Outputs.** “Outputs” are the results of transformation. One crucial output is attainment of the program’s goals, which alone justifies the existence of the program. For example, an output of a treatment program for those engaging in spousal abuse is whether or not the abuse ends.

**Environment.** The “environment” refers to any factors that, despite lying outside a program’s boundaries, can nevertheless foster or constrain that program’s implementation. Such environmental 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 the way the general environment perceives program outputs. Are the outputs valuable? Are they acceptable? A day care program provides an example. If its staff is suspected of abusing children, the environment would find that output unacceptable. Parents would immediately remove their children from the program; the community might press criminal charges or at least boycott the program.

**Feedback.** To succeed—to correct any problems or adjust course effectively—an open system requires information about inputs and outputs, transformation, and the environment’s responses to these components. This information is called “feedback.” Feedback is what program evaluation is all about. Programs need information to gauge whether inputs are adequate and organized, interventions are implemented appropriately, target groups are reached, and clients receive quality services, and outputs demonstrate the attainment of goals and meeting of funding agencies’ and decision makers’ (e.g., stakeholders’) expectations. Without feedback, a system flies blind and is bound to deteriorate and eventually die. Without insightful program evaluation, programs fail. The action of feedback within the system is indicated by the dotted lines in Figure 1.1.

The five components of an open system can also be identified in any given “policy,” which is a concept closely related to a program. Although
policies may seem grander than programs—in terms of the envisioned magnitude of intervention, the number of people affected, and the legislation process—the principles and issues this book addresses are relevant to both. Throughout the rest of the book, in fact, the word *program* may be understood to mean program or policy.

**THE NATURE AND CHARACTERISTICS OF EFFECTIVE EVALUATION**

The transformation of inputs into outputs, and the interaction between program and environment, are nothing if not dynamic. Their fluctuations often make the evaluation of a program as challenging as it is necessary. The program evaluator can help ensure the quality of feedback about a program by making certain that a program evaluation is *future action-directed*, carries both scientific and stakeholder *credibility*, and takes a *holistic approach*.

**Future Action-Directedness**

One popular view of program evaluation is the assessment of the merits of a program. However, merit assessment is just one possible approach to program evaluation; often, program evaluation needs to go beyond merit assessment. When actually practicing an evaluation, all program evaluators quickly learn that stakeholders are eager to figure out what to do next. Stakeholders find evaluations useful if they both offer conclusions on how well programs have worked and provide information that assists the stakeholders in figuring out what must be done next to securely attain—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 those 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 actual determination of needed improvements. Just as, in the industrial and business world, information on product improvement is provided by research-and-development engineering and by market research, 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 those 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: Consumers reject the dissatisfactory product and choose a better quality alternative. However, the great majority of action programs do not encounter any market competition (Chen, 1990). Drug abusers in a community may find, for example, that there is only one treatment program available to them. In the absence of an alternative, the treatment program is likely to continue whether or not its outcomes justify to do so. Furthermore, well-known good-intention programs, 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, in order to be responsive and useful to stakeholders, program evaluation should meet both assessment needs and improvement needs rather than confine itself solely to merit assessment. Stakeholders need to know if the program is reaching the target group, if treatment/intervention is being implemented as directed, if staff are providing adequate services, if clients are making a commitment to the program, and if the environment seems to be helping or hindering the delivery of services. Whereas any part of this information can be difficult for stakeholders to collect, program evaluators have the necessary training and skills to gather and synthesize it all systematically.

Merit assessment remains an important approach or activity within program evaluation. In a broad sense, however, 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.
Scientific and Stakeholder Credibility

The scientific credibility of a 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 higher its credibility. However, scientific credibility is just the first kind of credibility that program evaluation must establish. As an applied science, program evaluation also exhibits varying degrees of stakeholder credibility. Stakeholders are those who have vested interests in the evaluation. 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. In general, the more an evaluation responds to stakeholders’ views, concerns, and needs, the higher its stakeholder credibility will be.

The ideal evaluation achieves both high scientific and high stakeholder credibility. And yet the two do not go automatically 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 simply as they see fit. Their evaluation is easily dismissed or rejected by stakeholders, despite scientific credibility, because it fails to reflect the stakeholders’ intentions for the program. For example, there are good reasons for African-Americans to be skeptical of scientific experiments without community inputs, due to incidents such as the Tuskegee syphilis experiment (Jones, 1981). Researchers in the experiment withheld effective treatment from African-American men suffering from syphilis so that 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, ending in evaluation results that are suspect.

The challenge in program evaluation is to achieve a balance of both kinds of credibility. A helpful strategy is to pursue stakeholder credibility in the earliest phases of evaluation design but to yield to scientific principles later in the process (Chen, 1990). 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 clear 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 very beginning of the process (Perry & Backus, 1995). 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).

### Holistic Approach

Many disciplines or areas of inquiry have an evaluative ingredient. Some disciplines gauge consumer satisfaction with products (product evaluation). Some assess an employed individual’s performance in a company or organization (personnel evaluation). Some weigh a person’s skills or qualifications (college entrance examinations). Some assess the effects of new drugs (biomedical experimentation). Some judge performance in a sport (competitive figure skating or diving). Some assess the appropriateness of the handling of money (public accounting). The nature of a program, though, as discussed earlier in this chapter, suggests that programs are clearly distinct from the objects of these other disciplines, and thus both the nature of a program evaluation and its associated principles and strategies belong to it alone.

One important characteristic distinguishing program evaluation is its need, rarely shared by other disciplines, to use a holistic approach to assessment. The holistic approach to assessment finds it imperative to include contextual or transformation information when assessing the merit of a program. By comparison, product evaluation is more streamlined: It may get away with an exclusive focus on the intrinsic value of its object. Products like televisions can be assessed in terms of intrinsics such as picture, sound, durability, price, and so on. In many situations, however, the value of a program may be ecological 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. Take as an example the case of an educational program that, according to strictly performance-based evaluation, has attained its goals (which are its intrinsic values). But in what context was the performance achieved? Perhaps goals were attained by “creaming,” which is the deliberate admission of only those students expected to meet with success
in the program and the accompanying rejection of students not so clearly
destined to succeed but perhaps more in need of the program. Does the
program’s performance still deserve loud applause? Probably not.

Similarly, what about a case in which program success is due to partici-
pation 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. In program evaluation, how a program achieved its
goals is 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 it was misleading to make such a claim.
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, actually ordered soldiers to seize men on the streets and take them to
be sterilized. An evaluator with an approach that was less than holistic might
have declared that the goals of the program were attained, whereas all around
were people whose personal knowledge led them to condemn the program as
inhumane. Lacking a holistic orientation, program evaluation may reach very
misleading conclusions.

THE UNIVERSALIST VIEW OF THE
PRACTICE OF EVALUATION (AND ITS LIMITATIONS)

The approaches to and methods of evaluation that have been developed within
program evaluation are plentiful (Rossi, Lipsey, & Freeman, 2004; Shadish
et al., 1991). An important issue for evaluation practitioners is how to select
an approach or method for their evaluation. There are two general points of
views about making this determination: the universalist and the contingency.
In the following section, the universalist view is discussed and its limitations
reviewed, providing the basis for the next section’s discussion of the contin-
gency view.
The universalist view insists on the universal superiority of certain evaluation approaches and methods over others; in light of this precept, evaluators must strive to apply the endorsed “best” evaluation approaches and methods in undertaking any evaluation. The research methods involved in program evaluation provide an example. The universalist view argues that there is always a “best” research method. For instance, some evaluators advocate quantitative methods (such as randomized experiments) for use in program evaluation, whereas others insist that qualitative methods, or mixed methods, are best. Whatever his or her choice, the evaluator who advocates one best method takes the universalist view of program evaluation research methods. Formative versus summative evaluation provides a similar example. Evaluators who argue that summative evaluation is better than formative evaluation—as well as those who argue the opposite—demonstrate universalist principles for performing evaluation. One reason for the popularity of the universalist view (notably during program evaluation’s infancy) is that universalist principles appear straightforward, forceful, and easy for practitioners to follow. Another reason is the convenience the universalist view provides to evaluation theorists seeking to frame their arguments, highlight advantages of their approaches, and attract attention.

In reality, though, particular programs vary considerably with regard to structure (e.g., fluid vs. routine, large scale vs. small scale) as well as to expressed evaluation needs, availability of pertinent data, amount of funding earmarked for evaluation, and other realities. Given the plethora of such variables, despite the universalist view, it seems questionable at the very least whether there is one, best evaluation approach or research method covering every evaluation. The application of the universalist view and ensuing principles makes evaluators vulnerable to the failures of an indiscriminate, one-size-fits-all type of evaluation.

THE CONTINGENCY VIEW AND THE PROGRAM THEORY PERSPECTIVE

An alternative point of view from which to conceptualize, develop, and apply program evaluation principles, approaches, and methods is the contingency view, and it postulates that there is no single best way to conduct program evaluation; rather, the choice of approaches and methods for program evaluation should be situational. The basic principle of the contingency view is that the
individual natures of programs and the uniqueness of evaluation purposes and contextual circumstances require use of a range of evaluation approaches and methods. For example, taking a contingency view, Chen (1996) specified the conditions (e.g., depth of information, credibility of data, breadth of program boundaries) under which quantitative, qualitative, and mixed methods will prove most serviceable to program evaluators. From the contingency viewpoint, guidelines for evaluation may be less straightforward than under the universalist view, but this creates the advantage of avoiding simplistic assumptions. In general, the contingency view appears to fit better with the reality in which practitioners currently conduct even state-of-the-art program evaluations.

One key evaluation perspective incorporating the contingency viewpoint in its principles and strategies is the program theory evaluation perspective, often called simply the program theory perspective (Chen, 1990, 1994, 1996, 1997). The program theory perspective has benefited greatly from the intensively discussed topics comprised by theory-driven or theory-based evaluation (e.g., Bickman, 1987a, 1987b, 1990; Chen & Rossi, 1980, 1992; Connell et al., 1995; Donaldson, 2003; Fulbright-Anderson et al., 1998; Pawson & Tilly, 1997; Rogers et al., 2000; Suchman, 1967, 1969; Weiss, 1997, 1998; Wholey, 1987). The program theory perspective can help guide practitioners or students toward an understanding of the circumstances in which a particular set of evaluation approaches and methods is appropriate for evaluating a particular program.

The program theory perspective calls the practitioner’s attention to the following major factors influencing selection of evaluation approaches and methods: (a) Which stage or stages of the program cycle will be the focus of the evaluation? (b) What do stakeholders want from the evaluation—assessment-oriented information, improvement-oriented information, or both? (c) What evaluation options potentially fit the given program’s environmental and other circumstances, as well as stakeholders’ needs? (d) What trade-offs among these options will be most profitable? The program theory perspective provides a conceptual framework that practitioners can use to find effective answers to these four questions. Detailing this conceptual framework is one of two major purposes of this book. The second purpose is to provide systematic advice on how to fruitfully conduct a program evaluation once an evaluation approach is identified. The program theory perspective and its conceptual framework are discussed at length in Chapter 2.
AUDIENCES AND USES OF THIS BOOK

This book was prepared for two audiences. The first 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, enhancing their skills for designing evaluations fitting a variety of program circumstances and evaluation purposes. Seasoned program evaluators may find in this book both valuable insights into established evaluation strategies and approaches, and new ideas for practice. The second anticipated audience is students, possibly students who completed “Evaluation 101,” but certainly those interested in issues of the actual practice of evaluation, including difficulties 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. This should help prevent evaluation students from feeling like mindless number crunchers. The book could even challenge them 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.2

Chapter 2 discusses the concepts and conceptual framework of the program theory perspective; these provide the foundation for the materials in the rest of the book. In Chapter 3, a road map of evaluation options—the “evaluation taxonomy”—is presented. The evaluation taxonomy can guide evaluators and stakeholders in selecting the approaches and methods best suited to 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 10.

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.

The theoretical and methodological roots of this book could be traced back to my book, Theory-Driven Evaluations (Chen, 1990). Readers who are interested in program theory and theory-driven evaluation are encouraged to read this book.
NOTES

1. It is important to note that, although some components identified in systems theory bear the same labels as components found in the logic model, the definitions of each are quite distinct.

2. Chapter 3 includes a more intensive discussion of the concepts of these program stages.