A PRACTICAL EVALUATION TAXONOMY

Selecting the Evaluation Approach That Works

It is vital that practicing program evaluators know which evaluation strategy and which evaluation approach, out of the many available, will be best suited to meet stakeholders’ fluctuating needs. The “fish story” that opens the next section illustrates the value of this knowledge. The story is followed later in the chapter by a taxonomy of practical program evaluation means and ends: comprehensive, systematic guidance to use while weighing the circumstances and needs of one’s evaluation assignment against the strengths and shortcomings of various evaluation strategies and approaches.

THE ART OF FISHING?
THE ART OF PROGRAM EVALUATION?

Whereas discussions of individual evaluation approaches and methods are encountered frequently in the literature (e.g., Mark, Henry, & Julnes, 2000; Rossi, Lipsey, & Freeman, 2004; Shadish, Cook, & Leviton, 1991), there is much less written about issues that affect the profitable selection of one
approach or strategy from among the many. The information that is available tends not to be systematically presented. An analogy of fishing suggests how this gap in program evaluation can make life more difficult for the program evaluator. To go fishing one needs, first of all, equipment—the poles and lines, hooks, sinkers, floaters, bait. Without this basic equipment, fishing is (for most humans) impossible. Possessing equipment and knowing how to employ it, however, do not guarantee success. Choosing the wrong equipment from all that is available—the wrong size fishing line, or the wrong hook, or an inappropriate bait for a given fishing spot—probably means turning up empty-handed, even if one handles that line, hook, or bait magnificently.

The vital, and yet limited, role of equipment in the art of fishing is seen clearly in fly fishing, the technique in which the angler continuously casts and retrieves a line tipped with, or baited with, an artificial fly. With tackle and casting know-how, any person can go through the motions of fly fishing. Those who can be counted on to catch fish, however, are those accomplished in the art. A good fly-fishing angler knows how to choose the right place and time as well as the right artificial fly. Fly fishing masters have learned to habitually consider such things as season, currents, play of light and shade, and types of surrounding vegetation, in addition to their equipment. These masters can select just the right fly to mimic whatever real fly would inhabit a given area at a given moment, so that the fish pursue the fly without suspicion.

Productive fishing is more than equipment and the ability to operate it. In order to catch fish, it helps to know fish habits and habitat: favorite foods, favorite pools or banks, responses to weather, and so on. Upon this kind of understanding is based “the art of fishing.” Mastery of this art may result from trial and error over a long period of time or, more efficiently, from instruction by someone experienced in fishing. It may also be developed by studying authoritative books.

What can program evaluators learn from the art of fishing? To begin with, consider evaluation approaches and methods as being analogous to fishing equipment. Only when we have our evaluation approaches and methods down pat can we set about catching a fish. But although this is a necessary condition, it remains an insufficient one. Like an average fishing enthusiast with a tackle box and some manufacturer’s instructions, an evaluator familiar with evaluation approaches and methods can try his or her luck. But there is no more guarantee of a decent evaluation than of landing a catch. Like fishing the wrong spot or using the wrong bait, missing important issues in evaluation design finds
the adventurer returning home with no prize. Evaluation becomes productive only when we go beyond methodology to ply the waters with theoretical and also with contextual knowledge.

THE EVALUATION TAXONOMY AS A SOURCE FOR THE ART ASPECT OF PROGRAM EVALUATION

Practical evaluators must be able to apply evaluation approaches correctly—the scientific aspect of program evaluation. But before they can do so, they must be able to select the evaluation approach that complements the needs and realities they face—the art aspect of program evaluation. The evaluation taxonomy around which this book is organized is offered as a means of enhancing understanding of evaluation’s art aspect. A taxonomy provides any discipline with an important tool for classifying phenomena, guiding the selection of research strategy, facilitating communication, and developing knowledge. In program evaluation, a taxonomy can help with conceptualization of evaluation needs, focusing of evaluation activities, and identification of available evaluation means that suit a program’s needs and realities.

An early example of classifying evaluation activities is the distinction of formative from summative evaluation proposed by Scriven (1967). Formative evaluation is concerned with designing and using evaluation to improve a program, whereas summative evaluation is concerned with designing and using evaluation to judge a program’s merit (Shadish et al., 1991). Chen (1996, 2004) proposed a taxonomy of program evaluation, one built around the program stage that is the desired focus of the evaluation, as well as around the desired function of the evaluation (either improvement or assessment). This section of the book furthers this earlier work on the notion of a taxonomy. It introduces a systematic taxonomy for daily use by program evaluators in the field. This practical taxonomy of program evaluation ends and means is represented in Table 3.1.

At left in the table are several “classes” of potential evaluation needs of stakeholders at each stage of program growth. These classes are linked to numerous evaluation approaches and strategies on the right. The remainder of this chapter fills out the table, with discussion of stakeholders’ needs and subsequent discussion of evaluation approaches and strategies available to meet them.
## Table 3.1 A Practical Taxonomy for Program Evaluation Means and Ends

<table>
<thead>
<tr>
<th>Program Stages and Evaluation Purposes</th>
<th>Evaluation Strategies</th>
<th>Evaluation Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Program Planning Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providing pertinent information and assistance to help stakeholders in developing program rationale and plan</td>
<td>Background information provision</td>
<td>Needs assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Formative research</td>
</tr>
<tr>
<td></td>
<td>Development facilitation</td>
<td>Conceptualization facilitation</td>
</tr>
<tr>
<td></td>
<td>Troubleshooting</td>
<td>Relevancy testing</td>
</tr>
<tr>
<td></td>
<td>Development partnership</td>
<td>Pilot testing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commentary or advisory</td>
</tr>
<tr>
<td><strong>Initial Implementation Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providing timely information on implementation problems and the sources for assisting stakeholders to fix the problem and stabilize the program</td>
<td>Troubleshooting</td>
<td>Formative evaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Program review/ development meeting</td>
</tr>
<tr>
<td></td>
<td>Development partnership</td>
<td>Bilateral empowerment evaluation</td>
</tr>
<tr>
<td><strong>Mature Implementation Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Providing information on implementation problems and the sources for improving the implementation process</td>
<td>Troubleshooting</td>
<td>Formative evaluation</td>
</tr>
<tr>
<td></td>
<td>Development facilitation</td>
<td>Program review/ development meeting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conceptualization facilitation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concept mapping</td>
</tr>
<tr>
<td>Assessing the quality of implementation for serving accountability needs</td>
<td>Performance assessment</td>
<td>Fidelity evaluation</td>
</tr>
<tr>
<td>Monitoring progress of implementation</td>
<td>Performance monitoring</td>
<td>Process monitoring</td>
</tr>
<tr>
<td>Holistic assessment of implementation process</td>
<td>Enlightenment assessment</td>
<td>Theory-driven process evaluation</td>
</tr>
<tr>
<td><strong>Outcome Stage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessing whether program is ready for outcome evaluation</td>
<td>Development facilitation</td>
<td>Evaluability assessment</td>
</tr>
<tr>
<td>Monitoring client’s progress of outcomes</td>
<td>Performance monitoring</td>
<td>Outcome monitoring</td>
</tr>
<tr>
<td>Assessing the effect on outcomes</td>
<td>Performance assessment</td>
<td>Efficacy evaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Effectiveness evaluation</td>
</tr>
<tr>
<td>Holistic assessment of the program effect for serving accountability and program improvement needs</td>
<td>Enlightenment assessment</td>
<td>Theory-driven outcome evaluation</td>
</tr>
</tbody>
</table>
A PRACTICAL TAXONOMY OF PROGRAM EVALUATION MEANS AND ENDS

Classifying Stakeholder Needs at Particular Program Stages

The first row of Table 3.1 (illustrating the practical taxonomy of program evaluation means and ends) lays out four stages of a program’s growth and the nature of each stage. Evaluation requirements associated with each stage are highlighted. Stakeholders’ evaluation needs vary across the stages of program growth (Chen, 2004). Evaluators can best understand stakeholders’ evaluation needs if the evaluators are provided with information on the stage(s) the stakeholders are interested in evaluating. There are four program stages in the taxonomy: planning, initial implementation, mature implementation, and outcome.

It is often assumed that a program will move sequentially through these stages. This makes sense, but, in reality, programs could go back and forth between stages in a nonlinear fashion. As stages skip around, evaluation needs do, too. As an example, consider a program in its mature implementation stage. The program has been troubled by several major problems with service delivery, and its stakeholders decide to revise the program plan and return to the initial implementation stage. Thus, their evaluation needs are different from what they once were. Similarly, if, in its outcome stage, a program is found to be ineffective, its stakeholders could move to redesign the program, returning to the initial planning stage.

Evaluators in the field are asked to conduct evaluations at any stage and for various combinations of stages. When the program is an established one, evaluation of its implementation and outcome stages is common; start-up programs, too, need evaluations at these stages. Start-up programs also frequently require evaluators during the planning stage and initial implementation stage. In the following paragraphs, the evaluation needs characteristic of each stage are discussed.

Program Planning Stage. The first of the four stages is the program planning stage. This is the very beginning. Stakeholders at this stage—for example, program designers—are developing a plan that will serve as a foundation for organizing and implementing a program at some future date. As we have seen, programs can be complex; stakeholders often seek considerable help from experts with the hope of winding up with a plan of truly good quality. Today, evaluators are often found among these experts. In the program planning stage, stakeholders’ primary evaluation need is to learn from evaluators the
evaluation concepts, strategies, and activities that can help in the design and development of a program rationale and a program plan.

Early in Part II of this book, we read that program evaluation has, across much of its history, focused on outcomes. Lessons from the field, however, have plainly taught that program failures are often essentially implementation failures, and evaluation focus has gradually broadened to include process evaluation. The current view is that a major part of implementation failure can be traced to poor program planning and development. Evaluators can make important contributions in these areas where attention is most needed.

Initial Implementation Stage. The second stage cited in Table 3.1 is the initial implementation stage. As a program plan begins to be put into action, much can go wrong. During the initial implementation stage, a program’s course can be highly fluid and unstable. At this point in development, stakeholders’ evaluation needs include timely feedback on the major implementation matters and also identification of the sources of problems. These kinds of data can help the stakeholders to troubleshoot the implementation problems and to quickly stabilize the program.

Mature Implementation Stage. The mature implementation stage follows the initial implementation stage at a point when implementation of the program has settled into routine activities. Rules and procedures for conducting program activities are now well established. Stakeholders are likely to be interested in one or more of the following: continued unearthing of the sources of immediate problems, generation of data reassuring to those to whom stakeholders are accountable, and program improvement. Even in maturity, a program is subject to problems such as clients’ dissatisfaction with services. A wise course for stakeholders in a case like this is to seek timely information from evaluators on the cause of problems. Identifying and resolving problems are key to improving a program. Furthermore, as a program matures, stakeholders may think more about their accountability. Data illustrating the effectiveness of implementation, or the efficiency of service delivery, are useful to stakeholders, who often ask evaluators to find such data if they exist. Finally, within the mature implementation stage, stakeholders begin to look for strategies of improvement (tying in to their need to be accountable, perhaps). They call on evaluators to provide information from a holistic standpoint through process evaluation that goes beyond assessing the quality of implementation to strengthening the program processes.
Outcome Stage. The fourth stage of program growth is known as the outcome stage. Following a period of program maturity, stakeholders inside and outside the program want to know whether the program is achieving its goals. An evaluation at this point can serve any of five primary evaluation needs. First, stakeholders may rely on evaluators to determine if a program is ready for outcome evaluation. It may not be and, if it is not, evaluators may be asked for help in building the program’s “evaluation capacity.” Second, stakeholders may want to monitor their clients’ progress. Third, stakeholders may ask for information on what the program would be achieving if it existed in the ideal environment. (Such information can also help stakeholders decide if a program should be expanded to other people or settings.) Fourth, stakeholders may seek to know in detail the program’s effects in its real-world setting because these, obviously, have a direct bearing on practice. Finally, some stakeholders may ask evaluators to go beyond traditional evaluation and its single-minded focus on assessment. They may want an evaluation that serves both accountability and program improvement needs.

Dynamics of Transition Across Program Stages

Intervention programs are goal-oriented activities. Ideally, the program moves directionally through the following stages: program planning; initial implementation; mature implementation; and, finally, outcome. For the convenience of the reader, the remaining chapters of this book are arranged according to an ideal sequence of program stages. In reality, however, the program stages may not be linear at all. For example, due to a stakeholder’s dissatisfaction with the direction of a program, or due to political pressure calling for a change, a program could move from the mature implementation stage back to the planning stage rather than moving forward to the outcome stage. Similarly, a program at the outcome stage may make an overhaul of its operational procedures and move back to initial implementation stage. This book accommodates the nonlinear transition of program stages and the evaluations related to each stage. Readers can pick and choose from the chapters and rearrange the sequence of program stages to fit their programs.

Classes of Strategies and Approaches
Serving Stakeholders’ Various Needs

An evaluation strategy is a general direction or orientation the evaluator and stakeholders take in order to fulfill a given evaluation’s purpose. For
example, merit assessment is one general strategy employed to serve stakeholders’ accountability needs. Other evaluation strategies include the development and enlightened strategies. The overall evaluation strategy must be closely related to the stakeholders’ evaluation needs. In contrast, the evaluation approach constitutes a systematic set of concrete procedures and principles to guide the designing and conducting of an evaluation. The evaluation approach determines the evaluation’s focus; it affects the research methods applied to collect and analyze data, as well as the interpretation of data. Typically, several evaluation approaches are harnessed together within one evaluation strategy (see Table 3.1). Well-known evaluation approaches include experimental and quasi-experimental approaches, monitoring, needs assessment, and qualitative evaluation.

Whereas most stakeholders are unfamiliar with individual research procedures and techniques (familiarity with which is assumed of evaluators), they are usually acquainted with the general directions that evaluation strategies comprise. For example, stakeholders may not know what a quasi-experiment is, but they do understand generally what a strategy such as merit assessment entails. The easiest and best course for evaluators is to determine the appropriate program evaluation approach by communicating with stakeholders about the strategies the evaluators think will fit the stakeholders’ evaluation needs. With their input in mind, the evaluators can then lead a discussion of various appropriate evaluation approaches. Too often, evaluators ignore the dialogue on evaluation strategies, launching right into selection of evaluation approaches. The fallout from this practice can be stakeholders’ uninformed consent to employ whatever evaluation approach the evaluator recommends. With little or no understanding of what that approach consists of, stakeholders may, when handed the final report of the evaluation, realize that it is not what they wanted and does not provide the information they need.

My view on evaluation strategies and approaches has been greatly influenced by my acquaintance with the following case of misdirected evaluation. The client was a group of high-performing community-based organizations seeking to provide capacity-building services to similar but less accomplished organizations. A skilled and respected evaluator carried out the project. Before beginning, this evaluator met with stakeholders several times to discuss potential evaluation approaches. The parties decided to adopt mixed methods to assess the merit of the capacity-building program. The final evaluation report provided a detailed pros-and-cons assessment of the program, expressing in general that the evaluation favored the project. Unfortunately, those anticipated
to be the program’s service providers complained that the evaluation offered few insights into improving their program. The generated information was of the wrong kind, they protested, because the evaluation failed to reflect their needs and views. In the end, it became clear that the service providers had wanted a development-oriented evaluation, whereas the evaluator had conducted a merit-assessment evaluation.

This is not a case easily dismissed by blaming the service providers for misstating their evaluation needs at the beginning, or for changing their minds later on. In reviewing the project with the evaluator and the service providers, it appears that the heart of this problem was the absence of effective tools with which to voice actual evaluation needs and identify suitable accompanying evaluation approaches.

A superabundance of evaluation strategies and approaches is cited in the literature (e.g., Mark et al., 2000; Rossi et al., 2004; Shadish et al., 1991; Stufflebeam, 2001). This practical taxonomy (Table 3.1) does not include every strategy and approach ever conceived. Evaluation strategies and approaches discussed frequently in the evaluation literature are not necessarily those used frequently by evaluation practitioners, and vice versa. In contrast, the practical taxonomy of program evaluation means and ends includes only strategies and approaches with marked potential to be used successfully in the field regardless of its popularity in the existing literature. Readers for whom the terminology in Table 3.1 is new should note that subsequent chapters of the book explore and explain the strategies and approaches in detail. The bird’s-eye view provided in this chapter is simply a foundation.

To begin, there are four general categories of evaluation strategy included in the practical taxonomy of program evaluation means and ends: merit assessment, development, enlightenment, and partnership.

**Merit Assessment Strategies**

Merit assessment strategies are those that can provide information on the performance or merit of a program. Two merit assessment strategies frequently used by evaluators are the performance assessment strategy and the performance monitoring strategy. (Only a brief introduction to these follows here; Chapters 8 and 9, however, discuss at length the differences and relationships between performance evaluation and performance monitoring.) Performance assessment is the employment of rigorous designs to provide credible information about a program’s merit in terms of either its implementation process or its
outcomes. The performance assessment strategy is part of a long-standing, influential tradition in program evaluation, one discussed in Mark et al. (2000) and Shadish et al. (1991). It is often more expensive and time-consuming to employ the performance assessment strategy than other comparable evaluation strategies. The most common application of performance assessment strategy is with programs in the mature implementation stage or outcome stage. Performance assessment strategy is commonly affiliated with the use of a popular evaluation approach called \textit{fidelity evaluation}. The fidelity evaluation approach assesses whether a program has been or is being implemented according to expectations. Another approach popularly used with the performance assessment strategy is the \textit{outcome evaluation}, which assesses a program’s success in reaching its goals. The outcome evaluation approach is the major tool within traditional outcome evaluation (e.g., Cook & Campbell, 1979).

The \textit{performance monitoring strategy}, the second assessment strategy included in the practical taxonomy, involves using indicators to follow the implementation process and outcomes of a program across time. In a drug treatment program, for example, evaluators might monitor clients’ drug use both before and after they experience the intervention. The performance monitoring strategy comprises two well-known approaches: \textit{process monitoring} and \textit{outcome monitoring}. Process monitoring cannot produce in-depth information about a program’s implementation as process evaluation does; neither is outcome monitoring likely to produce convincing data about an intervention’s effect on outcomes the way outcome evaluation can. In their defense, however, process monitoring and outcome monitoring are useful for managing a program and likely to cost less than typical process evaluation and outcome evaluation.

\textbf{Development Strategies}

Development strategies collect evaluative data relatively quickly in order to assist stakeholders with program planning or development. Three development strategies are well established in program evaluation: the \textit{background information provision strategy}, the \textit{troubleshooting strategy}, and the \textit{development facilitation strategy}.

\textit{Background Information Provision Strategy}. To use the background information provision strategy is to research the background of a program in terms of community characteristics and needs, target population characteristics, and/or intervention options. The information gathered should help program designers
and other stakeholders plan or strengthen a program. Evaluation approaches suited to this strategy include needs assessment and formative research. Determining and prioritizing the needs of a community or target population that warrant intervention is called needs assessment, such as when an agency asks what kinds of youth services are most needed in a community. In such a case, program evaluators might systematically interview youths, parents, and community leaders to help the agency answer its question. Formative research differs from needs assessment in the greater emphasis it places on an identification or prioritization of needs. Formative research consists of gathering empirical information on community and target population characteristics, as well as intervention options, in order to help stakeholders plan and develop programs. For example, program designers uncertain about what kind of drug prevention program would be best received by new immigrants might engage program evaluators to manage a survey or focus group meeting, obtaining information that enables the program designers to make a decision.

Troubleshooting Strategy. The troubleshooting strategy is a system for identifying trouble spots in programs and addressing them. The troubleshooting strategy is used, first of all, to provide timely assessment of barriers and/or problems facing a program; its second use concerns options available to stakeholders to address difficulties. The value of the strategy lies in its ability (not always guaranteed) to effectively identify any implementation problem before it gets away from stakeholders and major damage occurs. Evaluators using this strategy must also provide stakeholders with information that facilitates resolution of the problem. The troubleshooting strategy is associated with use of the formative evaluation, relevancy testing, piloting, and commentary and advisory approaches. Formative evaluation is associated with research methods that are flexible to use; are easy to adopt in the field; and have a short turnaround time, such as focus groups and participant observation to collect, in timely fashion, facts about barriers and problems in implementation that promise to strengthen it. Having chosen to target newly arrived immigrants, for instance, an HIV prevention program further decides to serve them with group counseling. After the implementation is carried out, evaluators are contracted to look for potential problems in the recently completed process. Using formative evaluation, evaluators interview a sample of the clients and quickly learn that some clients—Asian immigrants—are uncomfortable in group discussions of sexual behavior. The quick feedback made available to the program director by the formative
evaluation approach prompted modification of the program to better serve this particular immigrant group.

Formative evaluation and formative research (an approach affiliated with the background information provision strategy) are both research activities, yet with an important difference: Whereas formative evaluation examines directly the program’s implementation, formative research is usually carried out before implementation and produces background information related to program planning. For example, the evaluator tackling the above assignment from a formative research approach might study a target group’s cultural background as it relates to sexual behavior in hopes of facilitating program design decisions. An evaluator using the formative evaluation approach would evaluate the given target population’s experience with the program itself.

The troubleshooting strategy also includes relevancy testing. Relevancy testing is the small-scale assessment of causal assumptions underlying a program and whether these assumptions hold up in the field. Relevancy testing can be used to strengthen the soundness of a change model. In contrast, the pilot-testing approach to the troubleshooting strategy involves actually operating the program on a very small scale. Unlike relevancy testing, pilot testing usually focuses on the action model. The information and experience gained from pilot testing can help strengthen a program before formal implementation begins because areas needing modification can be fixed early and prevented from affecting the full-scale implementation. Another troubleshooting strategy is the commentary and advisory approach, which is an approach that does not collect data from the field. Instead, the expertise of evaluators is tapped as they review and comment on an existing action model and change model. They advise stakeholders about probable strengths and weaknesses of the models and offer suggestions for improvement. Finally, the program review/development meeting approach generates insights through systematic discussions in a meeting format among a group of program implementers and staff. With the evaluator providing facilitation, the experiences of a program are discussed, any implementation problems are dissected, barriers to and facilitators of these problems are identified, and strategies are created to strengthen the program.

*Development Facilitation Strategy.* Evaluators’ knowledge and skills are also central to the development facilitation strategy, which is defined in this book as the use of such expertise to help key stakeholders in a meeting/workshop setting. The development facilitation strategy functions to facilitate the stakeholders’ efforts to develop or fine-tune the logic of a program, or to identify its problems
and seek programmatic solutions for them. Using this strategy, evaluators become facilitators and consultants, essentially; Guba and Lincoln (1989) and Patton (1997) have emphasized the value of this method for solidifying a common vision, winning support, and broadening a program’s capacity.

Expert evaluators can draw on their program evaluation skills to contribute greatly to the development of coherent programs that are logical in their foundations and feasible to implement.

Some evaluation approaches associated with the development facilitation strategy are the conceptualization facilitation approach and the concept-mapping approach. The conceptualization facilitation approach requires evaluators to work as facilitators and consultants, clarifying stakeholders’ ideas, especially those concerning action and change models—and then, quite likely, facilitating their effort to develop these models. Under the concept-mapping approach, quantitative methods provide help to stakeholders sorting out the ends and means of a program (Trochim & Cook, 1992).

Enlightenment Strategy

Stakeholders may, of course, seek program evaluation in response to accountability, as well as program improvement, needs. As discussed in Chapter 1, program improvement remains the ultimate goal of program evaluation, and pure performance evaluation has little to say about improving programs; yet it is possible to design evaluations to meet both kinds of needs. The key is extending the evaluation beyond merit assessment by examining underlying assumptions and mechanisms that mediate the effects of the program. Evaluators with this orientation are practicing the enlightenment strategy. Enlightenment strategy is discussed at length in the literature (e.g., Mark et al., 2000), and, in general, it takes the position “Assessment is means, program improvement is end.” The work by Cronbach (1982), Chen (1990), Chen and Rossi (1992), and Shadish, Cook, and Levinton (1991) is related to enlightenment strategy. An evaluation approach that suits this strategy is called the theory-driven (or theory-based) evaluation approach. Theory-driven evaluation has been applied to assess implementation processes as well as program outcomes (Chen, 1990).

Partnership Strategy

The final strategy presented in this section of the book is the partnership strategy, in which stakeholders invite evaluators to be partners in planning and
implementing programs. The parties work together closely at every step, with evaluation information introduced regularly to support their effort to develop and implement a program. This strategy, and the \textit{bilateral empowerment approach} that accompanies it, comprise something of a challenge to the traditional foci of evaluation. Bilateral empowerment means that the participating evaluators are granted membership on the development team. Accordingly, they have direct input as to the handling of development and evaluation issues; that is, evaluators participate in the decision-making process. Bilateral empowerment may work best with programs that have vague notions about goals, interventions, and implementation. This strategy and this approach have gained momentum in the recent literature concerning community coalition evaluation (e.g., Goodman, Wandersman, Chinman, Imm, & Morrissey, 1996).

\textbf{STEPS TO TAKE IN APPLYING THE PRACTICAL TAXONOMY}

The purpose of the practical taxonomy of evaluation ends and means is to associate particular evaluation strategies and approaches with particular program stages and stakeholder needs (see Table 3.1). The stakeholders of a program in its initial implementation stage, for example, need an evaluation strategy and approach that move quickly to tackle immediate implementation problems. The taxonomy demonstrates a very clear bearing: that \textit{program evaluation is situational}. No single evaluation strategy, approach, or method can succeed with every possible evaluation need or situation. Means of evaluation that are fruitful in one case may be fruitless—or even misleading—in others. The performance assessment strategy, for instance, although plainly useful when the need is for accountability of a program in its mature implementation stage, could produce questionable results if employed with an immature program. This is because the only input stakeholders can actually use early on is timely information that helps to stabilize early implementation.

The practical taxonomy as it appears in Table 3.1 was crafted as a “map” of the art of evaluation for evaluators and stakeholders to review together. An evaluator might want to proceed through the taxonomy (from left to right) with stakeholders, identifying the evaluation strategies and approaches best suited to the evaluation they seek. Taking the following steps \textit{in sequence} should bring the evaluator to the finish line in good shape.
1. Identify the program stage that is of interest. Evaluation needs are expressed by stakeholders in general, abstract terms. The evaluator must create precision in the discussion by facilitating a choice about exactly which program stage(s) should be the focus of investigation. When stakeholders request evaluation of a program implementation, they must decide if they mean its initial implementation or its mature implementation, because the two are not identical. Lack of expressed stage-specific needs, understood both by stakeholders and evaluators, can end in the choosing of mismatched strategies and approaches, producing a useless evaluation. Stakeholders cannot be blamed for misunderstandings about evaluation needs because it is the evaluator’s responsibility to thoroughly grasp stakeholders’ intentions before designing an evaluation. Information obtained in the course of articulating or clarifying stakeholders’ needs will advance the effort to select the best evaluation strategy and approach for the task.

2. Choose an evaluation strategy that matches stakeholders’ internal/external purposes. Having settled the issue of program stage, the evaluator must quiz stakeholders about the eventual audience for the evaluative information. Does it have an internal purpose, external purpose, or both? This is crucial when selecting an evaluation strategy. In general, if the information mainly will be used internally to find and fix implementation problems, then the development facilitation strategy is a good choice. For example, stakeholders desiring to troubleshoot their programs will find formative evaluation to be valuable. For an audience beyond the internal, however, evaluators and stakeholders might use an assessment strategy because assessment strategies provide much information that satisfies accountability requirements. A performance assessment strategy used at the outcome stage, for example, can be used to rigorously assess the effects of a program. But should the stakeholders need evaluative information that serves program improvement needs as well as accountability needs, then the enlightenment strategy is the best choice.

3. Choose evaluation approaches and research methods that provide acceptable trade-offs. With the strategy question answered, it is time to choose an appropriate, stage-specific evaluation approach (or approaches). Each strategy included in the practical taxonomy is linked to one or more evaluation approaches, and each of those is, in turn, affiliated with a number of research methods. All of these options demonstrate strengths and weaknesses in terms of the basic qualities of evaluations: timeliness, rigor, thoroughness, and cost.
Stakeholders must be willing to make trade-offs among these qualities, with adequate understanding of pros and cons. Two examples of trade-offs are illustrative. First is the forming of an acceptable compromise concerning the timeliness, rigor, and cost of evaluation. We accept that there is a tendency among evaluators (or at least a desire) to take whatever evaluation approach is the most rigorous. Rigorously designed evaluations with stringent methodologies are likely to be accepted by the scientific community and perhaps published in prestigious journals. But rigorously designed evaluations with stringent methodologies are usually expensive, and stakeholders may not have the necessary funds. Similarly, rigorous designs are not completed quickly, and stakeholders may be working within a window that accommodates client or community needs rather than scholarly ones. To make a generalization, the evaluation approaches and research methods within the assessment and enlightenment strategies of the taxonomy demand more scientific rigor and so take more time to finish. On the other hand, the evaluation approaches under the taxonomy’s development strategies, although they manifest a brevity that loves deadlines, also embrace “flexible” methods like the focus group, which can be construed as departing from the rigor of the established scientific standard. This book certainly endorses the use of rigorous designs and methods where and when feasible. It equally reiterates that program evaluation is an applied science. Serving stakeholders’ needs as responsively as possible must remain a paramount concern as the evaluation approach and research method are selected. Rigor is a major factor, not the major factor, for the evaluator’s consideration. So, if stakeholders offer sufficient money to support rigorous designs and methodologies, evaluators should exploit this. When money or time is necessarily limited, however, evaluators should not feel compelled to advocate an evaluation approach and research method that would be a financial burden or come to its conclusions belatedly.

Here is an example. The methods of outcome evaluation (such as efficacy evaluation or effectiveness evaluation) are rigorous and lengthy, whereas those of outcome monitoring are less demanding. Stakeholders whose priority is highly credible and precise information about a program’s effects want outcome evaluation. Stakeholders on a tight budget of cash, time, or both, want something else. If they want simply some rapid feedback about clients’ progress, it would be inappropriate for an evaluator to advocate an expensive outcome evaluation when less costly outcome monitoring can also provide that feedback. The evaluator’s role is to inform stakeholders that such an
option exists, and that it represents a trade-off, but one that will conserve their time and money. A second illustrative example lies in the trade-off between cost and thoroughness. Evaluative information can be costly; the deeper an evaluation delves, the costlier it becomes. Programs almost always are constrained by cost, and the results, when they are evaluated, are a trade-off between evaluative product and price. Stakeholders with a program in the planning stage need to realize that they can save money by seeking only an evaluator’s comments on a program plan if they can forgo the deeper data that costlier formative research or needs assessment would provide. Of course, if evaluator comments are unlikely to shed any new light on the program plan, the stakeholders might be better off waiting until they can afford the more expensive option and its data bearing directly on the decisions they must make.

4. Communicate to stakeholders facts about the chosen evaluation strategy/approach and research method. When the evaluator has determined which evaluation strategy, evaluation approach, and research method fit the assignment best, he or she must explain them carefully to the stakeholders. Stakeholders should be especially well instructed about the kind of information that will be the final product. Communication helps prevent misunderstanding between stakeholders and evaluator. It gives stakeholders an opportunity to voice any doubt about the proposed evaluation’s capacity to meet their needs. (Any evidence of such doubt should cause the evaluator to reexamine the options.) Finally, free communication with stakeholders also gives evaluators a forum for detailing the kind of support expected from stakeholders throughout the evaluation process.

EVALUATION RANGING ACROSS SEVERAL PROGRAM STAGES

Program evaluators are frequently engaged to conduct multiple-entry evaluations; that is, evaluations across program stages. Before beginning, conflicts of interest that could be created by such multitasking must be addressed. Generally speaking, when the various tasks all fall within the domain of the development strategies, or the domain of the assessment/enlightenment strategies, conflict of interest is negligible. Evaluators can, for example, carry out evaluation activities that assist in the development of a program plan and
also, later on, provide the data to facilitate program implementation. Because each evaluation is confined to one phase and thus is of a consistent nature, the evaluations complement each other instead of competing with each other. Similarly, no conflict results when an evaluator performs assessment evaluation during the implementation stage and goes on to assess the program’s effectiveness in the outcome stage. The natures of the two evaluations are compatible.

Attention to conflict of interest is warranted when evaluators doing development-oriented work with programs in their early stages subsequently become responsible for assessing program performance/merit in later stages. Whether an actual conflict exists depends on the strategies and approaches involved and on whether evaluators had a direct role in the decisions made about program planning and implementation. Conflicts of interest are quite likely to occur when evaluators conduct bilateral empowerment evaluation, becoming active members of design/development teams (as in the development partnership strategy described above), then later assume responsibility for assessing program merit. A team member-evaluator is seen as having a vested interest in the program. If he or she were to declare the program successful, the credibility of the outcome could well be suspect. Following completion of empowerment-based evaluation projects, it is much better to secure new evaluation professionals to carry out any assessment or enlightenment type of evaluation.

Evaluators whose involvement in the development facilitation strategy is limited to facilitating the work of stakeholders are not prohibited from conducting assessment or enlightenment types of evaluation of the program during later stages. “Vested interest” is not applicable in cases in which evaluators conducted needs assessment, formative research, or formative evaluation (in the program planning stage) for the benefit of stakeholders designing or developing their program. This facilitation experience is not grounds to exclude them from evaluating the program’s implementation and effectiveness later on. In the same way, an evaluator who has worked to facilitate stakeholders’ development of a logic model or program theory is not barred from later assessments of the program. However, as a precaution protecting the perceived credibility of an evaluation, evaluators in these situations need to do three things. They must first offer up for discussion and scrutiny the fact and the nature of their earlier involvement in development activity. Second, they must make it clear to stakeholders that the requirements for evaluating
programs in later stages differ from requirements for development-oriented evaluations. As a final condition, they must document explicitly how they arrived at the major conclusions of their evaluations.

HELPING STAKEHOLDERS GEAR UP
(OR CLEAR UP) THEIR PROGRAM THEORY

As the practical taxonomy suggests, when evaluators set about reviewing a program using an approach associated with development facilitation strategy or enlightenment strategy (strategies discussed further in the chapters that follow), a frequent first requirement is clarification of the stakeholders’ program theory (Chen, 2003). At times, the evaluator may even need to help the stakeholders with the initial draft of a program theory. This section explores ways to clarify or help develop stakeholders’ program theories.

Reviewing Existing Documents and Materials

To start the process, evaluators need to study existing documents or materials related to the program—brochures, pamphlets, grant applications, memos, and so on. This provides general information, preparing the evaluator for subsequent interviews with stakeholders and ensuring that these will be conducted efficiently. Evaluators might also consider visiting program sites to increase familiarity with programs that have already been implemented.

Clarifying Stakeholders’ Theory

As the evaluator begins to clarify stakeholders’ program theory, or as such a theory begins to be developed by stakeholders with assistance from the evaluator, an important issue must be resolved. What role should the evaluator play in this process? How can he or she best contribute to the work? The evaluator should remember that a program theory belongs to the stakeholders; the evaluator’s function is that of facilitator and consultant. Evaluation skills and knowledge should be brought to bear to increase the productivity of the meetings at which various stakeholders attempt to articulate and refine their ideas about the program theory. Stakeholders are sure to have divergent backgrounds, concerns, and interests. It is easy for them to eat up time with
freeform discussions that never even approach agreement. The evaluator’s job as facilitator is to outline for the group the salient issues to discuss, showing stakeholders where to fill in with their own experiences, thought, and expertise. Next, the evaluator can synthesize the discussions and build consensus. The evaluator’s concurrent job as consultant means filling in with his or her own evaluation expertise when stakeholders ask for advice. The evaluator is present to lay out options for stakeholders to consider and should avoid imposing his or her own values upon stakeholders. The evaluator should also present ideas drawn from his or her own expertise for stakeholders to discuss.

**Participatory Modes for Development Facilitation**

Evaluators can assist stakeholders whose program theory is under development by adopting either of two general participatory modes: the **intensive interview mode** or the **working group mode**. Choosing a mode is a prerequisite for stakeholders and evaluators preparing to work together. The **intensive interview mode** centers on individual, intensive interviews that the evaluator holds with representatives from each key stakeholder group. The aim is to record systematically the individuals’ perceptions about issues within the incipient program theory. Based upon these interviews, the evaluator formulates a first draft of the program theory, which will be read by the representatives and other stakeholders. Their comments are considered as the final draft is prepared. In addition, evaluators can meet with these individuals for the purpose of fine-tuning and finalizing the program theory. The **working group mode** similarly involves representatives from key stakeholder groups. However, in this mode, the representatives are not interviewed individually but instead meet together with the evaluator to develop the program theory. Group members need to consist of those who will be most deeply involved in formulating and designing the program, those who will be most deeply involved in implementing the program, and other key constituencies whose input will be influential as to the direction the program will take. The facilitator, of course, is another member.

This list actually creates relatively few participants when the planned program is a small one. With large programs, however, the working group tends to be too large. A group that is too large can discourage members’ full participation, at the same time necessitating many more sessions to finish the work. A good rule of thumb is to limit a group to no more than 15 members. Small groups can
foster a casual atmosphere for discussion, enabling the evaluator to serve as both facilitator and consultant. A large group, especially one with a highly diverse and vocal membership, makes it difficult for the evaluator to be facilitator and consultant at once. With large groups, at least two evaluators may need to participate in the meetings—one as facilitator, the other as consultant.

How should one choose a participatory mode? Each has its advantages. The intensive interview mode tends to be less challenging logistically because group meeting arrangements are needed only infrequently. In addition, the interview setting may strike some participants as being much more comfortable and secure than a typical meeting. The interview also tends to better promote probing of stakeholders’ views by the evaluator. A potential limitation of the intensive interview mode, however, is some stakeholders’ perception that they have participated in only one part of the theorizing process. This is especially problematic in large programs with many powerful stakeholders. In contrast, the working group mode tends to demonstrate that the program theory is being developed in an open, inclusive manner, which could increase some stakeholders’ buy-in. But again, work with a group often requires more time to finish than work done in interviews. Furthermore, it is possible in working groups for a few highly vocal stakeholders to dominate discussion. This problem might be alleviated if the evaluator sets clear rules of discussion from the first meeting. Rules should encourage full participation by all members. An even more serious problem with the working group mode is that some stakeholders—those in the lower ranks of the implementing organization(s)—may worry about expressing their actual opinions, choosing instead to simply echo what higher-ranking officials say. In such a case, the final program theory could reflect only the views of those in authority. If this is a concern, the intensive interview mode is the better choice.

**Theorizing Procedures for Development Facilitation**

As with the participatory mode, a theorizing procedure must be selected in order to help stakeholders develop their program theory. So-called forward reasoning, backward reasoning, and forward/backward reasoning are the three general options for evaluators working within the development strategies. Backward reasoning is an approach that begins with the change model, then moves backward step by step to the action model in order to obtain the program theory. It is “backward” reasoning in that the process moves in the direction
opposite of the sequences shown in Figure 2.1. More specifically, backward reasoning starts from the question of what goals the program seeks to achieve. Other questions are the following: On which determinants of these goals should the program focus? What intervention will affect these determinants in appropriate ways? When a change model has been completed, evaluators can facilitate stakeholders’ development of the corresponding action model with questions such as these: Which groups need to be reached and served? What kind of program implementers and implementing organizations will suit? What types of intervention and implementation protocols seem best? Should there be collaboration with other organizations? Will the program require ecological support?

Forward reasoning, on the other hand, means formulating a program theory in accord with the logic flow outlined in Figure 2.1—action model first, then change model. Forward reasoning produces general program goals and grows from initial thought on what kind of action model is needed. Questions like these are important in forward reasoning: At which intervention and implementation protocols will the implementing organizations excel as they try to solve particular problems or reach certain goals? What group needs to be reached with the intervention, and how can it be reached? What setting and delivery mode make sense? Do clients face barriers to receiving services, and can the program alleviate these? How and where should contextual support be sought for the intervention, if needed? When they have completed the action model, evaluators and stakeholders can develop a change model by asking two questions, in sequence: What determinants will be changed by the intervention? What outcomes will be achieved by changing these determinants?

Forward reasoning and backward reasoning alike can be used successfully in the formulation of program theories. In certain circumstances, however, one of the two theorizing procedures is clearly the better choice. Some rules of thumb can guide the evaluator.

The first rule says that, generally speaking, when program designers and other key stakeholders are familiar with social science methodology, backward reasoning works best. It is the procedure that starts with discussion of a program’s goals, a subject stakeholders enjoy discussing and that can help break the ice. Subsequent inquiries within the backward reasoning procedure (e.g., What are the causes of the problem? Which intervention seems to offer promise? What is an appropriate design for the intervention?) are well within the stakeholders’ capability to debate. On the other hand, when program designers and other key stakeholders are not familiar with social science methodology, forward reasoning should be preferred. The reason is that theorizing
procedures need to start with a topic that stakeholders feel comfortable discussing. Forward reasoning starts with the specification of programming issues, about which stakeholders have many ideas to voice. Forward reasoning aptly suits efforts to clarify or develop stakeholders’ views on the actual steps their program should take: what to do first of all, what to bring in next, building up to the third and fourth and fifth steps, and so on through culmination in delivery of a service or services. Whether an evaluation begins with forward or backward reasoning, if the evaluator and stakeholders come to realize that continuing in that mode will be difficult, they are always free to switch to the other procedure to resume their discussions.

It is also important to note that forward and backward reasoning are not mutually exclusive. The forward/backward reasoning is a use of forward and backward reasoning, back and forth, to facilitate stakeholders and make explicit their program theories. The forward/backward reasoning is more time-consuming than the other two approaches, but may have the best of both worlds. In using this technique, evaluators and stakeholders often apply backward reasoning first and then use forward reasoning to compensate for weaknesses in backward reasoning. For example, an evaluation focused on both action and change models might begin with the forward reasoning procedure to construct an action model, take up backward reasoning to establish a change model, and finally integrate the two to arrive at an overall program theory. This dual procedure is a good choice when program stakeholders and evaluators believe that unintended outcomes will be of import. Employing the theorizing procedures in both directions may make it more likely that a working group will be alerted to potential unintended desirable or undesirable effects. The evaluator should facilitate discussion of any unintended effects and their prevention, should they be undesired.

Preparing a Rough Draft that Facilitates Discussion

The act of developing a useful program theory is often time-limited. The work’s usefulness may dwindle with the passing of a deadline and, more often than not, deadlines come sooner than is desirable for the planning team. Scheduling, preparing for, and executing either interviews or meetings, and then compiling the information obtained and soliciting comment on it, is very time-consuming (and especially so if every element and issue needs to be broached, examined, and ruled on—from scratch—in these meetings or interviews). To shorten the period required, it is not unusual for evaluators to scour
existing information about a program and use what they learn to prepare a rough draft of a program theory for discussion by the working group. The rough draft should include the elements of a program theory stated in the existing information, the elements that may be implicit in the existing information but are not communicated straightforwardly there, and the significant elements not yet touched on that will require intensive discussion. The rough draft provides a focus for stakeholders’ thoughts and suggestions. It should be distributed to members of the working group (or to individuals scheduled for interview) well in advance of the meeting date, giving them time to digest the contents. The rough draft is a tool to streamline discussion, focus comment, and foster specificity and usefulness in the work.

DETAILS OF THE WORK:
DESIGNING AN EVALUATION

The taxonomy displayed in Table 3.1 and discussed above is useful for identifying stakeholders’ evaluation needs at certain program stages. That task concluded, evaluators may consult Chapters 4 through 10 for the detailed principles and guidance helpful in making a final choice of evaluation strategy and approach. With a strategy and approach (or approaches) determined, the time has arrived to design the evaluation and launch it in the field. Chapters 4 and 5 (and beyond) discuss evaluation design in depth and are tailored to cases in which stakeholder needs pertain to the planning stage. Chapter 6 concentrates on designing evaluations at the initial implementation stage, and in Chapter 7 (and in the process-monitoring portion of Chapter 8) the focus is evaluation design at the mature implementation stage. The outcome-monitoring portion of Chapter 8, and Chapters 9 and 10 as well, are tailored for evaluators involved in outcome stage evaluation.

DYNAMICS OF EVALUATION ENTRIES
INTO PROGRAM STAGES AND HOW TO APPLY
THIS BOOK TO CONDUCTING EVALUATION

The application of evaluation along program stages is dynamic in nature. Evaluators might be asked to conduct an evaluation focusing on either any one
stage or a combination of stages. Figure 3.1 highlights the dynamics of such evaluation application.

The center section of Figure 3.1 indicates that program stages ideally move from planning to initial implementation to mature implementation and, eventually, to outcomes. It is possible, however, to move in a nonlinear fashion. Figure 3.1 demonstrates both single entry and multiple entry evaluation. The definitions of these two types of evaluation, as well as how to apply this book to conducting these two types of evaluations, are presented below.

1. Single Entry Evaluation

In single entry evaluation, evaluators focus their evaluation on a single program stage. This book is organized in a way that accommodates a single entry evaluation. As long as readers have a basic knowledge of the information found in Chapters 1 through 3, they can go directly to the chapter that applies to the stage of evaluation in which they are interested. For example, if evaluators are interested in outcome evaluation, they can move from Chapter 3 directly to Chapters 9 and 10, which discuss the major issues of outcome evaluation; they do not have to refer to Chapters 4 through 8 in order to conduct an outcome evaluation, though these chapters may still prove helpful. Similarly, readers that are interested in program planning can move from
Chapter 3 to Chapters 4 and 5, which discuss evaluation approaches used in the planning stage. Readers interested in conducting an evaluation at the initial implementation stage can move from Chapter 3 to Chapter 6; readers interested in process evaluation can move from Chapter 3 to Chapter 7; readers interested in program monitoring can move from Chapter 3 to Chapter 8; and readers interested in outcome evaluation can move from Chapter 3 to Chapters 9 and 10.

2. Multiple Entry Evaluation

In multiple entry evaluation, evaluators are concerned with conducting an evaluation that focuses on two or more program stages. For example, at the beginning of a program, evaluators may be asked to conduct an evaluation of any two or more program stages, from planning to outcome. Similarly, in an established program, evaluators may be asked to conduct an evaluation covering both the implementation and outcome stages. This book can be used effectively to guide multiple entry evaluations. After evaluators and stakeholders have decided which combination of program stages or evaluation approaches are to be used, evaluators could read the chapters relevant to these stages. For example, if evaluators are asked to conduct process evaluation and outcome evaluation, they could refer to Chapters 7, 9, and 10. Similarly, if they are asked to conduct evaluation in program planning and initial implementation stages, they could refer to Chapters 4, 5, and 6.

The entries of program stages in a multiple entry evaluation could be non-linear. For example, imagine that stakeholders are not happy with their existing program. They may ask evaluators to conduct an evaluation at the mature implementation stage to learn from their mistakes and then ask evaluators to conduct evaluations at the planning stage to facilitate their development of a new program. In this case, the evaluators could refer to Chapter 7 first and then to Chapters 4 and 5.