EVALUATION FORMS

Toward an Inclusive Framework for Evaluation Practice

John M. Owen

The title and contents of this chapter should reinforce the decision to place me on the utilization branch of the evaluation theory tree. Contributing to the field later rather than earlier means that I have had the benefit of prior contributions of eminent scholars. I belong to a group of evaluators described in Chapter 1 as "evaluation interpreters and teachers," for the ideas presented here are heavily dependent on prior substantial conceptual writings of others who have contributed to this book.

The starting point for this discussion is an assumption that evaluation is undertaken to produce knowledge of use to identified audiences and program stakeholders.[1] Furthermore, a *knowledge utilization* focus in evaluation acknowledges the primacy of negotiation between evaluators and stakeholders as the first stage of evaluation, which fashions the way the rest of the evaluation will be undertaken. These are not new ideas and are the basis of accepted evaluation practice consistent with the notion of "tailored" evaluations (Rossi & Freeman, 1993, chap. 3) and the importance of evaluation for program decision makers (e.g., Alkin, 1985).

Returning to the "tree" metaphor, I believe that what follows is novel and takes me way out on a limb in terms of evaluation theory. For in this chapter,

I suggest that while the knowledge produced in any single evaluation study will be particular to the issues that need to be addressed in the evaluation, it is possible to classify the knowledge produced by evaluations into five distinct epistemological clusters. Furthermore, different styles or *forms* of evaluation produce knowledge products that are unique to each cluster.

These evaluation forms are based on different generic purpose, as follows:

- *Proactive evaluation,* which takes place before a program is designed. The purpose is to synthesise knowledge for decisions about how to best develop a program in advance of its planning and implementation.
- *Clarificative evaluation,* which takes place early in the delivery of a program. The purpose is to provide knowledge that identifies and documents the essential dimensions of a program to make them explicit to stakeholders.
- *Participatory/Interactive evaluation,* which takes place during the delivery of a program. The purpose is to provide knowledge for decisions related to continuous improvement by involving program providers in the evaluation process.
- *Monitoring evaluation,* which takes place over the life of a program that is well established and ongoing. The purpose is to provide knowledge to check that the program is "on track" and to provide a basis for its refinement.
- *Impact evaluation,* which is used to assess the impact of a settled program. The purpose is to determine the effects of the program in terms of the criteria selected to judge its success.

This chapter describes the development of a metamodel of evaluation based on these purposes. In keeping with the theme of this book, the next section of the chapter identifies the roots of this development. Later sections provide an extended description of the forms, and links between the forms and other conceptualisations of evaluation practice.

GENESIS OF THE FORMS

About 15 years ago, the Centre for Program Evaluation (CPE) at The University of Melbourne began a graduate teaching program in evaluation, the first of its kind in Australasia. The program was a response to a need for

knowledge and skills in evaluation to those wanting to commission or undertake evaluations in organisational settings. The typical student entering this program already had an undergraduate degree, some years of experience in middle management, and had recently been given responsibility for evaluation in the organisations in which he or she was employed. Enrolling in the program signified a change in career direction for the student.

The range of evaluation-based work experiences that these students brought to class varied enormously. At this stage, the academic staff of the CPE (the CPE team) had a good knowledge of the works of notable writers in evaluation, particularly those that Shadish refers to as "Stage 1 theorists," those interested in "bringing truth to social problem solving" (Shadish, Cook, & Leviton, 1991). At the time, the CPE team had extensive expertise in the social science research methods and experience in undertaking evaluations that concentrated on outcomes or impact (process-outcome studies).

However, working with our students, we began to find that the range of their knowledge needs could not always be satisfied by the existing knowledge base of the CPE team. One reason was that the students were thinking about and undertaking evaluations within the organisations in which they worked. These evaluations had an internal focus and reflected the need for information that could be used in context. Often, a decision needed to be made under pressure of time. Associated with time-related decisions, many interventions, policies and programs, and activities were not mature or established. The CPE team had been relying on theory that assumed that evaluation had to do with assessing the worth of mature programs. This was clearly not the situation for most of the evaluative decisions that were faced by our students.

In Chapter 2, the influence of Michael Scriven on valuing in evaluation has been acknowledged. For Scriven, and Deborah Fournier (Fournier, 1995), the judgement of the merit or worth of an evaluand IS evaluation. An important outcome of theoretical writing on valuing was the explication of "evaluation logic" (Scriven, 1980). This logic involved the selection of criteria, the construction of accepted levels of performance, measurement of actual performance and comparison with accepted level, and synthesising the evidence into a judgement of worth. Usually, this logic assumes that the program is mature, and that it is up to the evaluator to act independently of program stakeholders in determining program worth.

If one takes the view that this evaluation logic should underlie all evaluation practice, it is limiting. Scriven himself acknowledged this at an
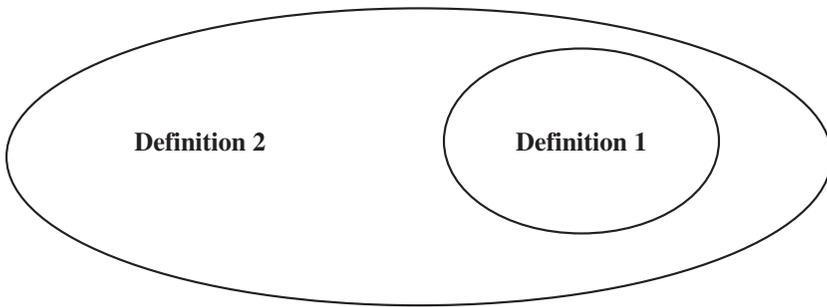
Figure 32.1   Definitions of Evaluation

international meeting of evaluators in Vancouver in 1995, where he made the distinction between evaluation based on the principles described above and "what evaluators do."

My concern was that as an evaluation theorist who needed to be grounded in practice and provide relevant conceptual perspectives to our students, more theoretical attention needed to be given to "what evaluators do." This was the genesis of a framework that acknowledges the real-world diversity of evaluation practice.

DEFINING EVALUATION MORE BROADLY

As a first step, there was a need to expand the definition of evaluation over and above the judgement of worth. My view was that was evaluation should be seen as a knowledge production activity, bounded by the need to provide program-related information to identified audiences who had a legitimate interest in that program. This expanded role of evaluation over that of providing information about the worth of a program is depicted in Figure 32.1.

Figure 32.1 shows that judgement of program worth is just a part, albeit an important part, of what evaluators do. An important element of both definitions is that the knowledge produced has to be drawn from systematic enquiry; that is, it must be the result of the application of the canons of social science research.

The judgement of worth perspective seems to be based on several assumptions about the creation and use of programmatic interventions in a

sociopolitical milieu. The first was that these interventions could and should be developed in a trial or prototype format and that impact evaluations would find out whether or not they worked. Then, those that did work would be disseminated across social systems, leading to widespread amelioration of the problem this program was meant to address. This implies that the development of social interventions was in the hands of central authorities and consistent with a Research and Development perspective on change and improvement. It implied also that the social system would be prepared to wait until programs could be scientifically validated. Valued knowledge was that which would be produced by procedures that originated in the physical sciences and relied on the application of the scientific method.

These assumptions became the subject of disputes among evaluators and social planners from the beginning of the 1970s, and the arguments are well-known. However, one point of agreement was that social interventions are not "black boxes" and implementation cannot be taken as a given. Also, it is now more or less agreed that there is far more devolution of responsibility for policymaking and program development to local authorities in most Western countries. This makes the notion of trialing and evaluating centrally developed interventions, which are then used universally across a social system, an anachronism.

An important outcome of these developments meant that more attention needed to be given to the information needs of those who had responsibilities for planning and implementing social programs, *as they were being delivered*. Some evaluators moved to provide this information. While impact evaluations cannot, by their nature, have an effect on the program being delivered, other kinds of evaluation were designed to interact with the program as it occurred. For me, this is one of the defining distinctions in the expansion of evaluation theory and practice during the past 30 years.

Other chapters in this book have addressed the expansion of this practice. A result has been a proliferation of "evaluation models," each of which is generally associated with a well-known North American evaluator. These have been admirably reviewed in a recent volume of *New Directions for Evaluation* (Stufflebeam, 2001).

Ironically, the expansion of theories and the practice of evaluation meant that neophyte evaluators, such as those who enrolled in the program offered by the CPE evaluation team, had great difficulty in marrying these models to the practical needs for evaluation in their workplaces. For me, there was a

challenge to provide some overarching structure to a theoretical world that seemed full of complexity. One of the dangers was that new practitioners would latch on to one of these models and use it no matter what the circumstances, and regardless of the lack of congruence between what the model proposed and the knowledge needs of stakeholders. In the next sections, I outline the developments of the metamodel of evaluation that was designed to provide a solution to these concerns.

SALIENCE OF EVALUATION NEGOTIATION

Earlier in this chapter, I argued that the most appropriate definition of evaluation is in terms of the production of knowledge that is useful to identified audiences for the purpose of decision making. How does one encourage use by these stakeholders? My view is that use is considerably enhanced by extended up-front negotiation with stakeholders before any evaluative data are collected (Owen, 1998). Involvement of stakeholders is now recognised as an essential aspect of evaluation practice in the context of large-scale policy research (e.g., Weiss, 1983) and in smaller-scale program evaluations (e.g., Guba & Lincoln, 1989).

A *knowledge utilization* definition of evaluation (as distinct from an evaluation logic definition) involves three stages:

- The negotiation of an evaluation plan with key stakeholders
- The implementation of an evaluation design that acknowledges the plan
- Dissemination of evaluation findings to audiences identified during the negotiation (see Owen & Rogers, 1999)

The importance of this view of evaluation is that utilization issues need to be thought about at the very beginning of an evaluation. Negotiation has been defined in the conflict resolution literature as "a discussion between two or more parties with the apparent aim of resolving a divergence of interest and thus escaping social conflict (Pruitt & Carnevale, 1993, p. 27). Levinger and Rubin (1994) define negotiation as a "process where two or more interdependent parties use the give and take offers and counter-offers in an effort to build a mutually acceptable settlement" (p. 23). This involves the evaluator or

evaluation team in working with identified stakeholders to set the parameters of what will follow. A major element of this negotiation is to determine what is possible given limitations set by factors such as the budget, by the timing of the evaluation in relation to program delivery, and by ethical considerations.

EVALUATION FORMS: TOWARD A METAMODEL OF EVALUATION

The metamodel I flagged earlier in this chapter was designed, in the first instance, to be used in the negotiation phase of an evaluation. I saw it as an educative tool, to provide some order into the range of evaluation concepts and models that existed. I drew on existing literature, meetings with key theorists, and information obtained at evaluation conferences (attendance led to some notable "aha" moments). I found that the closest existing framework to the one that was needed was the context input process and product (CIPP) model, developed by Daniel Stufflebeam (Stufflebeam, 1983). This had a major influence on my thinking, combined with a detailed knowledge of the range of "grounded" evaluation work being undertaken in Australasia.

Table 32.1 provides an overview of the metamodel, which includes generic kinds or *forms* of evaluation. As indicated earlier, they are titled "Proactive," "Clarificative," "Participatory/Interactive," "Monitoring," and "Impact." The framework assumes that the broad sweep of evaluation practice or "what evaluators do" can be summarised by these broad epistemologies. In addition to purpose, each of them has each its own characteristics, assumptions, and imperatives. Furthermore, it is assumed that the application of a given form (or forms in some cases) to an evaluation will encourage the utilization of the findings in a given organisational context.

Let me take an example. The most familiar form of evaluation for most readers is the impact form, in the right-hand column of Table 32.1. As I discussed earlier in this chapter, impact evaluations can be defended on the basis of the need to find out whether interventions work and why they work. In knowledge utilization terms, the imperative is that evaluation can lead to two kinds of use. First, the information can be used to give us confidence that the same or similar programs can be implemented with a high chance of success in a similar context. Second, evaluations can contribute to the "funded knowledge," the accumulation of knowledge about a policy or general

Table 32.1   Forms of Evaluation

Form	*Proactive*	*Clarificative*	*Participatory/ Interactive*	*Monitoring* Checking/ Refining	*Impact* Learning/ Accountability
Purpose	*Synthesis*	*Clarification*	*Improvement*		
Assumption	What is already known should influence action.	Program rationale and design needs to be laid out.	Those close to action need information for ongoing change.	Programs need to be monitored to ensure quality.	Need to know what works and why.
Imperative	Importance of external frame of reference	Importance of making intervention explicit	Importance of provider involvement	Importance of quality control	Importance of transfer-ability: contribution to funded knowledge

phenomenon, for example, the literature on effective training. There is also often an accountability focus underlying impact evaluations, to show that funds spent on them have been used wisely.

It is worth commenting on the purposes, assumptions, and imperatives of the other forms briefly. To the right in Table 32.1 is the Monitoring Form. This is based on a management view of evaluation practice in which ongoing programs are monitored regularly over time or across sites. This form is currently strongly favoured by government in countries such as Australia and the United States. There is a strong emphasis on the delivery of "quality" programs, and the use of performance measurement is seen as one way to ensure quality.

The Participatory/Interactive Form takes a diametrically opposite view in that it relies on ongoing cooperative program development within organizations and a symbiotic relationship between developers and evaluators. The evaluators' work can spread over into program decision making. The act of the evaluation can be just as important as the findings, and there is evidence that process use has a beneficial staff development effect on program staff involved.

The Clarificative Form takes the view that evaluators provide a service to program managers and deliverers by exposing the key features of the intervention. Employment of the form is based on the need for all stakeholders to have an explicit rather than tacit knowledge of program plans and how they should

or are being implemented. Use of this form is likely to be most useful at an early stage of program delivery.

The Proactive Form encourages evaluators to provide relevant knowledge to program deliverers before or in conjunction with program planning. The form is based on the assumption that what is already known about the problem to be ameliorated by the intervention should be used. For those using this form, an imperative is to look "outside" for evidence on which to make program decisions.

While Table 32.1 provided a conceptual basis for the forms, there was also a need to make them operational. A question was: What elements would be needed in the metamodel to make it useful for evaluators when negotiating with stakeholders? Experiences with negotiation in evaluation studies by the CPE team had shown that evaluations progressed most effectively and efficiently when the questions or issues that guided the study were made clear. This conclusion has backing in the evaluation literature; for example, studies have shown that the nature of the question should determine the types of data and analyses that are chosen (Smith, 1992).

Table 32.2 provides an expanded version of the metamodel. The key elements are as follows:

- Typical issues or questions that are asked within each Form
- Examples of methods that could be used to answer the questions chosen to structure the evaluation
- Examples of existing or emerging evaluation models that have been assigned to each Form

In the Proactive Form, for example, the issues reflect the need to provide evidence to assist in the development of policies and programs. Thus, there is an emphasis on determining what is already known about the problem and whether solutions have already been discovered. Three approaches that have been used to address these issues are needs assessment, research review, and review of best practice or benchmarking. In passing, it should be noted that while needs assessment theory and practice have firm roots in the evaluation community through the contributions of Belle Ruth Witkin, James Altschuld, and others, the other approaches have been sourced from the broader social science and management practice. It should be noted that there is no necessity to use just one of these approaches; indeed, one of the advantages of the

Table 32.2   Forms of Evaluative Enquiry: Issues, Methods, and Approaches

Form	Proactive	Clarificative	Participatory/ Interactive	Monitoring	Impact
Purpose	Synthesis	Clarification	Improvement	Checking/Refining	Learning/ Accountability
Typical issues	• Is there a need for the program? • What do we know about this problem that the program will address? • What is recognised as best practice in this area? • Have there been other attempts to find solutions to this problem? • What does the relevant research or conventional wisdom tell us about this problem? • What could we find out from external sources to rejuvenate an existing policy or program?	• What are the intended outcomes, and how is the program designed to achieve them? • What is the underlying rationale for this program? • What program elements need to be modified to maximise the intended outcomes? • Is the program plausible? • Which aspects of this program are amenable to a subsequent monitoring or impact assessment?	• What is this program trying to achieve? • How is this service going? • Is the delivery working? • Is delivery consistent with the program plan? • How could delivery be changed to make it more effective? • How could this organisation be changed so as to make it more effective?	• Is the program reaching the target population? • Is implementation meeting program benchmarks? • How is implementation going between sites? • How is implementation now compared with a month ago? • Are our costs rising or falling? • How can we fine-tune the program to make it more efficient? • Is there a program site that needs attention to ensure more effective delivery?	• Has the program been implemented as planned? • Have the stated goals of the program been achieved? • Have the needs of those served by the program been achieved? • What are the unintended outcomes? • Does the implementation strategy lead to intended outcomes? • How do differences in implementation affect program outcomes? • Has the program been cost-effective?

(Continued)

Table 32.2 (Continued)

Form	Proactive	Clarificative	Participatory/ Interactive	Monitoring	Impact
Purpose	Synthesis	Clarification	Improvement	Checking/Refining	Learning/ Accountability
Methods	Review of documents and databases, site visits, and other interactive methods. Focus groups, nominal groups, and delphi technique useful for needs assessment.	Generally relies on combination of document analysis, interview, and observation. Findings include program plan and implications for organisation. Can lead to improved morale.	Relies on intensive on-site studies, including observation; degree of data structure depends on approach. May involve providers and program participants.	Systems approach requires availability of management information systems (MIS), the use of indicators, and the meaningful use of performance information.	Traditionally requires use of preordinate research designs, where possible, the use of treatment and control groups, and the use of tests and other quantitative data. Studies of implementation generally require observational data. Determining all the outcomes requires use of more exploratory methods and the use of qualitative evidence.
Key approaches	• Needs assessment • Research review • Review of best practice (benchmarking)	• Evaluability assessment • Logic/theory development	• Responsive • Participatory • Quality review • Developmental • Empowerment	• Component analysis • Devolved performance assessment • Systems analysis	• Objectives based • Process-outcome studies • Realistic • Goal-free • Performance audit

metamodel is that potential users see the "bigger picture," in which the major purpose of the evaluation exercise is paramount.

DISCUSSION

A major aim of the development of the theory behind the forms was to explicate and categorise the range of evaluative activity that is practised around the globe. Each form has its own adherents and critics, and it is possible to identify key contributors, for example, Joseph Wholey in Clarificative, through his work on evaluability assessment; Bradley Cousins in Participatory/Interactive, through the participatory approach; and Harry Hatrey, on Monitoring, through performance assessment. These theorists maintain they are undertaking evaluative activity, yet up to now, we have lacked a metamodel that is inclusive of them all and allows them to be compared across common dimensions.

The metamodel is open to criticism in that within a given form, key distinctions between models or approaches might seem to be blurred. For example, in the Participatory/Interactive Form, Robert Stake would say that his responsive approach differs from Michael Patton's developmental approach in terms of the extent of involvement of the evaluator in decision making and in the choice of methodology.

My reaction to this is that the commonalities outweigh the finer details and that the general thrust of these approaches is sufficiently similar to cluster them together under the one purpose. In the first instance, clients wishing to commission an evaluation can appreciate the big picture through the use of the tables outlining the forms. However, my experience is that some clients are interested in differences between approaches *within* one form, in which case we have taken the opportunity to explain these differences as a means of firming up the evaluation design during the negotiation stage of the evaluation.

A key evaluation viewpoint presented here is that evaluation needs to be seen as a complement to effective program delivery. As such, it is a service to decision makers and is unashamedly "clientist," in Ernie House's terms (see Chapter 2). Evaluation findings need to provide relevant and timely information that can be applied to the program under review. This position has its roots in the Australasian context, where research practices continued to be employed that were remote from the real-life concerns of policymakers and providers,

even when they were labelled as "evaluations." Evaluations of this kind borrowed heavily from acceptable social science methods, and there was a reluctance to use or invent methods that were more likely to provide answers to questions that were central to program decision makers.

This concern is one of the reasons for promoting, through the forms, an emphasis on evaluation activity toward the "front end" of program decision making. This places more emphasis that conventional theorists would accept on models and methods in the Proactive and Clarificative Forms, which reflect a kind of quality assurance perspective to evaluation. It seems that we need an alternative term to evaluation to encompass the range of activities that "evaluators do," for example, *evaluative enquiry*.

A second key viewpoint is the pluralism in evaluation. Evaluative enquiry can no longer be presented to the client as a unitary concept. The forms outline five evaluation epistemologies, and it would be satisfying if theorists in the professional evaluation community were to acknowledge this diversity of practice. An implication is that evaluators should be willing to entertain evaluations that are framed according to the issues or questions within a given form that are of salience to the clients.

In this sense, the tables are deliberately arranged to reflect a hierarchy of elements, with purpose at the top and methods farther down. This is meant to signify that methods are just tools to meet the purposes and key issues of the evaluative enquiry and should be selected only after these issues have been developed.

This chapter has provided me with an opportunity to reflect on the development of the forms for a recent text on evaluation (Owen & Rogers, 1999). I noted earlier about going out on a limb to develop this view of evaluation practice. The fact that I was placed on the tree at all is heartening, as have been reactions to the forms by valued colleagues who have theoretical interests in evaluation and the students who provided the impetus for the conceptual development underlying the metamodel.

NOTE

1. "Program" is used throughout generically to describe the evaluand, which could be a policy, program intervention, or organization.

REFERENCES

Alkin, M. C. (1985). *A guide for evaluation decision makers*. Beverly Hills, CA: Sage.

Fournier, D. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions in Program Evaluation, 68,* 15-31.

Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. Newbury Park, CA: Sage.

Levinger, G., & Rubin, J. (1994). Bridges and barriers to a more general theory of conflict. *Negotiation Journal, 10*(3), 20-35.

Owen, J. M. (1998). Toward a theory of negotiation in evaluation. *Evaluation News and Comment, 7*(2), 32-35.

Owen, J. M., & Rogers, P. (1999). *Program evaluation: Forms and approaches* (2nd ed.). London: Sage.

Pruitt, D. G., & Carnevale, P. J. (1993). *Negotiation in social conflict.* Buckingham, UK: Open University Press.

Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach.* Newbury Park, CA: Sage.

Scriven, M. (1980). *The logic of evaluation.* Port Reyes, CA: Edge Press.

Shadish, W. R., Cook, T., & Leviton, L. (1991). *Foundations of program evaluation.* Newbury Park, CA: Sage.

Smith, N. L. (1992). Aspects of investigative inquiry in evaluation. *New Directions in Program Evaluation, 56,* 3-13.

Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 117-141). Boston: Kluwer-Nijhoff.

Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation, 85,* 8-98.

Weiss, C. (1983). The stakeholder approach to evaluation: Origins and promise. *New Directions in Program Evaluation*, *17,* 3-14.