ROLES FOR THEORY IN CONTEMPORARY EVALUATION PRACTICE: DEVELOPING PRACTICAL KNOWLEDGE

Stewart I. Donaldson and Mark W. Lipsey

Relative to many professions, evaluation has a brief but interesting history. Evaluation scholars often note the work of Ralph Tyler and his “Eight Year Study” of progressive education in the 1940s as one of the first landmarks in the development of the modern profession and discipline of evaluation (see, e.g., Alkin, 2004a). However, the first major boom in evaluation seemed to occur in the United States in late 1960s and 70s under the Kennedy and Johnson Administrations, when social programs were developed on a grand scale and heavily supported by federal funding under the policies of the “War on Poverty” and the “Great Society” (Rossi, Lipsey, & Freeman, 2004). Many of our most sophisticated experimental methods, quasi-experimental designs, and data analytic techniques for generalized causal inference were developed in response to the challenges of determining the net impact of these and subsequent large-scale government social programs and policies (Shadish, Cook, & Campbell, 2001).

Donaldson & Scriven (2003a) noted that we are now experiencing what could be called the “Second Boom in Evaluation,” which is more global and has notably different characteristics than our first growth spurt. For example, in the past decade, we have witnessed societies all around the world embracing the values of accountability and professionalism. This has led organizations and agencies of all shapes and sizes to commission professional evaluations at a
dramatically increasing rate. While these evaluations still include federally funded and other types of program and policy evaluations, interest in evaluating non-governmental programs, personnel, proposals, performance, technology, research, theory, and even evaluation itself fall under the growing domain of contemporary evaluation practice (American Evaluation Association, 2005). One indicator of the increasing demand for evaluation practice is the number of professionals now participating as members of organized evaluation associations and societies. In 1990, there were approximately five major evaluation professional associations, whereas today there are more than 50 worldwide (see Donaldson & Christie, 2006; Mertens, 2003), as well as an international alliance to link them together to share knowledge about how to improve the practice of evaluation (Mertens, 2005; Russon, 2004).

Another characteristic of the second boom is that new theories of evaluation practice, new evaluation methods, and new evaluation tools are being developed and refined to address a much broader and diverse range of evaluation practice challenges. Donaldson & Christie (2006) and Scriven (2003) describe some of these new challenges under the rubric of evaluation as a “transdiscipline” that supplies essential tools and techniques for a wide range of other disciplines, while retaining an autonomous research effort of its own, focused on advancing knowledge about how best to practice evaluation.

General organizing frameworks have been called for, and are beginning to emerge, as a way to recognize the new challenges faced by evaluators today (Mark, 2003). Mark, Henry, & Julnes (2000) provided a framework to guide the evaluation of programs and policies aimed at promoting social betterment. They proposed that four main evaluation purposes have now evolved in the program and policy evaluation domain:

1. Program and organizational improvement;
2. Oversight and compliance;
3. Assessment of merit and worth;
4. Knowledge development.

Keeping these purposes in mind, we will attempt to address in this chapter new and emerging roles for theory in “second boom” program evaluation practice. We will pay particularly close attention to those parts of the practice of evaluation that aspire to contribute to social programming and policymaking by contributing to the knowledge and theoretical base – in large part by using, developing, testing, or otherwise enhancing relevant theory.

**Theory for Evaluation Practice**

Reference to theory is widespread in the contemporary evaluation literature, but what is meant by “theory” encompasses a confusing mix of concepts related to evaluators’ notions about how evaluation should be practiced, explanatory frameworks for social phenomena drawn from social science, and assumptions about how programs function or are supposed to function. A newcomer to evaluation, and even a grizzled veteran, could have a difficult time sorting through the closely related and sometimes interchangeable terms that litter the evaluation landscape – theories of practice, theory-based evaluation, theory-driven evaluation, program theory, evaluation theory, theory of change, logic models, and the like.

Aside from considerable potential for confusion, the nature and role of theory in evaluation is often a contentious matter. On the one hand, such distinguished evaluators as Scriven (1998, 2004a, 2004b) and Stufflebeam (2001, 2004) have asserted that there is little need for theory or, at least, some forms of theory, in evaluation. Scriven (2004a, 2004b), for
instance, claimed "it's possible to do very good program evaluation without getting into evaluation theory or program theory," and declared that "the most popular misconception amongst currently politically correct program evaluators is the evaluation of a program (a) requires that you have, or (b) is much benefited by having, a logic model or program theory." Similarly, Stuffelbeam (2001), in a review of evaluation models and theories of evaluation practice, remarked that "there really is not much to recommend theory-based evaluation, since doing it right is usually not feasible and since failed or misrepresented attempts can be counterproductive." More recently, Stuffelbeam (2004) described the "now fashionable advocacy of 'theory-based evaluation'" as a situation

... wherein one assumes that the complexity of variables and interactions involved in running a project in the complicated, sometimes chaotic conditions of the real world can be worked out and used a priori to determine the pertinent evaluation questions and variables ... Braybrooke & Lindblom (1963) discredited this notion 40 years ago, and their message clearly needs to be revisited. (p. 253)

In contrast, other eminent evaluators have argued that program theory, evaluation theory, and social science theory all do, and should, play important roles in modern program evaluation (e.g., Alkin, 2004a; Chen, 1990; Donaldson, 2003; Fettermann, 2003; Lipsey, 1990; Mark, 2003; Rossi et al., 2004; Shadish, Cook, & Campbell, 2004; Weiss, 2004a, 2004b). For example, in the case of evaluation theory, Shadish (1998) introduced his presidential address to the American Evaluation Association, entitled "Evaluation Theory Is Who We Are," with the following assertion:

All evaluators should know evaluation theory because it is central to our professional identity. It is what we talk about more than anything else, it seems to give rise to our most trenchant debates, it gives us the language we use for talking to ourselves and others, and perhaps most important, it is what makes us different from other professions. Especially in the latter regards, it is in our own self-interest to be explicit about this message, and to make evaluation theory the very core of our identity. Every profession needs a unique knowledge base. For us, evaluation theory is that knowledge base. (Shadish, 1998, p. 1)

One purpose of this chapter is to attempt to sort out some of the different meanings of theory in evaluation and the utility the different forms of theory might have for evaluation practice. This will be done in service of the larger purpose of discussing evaluation work that is carried out in an effort to generate knowledge and theory for program practice and policy-making. We will first distinguish the three most common types of theories encountered in evaluation – evaluation theory, social science theory, and program theory – and discuss their respective roles and limitations. We will then describe one way in which these different forms of theory intersect that defines a widespread and, we believe, especially fruitful approach to evaluation – where the theory of evaluation involves integration of social science theory in the development and use of program theory to guide evaluation practice and expand knowledge about how programs bring about social change. We suggest that, while both program theory and social science theory can enhance program planning and evaluation, the cause of knowledge development as a pathway to evaluation influence is particularly likely when both kinds of theory are integrated. Finally, we conclude the chapter by considering the implications of that perspective for the future of evaluation.


**Evaluation Theory**

Chen (1990) described theory as a frame of reference that helps humans understand their world and how to function within it. In program evaluation, he emphasized that it is important to distinguish between descriptive and prescriptive theories. Simply stated, descriptive theories characterize what is and prescriptive theories articulate what should be. *Evaluation theories* are largely prescriptive and “offer a set of rules, prescriptions, prohibitions, and guiding frameworks that specify what a good or proper evaluation is and how evaluation should be done” (Alkin, 2004a). They are thus theories of evaluation practice that address such enduring themes as how to understand the nature of what we evaluate, how to assign value to programs and their performance, how to construct knowledge, and how to use the knowledge generated by evaluation (Shadish, 1998).

In one sense, evaluation theories are less important than social science theories or program theories for a discussion of knowledge development as a purpose of evaluation. In another sense, evaluation theory is quite important in this regard: Contemporary evaluation theories differ in terms of their views of the appropriateness of and their emphasis on the four main purposes of evaluation. In particular, evaluation theories disagree about whether, and to what extent, priority should be given to theory building and testing, and more generally, to knowledge development, as a pathway to improving programs and policies. For this reason, and because our discussion of theory would be incomplete without some attention to evaluation theory, we give some attention to it in this chapter.

**The Evolution of Evaluation Theory**

Shadish, Cook, & Leviton (1991) characterized the history of theories of evaluation as a series of stages. Stage I theories emphasized the discovery of truth (e.g., the evaluation theories of Michael Scriven and Donald Campbell). Stage II focused on the way evaluation was used and its social utility (e.g., the theories of Joseph Wholey, Robert Stake, and Carol Weiss). In Stage III, theory development addressed the integration of inquiry and utility (e.g., the theories of Lee Cronbach and Peter Rossi).

It is only fitting that evaluation theories themselves be evaluated, and there is no shortage of critique in the writings and discussions among evaluators. The criteria Shadish et al., used specified that a good theory of evaluation practice should give a full account of the appropriate principles and practices relating to:

1. Knowledge: What methods to use to produce credible knowledge;
2. Use: How to use knowledge about social programs;
3. Valuing: How to construct value judgments;
4. Practice: How evaluators should practice in “real world” settings;
5. Social programming: The nature of social programs and their role in social problem solving.

In their final evaluation of these theories of evaluation practice, Shadish et al., (1991) concluded that only the Stage III integrative theories addressed all five criteria. However, they recognized the contributions of many perspectives to the evolution of theories of practice and argued that evaluators should not follow the same evaluation procedures under all conditions.

While the Shadish et al., framework has been widely cited and used to organize theories of evaluation practice for over a decade, it is controversial in a number of ways. Whatever
criticisms the theorists evaluated, omitted, and their respective followers might have, however; it provides one of the most fully developed perspectives available on the nature of evaluation theory.

In a more recent attempt to trace the roots of modern theories of evaluation practice, Alkin & Christie (2004) developed a different scheme for classifying evaluation theories (see Figure 2.1). Their "Evaluation Theory Tree" has a trunk built on the dual foundation of accountability and social inquiry and branches of use, methods, and valuing. They placed 26 evaluation theorists on this tree and then gave them an opportunity to reflect on their theory of practice and to suggest revisions to their classification. The resulting scheme promises to help evaluators understand fundamental differences and points of connection between some of the most common theories of evaluation practice, as well as illustrate how theorists' views change over time in light of experiences from evaluation practice. This process of reflection also underscores the point that the originators of ideas, evaluation theorists in this case, often disagree with how others interpret, describe, or classify their original works. This very different approach for understanding the history of evaluation provides a unique window on the development of the discipline and profession of evaluation.
The Role of Evaluation Theory

Many of those participating in evaluation do so within the confines of one theory of practice, often a general social science research paradigm, without much apparent reflection or concern about the underpinnings of that theory or the challenges posed by competing theories. Shadish (1998) provided at least six reasons why practicing evaluators should be thoughtful about evaluation theory:

- Evaluation theory provides a language that evaluators can use to talk with each other about evaluation;
- Evaluation theory encompasses many things in our field about which evaluators seem to care most deeply;
- Evaluation theory defines the themes of the majority of evaluation professional conferences;
- Evaluation theory provides evaluators with an identity that is different from the identity of other professionals.
- Evaluation theory provides the face that evaluators present to the outside world;
- Evaluation theory is the knowledge base that defines the profession.

Taken together, these points suggest that evaluation theory has become a central thread in the social fabric of the evaluation profession. Evaluation theory can facilitate communication amongst evaluators practicing across the globe, help evaluation practitioners understand and share best practices, and provide the rationale for the various procedures evaluators recommend and use in practice.

Knowledge of evaluation theory can also help evaluators become better ambassadors for the profession of evaluation and educators of potential clients. Because professional evaluation now offers a range of acceptable approaches and perspectives, it is critical that sponsors and users understand that there are variations and how they differ. Evaluation approaches and services may differ rather dramatically across evaluation teams. Finding an optimal fit between an evaluation team and the needs and interests of evaluation sponsors and stakeholders could arguably be one the most important factors in determining whether an evaluation will ultimately be useful (Donaldson, 2004).

In general, effective evaluation practice has the potential to help prospective clients and other stakeholders dramatically improve their work. For example, professional evaluation can help stakeholders make better decisions about service, policy, and organizational direction; build knowledge and skills, and develop a capacity for evaluative thinking; facilitate continuous quality improvement and organizational learning; and provide accountability or justify a program, policy, or organization's value to investors, volunteers, staff, and prospective funders.

Beyond the general benefits of evaluation, however, is the question of how appropriate a particular evaluation is for a particular program at a particular time. It is important to consider who could be negatively affected by an evaluation of a given sort, how much time and resources may be taken away from program services while the evaluation is being conducted, and the ways in which the evaluation process might be uncomfortable and disruptive for some project team members and other stakeholders (Donaldson, 2001b; Donaldson, Gooler, & Scriven, 2002). It must also be recognized that the questions a particular evaluation asks and the way in which it goes about answering those questions will have repercussions that will not always be constructive. When evaluators and stakeholders fully explore the potential benefits and costs of doing a specific evaluation and consider other options and approaches (based on other theories of practice), their expectations and plans become more realistic and the evaluation is much more likely to reach its potential (see Donaldson, 2001b).
Finally, evaluations are subject to critique, meta-evaluation, and even hostile attacks (most likely when negative findings surface—the "kill the messenger" phenomenon; Donaldson et al., 2002). Knowledge of evaluation theory can help evaluators better understand reactions to their work and help them defend against critics making different assumptions about evaluation design or unfairly using a fundamentally different theory of practice to discredit the work.

Of course, being knowledgeable about the history and "state of the art" of evaluation theory does not guarantee successful evaluation practice. Some theories of practice may not be effective or may even be harmful in certain circumstances. Moreover, every theory of practice is likely to be more effective in some settings than in others. Recognizing boundary conditions for the role of any one evaluation theory, and for the use of evaluation theories in general, is essential for good practice (Mark, 2003). In the end, logically consistent guiding frameworks for evaluation practice that are sensitive to the mainstream principles and standards in the field (e.g., American Evaluation Association, 2005; Joint Committee on Standards for Educational Evaluation, 1994) should play an important role in contemporary evaluation practice.

**Social Science Theory**

Social science theories are not especially concerned with methods or practices for evaluators. Rather, social science theories attempt to provide generalizable and verifiable knowledge about the principles that shape social behavior. When such theories address the social phenomena related to social programs and the social conditions they are intended to improve, however, they may be very relevant to evaluation.

Evidenced-based social science theories are often helpful for understanding the etiology of desired or undesired outcomes and for developing intervention strategies for influencing those outcomes. For example, social cognitive learning theory has been used effectively to design programs to promote positive social norms to prevent alcohol and drug use, risky sexual behavior, breast cancer, and a range of other social and behavioral problems (Bandura, 2006; Donaldson, Graham, & Hansen, 1994; Petraltis, Flay, & Miller, 1995). Indeed, many such social science theories have been, or might be, used for designing, improving, and evaluating programs: the theory of planned behavior (Ajzen, 2002; Bamberg & Schmidt, 1997); Prochaska & DiClemente’s (1999) theory of health behavior change, Kram’s (1985) theory of mentoring, Sternberg’s (2003) theories of practical intelligence and leadership, theories of re-employment training (Price, van Ryn, & Vinokur, 1992), theories of learned helplessness (Seligman, 2003), and Rosenthal's (2006) theory of interpersonal expectations and self-fulfilling prophecies, just to name a few.

While it is true that the conceptual and empirical evidence base is sometimes thin for any given problem, we would argue it is misguided for evaluators to ignore potentially applicable prior research and empirically based social science theory when it is available (as has been suggested by, e.g., Scriven, 1998, 2004a, 2004b, and Stufflebeam, 2001, 2004).

**The Role of Social Science Theory**

Social science theory can play several important roles in evaluation practice. First, such theory and prior research can be very informative for initial needs assessment and program design. Many, if not most, social problems have been encountered by others, and sometimes research or evaluation exists on efforts to prevent or solve these problems. A careful examination of available literature, including
primary studies, as well as syntheses such as meta-analyses, may turn up knowledge about effective strategies for dealing with the problems of concern, or just as important (and probably more likely), lessons learned about what does not work, which may save program designers and evaluators countless hours and resources. "Operating within a vacuum," when useful information exists, can be a very inefficient way to practice evaluation.

Another important role for social science theory and research is to help evaluators assess the likelihood that programs will be able to accomplish certain objectives. It is not uncommon to find stakeholders who have very unrealistic aspirations for their program, given what is known about the desired outcomes from prior research and evaluation. Such stakeholders might be best helped by a focus on program improvement rather than, say, a full-scale evaluation for outcomes they have little chance to produce. In other circumstances, stakeholders may believe there is something significantly different about their program, target population, and/or context relative to those evaluated previously. In this case, outcome evaluation might proceed, but findings from previous work could suggest important contextual or moderating factors to measure and include in the analyses (Donaldson, 2001a).

Finally, social science theory and research are sometimes useful for guiding evaluation measurement and design decisions, and can provide a context for interpreting evaluation findings. For example, it is sometimes possible to locate relevant and valid measures of constructs of interest or feasible designs that have lead to unequivocal findings, as well as measures and designs that are not likely to lead to valid results in your work. Previous theory and research can also provide a context, or suggest expectations for the range of effect sizes evaluators should expect. Estimated effect sizes in the current evaluation can be compared to previous findings in an effort to further explore the relative practical significance of the program under investigation (Donaldson et al., 2001; Lipsey, 1990).

One of the notable boundary conditions that must be acknowledged when using social science theory and research in evaluation practice is the limit of generalizability. Oftentimes the characteristics of previous research and evaluation are, to various degrees, different than those in the current investigation. It is very important not to assume, without serious critical reflection, that the results from previous work will necessarily generalize to the current application. Some evaluation theorists have suggested that evaluators should be much more concerned about producing local knowledge than concerning themselves with producing generalizable knowledge (e.g., Alkin, 2004b; Stake, 2004). While we agree with this prescription to a certain degree (i.e., it is important to meet contractual obligations and local needs first), examining local findings in relation to social science theory and research can be beneficial both for interpreting current findings and in adding to the knowledge base of how to prevent or solve societal problems more generally.

Scarcity of quality work is another common limitation encountered when trying to use social science theory and research in evaluation practice. That is, many areas of programming simply do not have much sound social science theory and research to draw upon. While somewhat related work might exist, it is important to be extremely careful not to overuse or generalize findings and lessons learned from only remotely similar efforts. However, testing the generalizability limits of related existing theory is another potential substantive benefit of incorporating social science theory in programs and their evaluation. As the discipline and profession of evaluation mature, our cumulative knowledge base in this area
should expand and the scarcity problem is likely to be less common (Weiss, 2004b).

**Program Theory**

While evaluation theory is concerned with how to practice evaluation, *program theory* focuses on the nature of the evaluand itself (i.e., the program, treatment, intervention, policy, etc. being evaluated). Program theory should not be confused with social science theory, nor should it conjure up images of broad concepts about the nature of social problems (Donaldson, 2003; Weiss, 1997). Program theory is much more modest and deals with the assumptions that guide the way specific programs, treatments, or interventions are implemented and expected to bring about change (Donaldson, 2001b; Lipsey, 1993). The following definitions of program theory capture the essence of how program theory is typically defined in evaluation practice today:

- The construction of a plausible and sensible model of how a program is supposed to work (Bickman, 1987).
- A set of propositions regarding what goes on in the black box during the transformation of input to output, that is, how a bad situation is transformed into a better one through treatment inputs (Lipsey, 1993).
- The process through which program components are presumed to affect outcomes and the conditions under which these processes are believed to operate (Donaldson, 2001b).

Rossi et al., (2004) describe program theory as consisting of three main components:

1. The *organizational plan*; How to garner, configure, and deploy resources, and how to organize program activities so that the intended service delivery system is developed and maintained.

2. The *service utilization plan*; How the intended target population receives the intended amount of the intended intervention through interaction with the program’s service delivery system.

3. The *impact theory*; How the intended intervention for the specified target population brings about the desired social benefits.

The organizational and service utilization plans together constitute the *program process theory* and the impact component is referred to as *program impact theory* (see Figure 2.2). This form of practical program theory is often referred to as the “program logic” and various logic modeling techniques and ways of depicting program logic have become commonplace in evaluation practice (Funnel, 1997; Gargani, 2003).

Program process theory must usually be developed from information that comes almost entirely from the program and its immediate context. Program impact theory, with its focus on the nature of the social, psychological, or behavioral change a program intends to bring about, however, may be informed by social science theory. This is a desirable circumstance; program impact theory gains plausibility if it is rooted in, or at least consistent with, behavioral or social science theory or prior research (Donaldson et al., 2001). Unfortunately, sound theory and research are not often available for the social problem of concern. The development of program impact theory must then rely on other sources of information such as the implicit theories held by those close to the operation of the program, program documents, observations of the program in action, and exploratory research to test critical assumptions about the nature of the program (Donaldson, 2001b; Rossi et al., 2004).

**The Role of Program Theory**

At the most practical level, a well-developed and fully articulated program theory can be very useful for framing key evaluation questions and designing sensitive and responsive
evaluations. By indicating clearly what is assumed and expected in the operation and outcomes of a program, program theory helps the evaluator and program stakeholders identify the performance dimensions most critical to the program’s success and, hence, those that may be most important to assess. For evaluation design, a detailed program theory informs the selection of variables, timing of measures and observations, identification of appropriate informants, and a host of other conceptual and procedural aspects of the evaluation plan. In many regards, it is the articulated program theory that links the evaluation design to the concerns and understandings of the stakeholders – it provides the common representation of the program that all parties can reference as the evaluation is planned and implemented, and its results interpreted.

As depiction of program theory has become more valued and common in evaluation practice, more and more clients and stakeholders are expecting evaluators to know how to develop and use it. Program theory is now playing an important role in needs assessment, program planning and design, and evaluability assessment; in providing a basis for informed decisions about measurement and evaluation methods; in enabling evaluators to disentangle the success or failure of program implementation from the validity of the program’s conceptual model; and for facilitating a cumulative wisdom about how programs work and how to make them work better (Donaldson, 2003, forthcoming).

The essential involvement of program stakeholders in the articulation of program theory has also emerged as one of the more significant aspects of an emphasis on theory as a tool for evaluation planning. Engaging stakeholders in discussions about the what, how, and why of program activities is often empowering for them and, additionally, promotes inclusion and facilitates meaningful participation by diverse stakeholder groups (see Centers for Disease Control Program Evaluation Framework, 1999; Donaldson, 2003). In involving stakeholders in this way can dramatically increase the chances that the evaluation will meet utility, feasibility, propriety, and accuracy
evaluation standards (Joint Committee on Standards for Educational Evaluation, 1994). The engagement of stakeholders around developing program theory is a very important role for program theory to play in evaluation practice.

As with all evaluation tools and approaches, there are boundary conditions that evaluators must consider when developing and using program theory in evaluation. Some of the circumstances that can prevent program theory from reaching its potential in practice include lack of stakeholder cooperation, disagreements about program theory, highly dynamic programs and program theories, and lack of consensus about evaluation questions. What constitutes credible evidence, and how to use and disseminate evaluation findings (Donaldson, 2003, forthcoming; Donaldson et al., 2002; Mark, 2003). It is important for evaluators to assess project conditions carefully to determine if the stakeholders, and the project in general, are ready for the commitment necessary for the effective development and use of program theory to guide evaluation efforts.

Where They Come Together: Program Theory-Driven Evaluation Science

As the prior discussion has made evident, one widespread variant of evaluation theory is centered on the notion of articulating and using program theory to shape those aspects of evaluation practice that are related to conceptualizing social programs, designing evaluation studies, and interpreting evaluation findings. Program theory and evaluation research based on it, in turn, are enriched when they can draw on relevant social science theory and themselves enrich that theory when they contribute to knowledge about how social intervention brings about social change. This confluence of evaluation theory, social science theory, and program theory constitutes a distinctive approach to evaluation one of us has recently broadly cast as “program theory-driven evaluation science” (Donaldson, 2005; forthcoming). We suggest that this integration of theory constitutes a (if not the) major way that evaluation contributes to social betterment by way of knowledge development.

As Gargani (2003) documented, the use of program theory in evaluation practice has a long history. He noted that the practice of articulating and testing program theory was introduced to evaluation by Ralph Tyler in the 1930s, but did not find widespread favor with evaluators at that time. It was in the 1990s that the notion of using program theory grounded in relevant substantive knowledge to guide evaluation practice took hold. The work of Chen & Rossi (1983, 1987) was especially instrumental in this development. They advocated for moving away from atheoretical, method-driven evaluation approaches to a “theory-driven” approach that they argued would both improve evaluation practice and make evaluation a more rigorous and thoughtful scientific endeavor.

Chen (1990) then provided the first text on theory-driven evaluation, which has been widely used and cited in the program evaluation literature. In their broad review of evaluation theory, Shadish et al., (1991) characterized the theory-driven evaluation approach as an ambitious attempt to bring coherence to a field in considerable turmoil and debate. In their view, this approach offered three fundamental concepts that facilitate the integration of concepts and practices in evaluation:

- **Comprehensive evaluation**: Studying the design and conceptualization of an intervention, its implementation, and its utility.
- **Tailored evaluation**: Evaluation questions and research procedures depend on whether the
program is an innovative intervention, a modification or expansion of an existing effort, or a well-established, stable activity.

- Theory-driven evaluation: Constructing models of how programs work, using the models to guide question formulation and data gathering; similar to what econometricians call model specification.

Some indication of the success of this perspective in providing an integrative framework for evaluation is its representation in contemporary evaluation textbooks. Most of the major texts in the field are based on or give significant attention to the role of program theory in evaluation practice (e.g., Posavac & Carey, 2003; Rossi et al., 2004; Weiss, 1998).

Serious dialog and applications of theory-driven or theory-based evaluation are now prevalent throughout most regions across the global evaluation landscape, and within most major evaluation associations. For example, recent volumes, papers, and visions for future applications of theory-driven evaluation have been discussed at length within the European Evaluation Community (e.g., Pawson & Tilley, 1997; Stame, 2004; Stern, 2004; Van Der Knaap, 2004). Published examples of theory-driven evaluations now exist from many parts of the world including but not limited to Taiwan (Chen, Wang, & Lin, 1997), Germany (Bamberg & Schmidt, 1997), Britain (Tilley, 2004), South Africa (Mouton & Wildschut, forthcoming), Canada (Mercier et al., 2000), as well as the United States (Bickman, 1996; Cook, Murphy, & Hunt, 2000; Donaldson & Gooler, 2002, 2003).

However, as mentioned earlier, there remains much confusion today about what is meant by theory-based or theory-driven evaluation, and the differences between using program theory and social science theory to guide evaluation efforts (Donaldson, 2003; Weiss, 1997). Rather than trying to sort out all the nuances of closely related or sometimes interchangeable terms in the evaluation literature, such as theory-oriented evaluation, theory-based evaluation, theory-driven evaluation, program theory evaluation, intervening mechanism evaluation, theoretically relevant evaluation research, program theory, program logic, and logic modeling, Donaldson (forthcoming) offers a broad definition in an attempt to be inclusive:

*Program theory-driven evaluation science* is the systematic use of substantive knowledge about the phenomena under investigation and scientific methods to improve, to produce knowledge and feedback about, and to determine the merit, worth, and significance of evaluands such as social, educational, health, community, and organizational programs.

The phrase, *program theory-driven* (instead of theory-driven), is intended to clarify the meaning of the use of the word “theory” in this evaluation context. It specifies the type of theory (i.e., program theory) that is expected to guide the evaluation questions and design. The phrase, *evaluation science*, is intended to underscore the use of rigorous scientific methods (i.e., qualitative, quantitative, and mixed methods) to answer key evaluation questions. A renewed emphasis on the reliance of evaluation on systematic scientific methods is especially important for overcoming negative images of evaluation as unreliable, soft, or a second-class type of investigation (Donaldson, 2001b). *Evaluation science* signals the emphasis placed on the guiding principle of *systematic inquiry* (American Evaluation Association, 2005) and the critical evaluation standard of *accuracy* (Joint Committee on Standards for Educational Evaluation, 1994).

We view *program theory-driven evaluation science* as essentially method neutral within the broad domain of social science methodology. Its focus on the development of program theory and evaluation questions frees evaluators...
initially from having to presuppose use of one method or another. This simple view that the choice of methods is contingent on the nature of the question to be answered and the form of an answer that will be useful reinforces the idea that neither quantitative, qualitative, nor mixed method designs are necessarily superior or applicable in every evaluation context (cf. Chen, 1997). Whether an evaluator uses case studies, observational methods, structured or unstructured interviews, online or telephone survey research, a quasi-experiment, or a randomized experimental trial to answer the key evaluation questions is dependent on discussions with relevant stakeholders about what would constitute credible evidence in this context, and what is feasible given the practical and financial constraints (Donaldson, 2005).

In an effort to incorporate various lessons learned from the practice of theory-driven evaluation in recent years (e.g., Donaldson & Gooler, 2003) and make this approach more accessible to evaluation practitioners, Donaldson (2003, 2005) presented a simple three-step model for understanding the basic activities of program theory-driven evaluation science:

1. Developing program theory;
2. Formulating and prioritizing evaluation questions;
3. Answering evaluation questions.

Simply stated, evaluators work with stakeholders to develop a common understanding of how a program is presumed to solve the social problem(s) of interest, develop the implications and priorities for what aspects of program performance need to be examined, and design an evaluation that will provide information about those aspects.

Relevant social science theory and prior research (if available) are used to inform this process and to assess the plausibility of the relationships assumed between a program and its intended outcomes. Each program theory-driven evaluation also has the potential to contribute to an evolving understanding of the nature of the change processes programs bring about and how they can be optimized to produce social benefits. This orientation contrasts sharply with a history of evaluation practice that is filled with "method-driven," "black box," and "input/output" studies that reveal little about the mechanisms of change embodied in social programs. A program theory-driven approach focuses on which program components are most effective, the mediating causal processes through which they work, and the characteristics of the participants, service providers, settings, and the like that moderate the relationships between a program and its outcomes.

One of the best examples to date of program theory-driven evaluation science in action is embodied in the Centers for Disease Control's six-step Program Evaluation Framework. This framework is not only conceptually well developed and instructive for evaluation practitioners, it has been widely adopted for evaluating federally funded public health programs throughout the United States. The CDC framework extends the more concise three-step program theory-driven evaluation science model described above to guide practitioners through six distinct evaluation steps:

1. Engage stakeholders;
2. Describe the program;
3. Focus the evaluation design;
4. Gather credible evidence;
5. Justify conclusions;
6. Ensure use and share lessons learned.

The second element of this framework is a set of 30 standards for assessing the quality of
Elements of the Framework

REFERENCE CARD

<table>
<thead>
<tr>
<th>Steps in Evaluation Practice</th>
<th>Standards for Effective Evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engage Stakeholders</td>
<td>Utility</td>
</tr>
<tr>
<td>Those involved, those affected, primary intended users</td>
<td>Serve the information needs of intended users</td>
</tr>
<tr>
<td>Describe the program</td>
<td>Feasibility</td>
</tr>
<tr>
<td>Need, expected effects, activities, resources, stage, context, logic model</td>
<td>Be realistic, prudent, diplomatic, and frugal</td>
</tr>
<tr>
<td>Focus the evaluation design</td>
<td>Propriety</td>
</tr>
<tr>
<td>Purpose, users, uses, questions, methods, agreements</td>
<td>Behave legally, ethically, and with due regard for the welfare of those involved and those affected</td>
</tr>
<tr>
<td>Gather credible evidence</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Indicators, sources, quality, quantity, logistics</td>
<td>Reveal and convey technically accurate information</td>
</tr>
<tr>
<td>Justify conclusions</td>
<td></td>
</tr>
<tr>
<td>Standards, analysis/syntheses, interpretation, judgment, recommendations</td>
<td></td>
</tr>
<tr>
<td>Ensure use and share lessons learned</td>
<td></td>
</tr>
<tr>
<td>Design, preparation, feedback, follow-up, dissemination</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.3** CDC framework (Centers For Disease Control Program Evaluation Framework, 1999).

The evaluation effort (adopted from Joint Committee on Standards for Educational Evaluation, 1994; see Figure 2.3). These standards are organized into categories relating to utility, feasibility, propriety, and accuracy. In addition, program theory-driven evaluations are designed

- **Systematic inquiry**: Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.
- **Competence**: Evaluators provide competent performance to stakeholders.
- **Integrity/honesty**: Evaluators ensure the 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 with whom they interact.
- **Responsibility for general and public welfare**: Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

Clearly we take issue with the notion that theory is not a useful tool for improving evaluation practice as has been asserted by, for example, Scriven (1998, 2004a) and Stufflebeam (2001, 2004). However, we should make it equally clear we are not claiming that program theory-driven evaluation science is a panacea for all that might all evaluation practice. Nor do we wish to argue that this approach is right for all situations. Like most of the tools in an evaluator's toolkit, theory can be very useful under some conditions and not needed in others.

It is important to note here that theories of evaluation practice, other than explicitly theory-related ones, can also play an important role in knowledge development and theory improvement for social programming and policy-making (see Alkin, 2004a; Donaldson & Scriven, 2003a; Stufflebeam, 2001, for recent discussions about evaluation theories and approaches). Indeed, as Mark and Henry suggest elsewhere in this *Handbook*, an outcome evaluation can sometimes be seen as akin to a hypothesis test, with potential to contribute to the knowledge base, even without any other of the characteristics of program theory-driven evaluation. Furthermore, a wide range of applied research applications and methods typically used in contemporary evaluation practice, such as needs assessment (Scriven, 2003), research synthesis and meta-analysis (Donaldson et al., 2001; Lipsey & Wilson, 1993), and a variety of qualitative approaches and methods (e.g., see Fettersman, 2003; Mertens, 2003, Patton, 1997), have much to offer for enhancing our knowledge and understanding of the effectiveness of social programs and policies.

**Looking Toward the Future**

While various perspectives on evaluation theory can provide important insights, practicing evaluators tend to be most concerned with current and emerging challenges for evaluation practice. To explore potential futures for evaluation practice, Donaldson & Scriven (2003b) invited a diverse group of evaluators and evaluation theorists to represent their own views at an interactive symposium. A “last lecture” format was used to encourage each participant to articulate a vision for “How We Should Practice Evaluation in the New Millennium.” These evaluators were thus asked to give a “last lecture” passing on advice and wisdom to the next generation about how we should evaluate social programs and problems in the twenty-first century. In addition to six presentations of future visions, five prominent evaluators gave lectures in which they reacted to those visions. These reaction lectures were followed by responses from the initial visionaries and an audience of more than 300 participants.

This approach to exploring theories of evaluation practice and directions for the future led to an exceptionally diverse and provocative set of ideas about how to evaluate social
programs and problems (Donaldson & Scriven, 2003b; Slater, 2006; Trilloli, 2004). For present purposes, however, one aspect of this rich discourse was especially striking – program theory was an integral component of most of the visions and practice theories articulated (Crano, 2003). That is, in addition to being a relatively distinct stand-alone theory of evaluation practice, many elements of the theory-driven approach were incorporated into other theories of evaluation practice.

For example, Scriven (2003) described a transdisciplinary view of evaluation that included enlightened use of a tripartite set of program theories – the alleged program theory, the real logic of the program, and the optimal program theory. Fetterman (2003) described how the central process of empowerment evaluation creates a program theory and Mertens (2003) noted that inclusive evaluation encourages evaluators to prevent program theories from contributing to the negative stereotyping of marginalized groups. Results-oriented management incorporates logic models and program theory to guide performance measurement (Wholey, 2003) and Mark’s (2003) integrative vision for evaluation theories suggested that theory-driven evaluation has a significant place in the toolkit for all evaluation approaches. In this regard, it is worth noting that recent presentations of “realist evaluation” focus on program mechanisms as part of program theory and evaluation practice (Mark et al., 2000; Pawson & Tilley, 1999) and that “utilization-focused evaluation” is highly compatible with a theory-driven approach (Christie & Alkin, 2003; Patton, 1997).

Another major theme that emerged from this discourse on future evaluation practice was the challenge posed by the vast diversity represented in modern evaluation theories. Mark (2003) observed that each vision for the future of evaluation gave a central place to one theory of evaluation practice and left scant room for others. One way out of the continuous polarization of perspectives that results from these diverse theoretical allegiances is the development of higher-order frameworks that integrate evaluation theories (Mark, 2003). Another option presented for achieving a peaceful and productive future for evaluation involved a strategy for embracing diversity in evaluation theory, recognizing that fundamental differences cannot be reconciled or integrated, and that clients may benefit most from having a diverse smorgasbord of options to choose from for evaluating their programs and policies (Donaldson & Scriven, 2003b).

Developing the Evidence Base

Theories of evaluation practice are based more on philosophy and experience than systematic evidence of their effectiveness. Though exhortations to test the validity of claims made by these theories have been made (e.g., Mark, 2003), there is little likelihood that the diversity of practice theories in evaluation will be greatly influenced by any accumulation of research results. Evaluation theories remain largely prescriptive and unverified.

Program theory, on the other hand, is amenable to test by systematic research. Great advances can be made in the future if there is continuing development of evidence about the validity of the various program theories evaluators encounter. Birckmayer & Weiss (2000) examined a range of theory-driven evaluations conducted over the past decade and concluded that in almost every case some aspect of program theory was not supported by the resulting evidence. With sufficient information of this sort, misguided programs that rely on untenable concepts can be more readily identified without extensive empirical investigation. In addition, an expanded evidence base will allow programs based on sound theory to be more
easily designed and the synthesis of research across evaluation studies will generate deeper understanding for evaluators, policy-makers, and social scientists about how social change can be brought about.

One purpose of program theory-driven evaluation science is to rigorously examine the validity of program theories. That is, once a program theory is developed and articulated, systematic, data-based evaluations would be used to determine if the program does bring about change in the manner theorized. Over time, program theory-driven evaluations will provide an expanding base of evidence about specific relationships that appear in the mechanisms of change assumed in social programs. This empirical evidence can be used to improve programs and program theories, as well as identify promising strategies for addressing problems in other program domains. Weiss (2004b) suggested it may well turn out, once the empirical evidence accumulates, that there are a limited number of effective change mechanisms that can be identified and applied to a wide range of social problems. A major goal of program theory-driven evaluation science, therefore, is to establish evidence-based program theories that can enhance efforts to promote social betterment.

Conclusion

While the following quote is dated and probably one of the most overused in the social sciences—we can’t resist:

Nothing is so practical as a good theory (Lewin, 1945). This observation is widely accepted and valued in most social science disciplines and professions today. In this chapter, we have affirmed and extended the spirit of this observation to the emerging global transdisciplinary evaluation. Although evaluation has been frequently criticized for being atheoretical and method-driven, this chapter has illustrated how theory is playing, and should continue to play, important roles in evaluation practice. In particular, we argue that moving toward program theory-driven evaluation science will substantially increase the extent to which evaluation contributes to the purpose of knowledge generation and testing. More generally, well-supported evaluation theories, program theories, and social science theories are likely to be among the most useful and effective tools evaluators will have in their efforts to promote social betterment in the years ahead.

References


