Improving Evaluation Theory Through
The Empirical Study of Evaluation Practice

NICK L. SMITH

Empirical knowledge about the practice of evaluation is essential for the development of relevant and useful evaluation theories. Although such studies are too seldom conducted, they provide a necessary connection between theory and practice.

Evaluation theory and practice have been linked in a variety of ways since the beginning of the field. Difficulties in responding to early mandates for evaluation gave rise to the development of alternative evaluation models. As new problems, contexts, and participants influenced the nature of evaluation practice, continual modifications were made to the emerging evaluation models, theories, and methods. Certain theoretical issues have been identified and developed as a direct result of practical problems. For example, theorists studied factors influencing evaluation utilization as a result of practical difficulties with the acceptance and implementation of evaluation recommendations. Similarly, practical problems in client/stakeholder communications and the use of evaluation results in policy formation have resulted in increased attention to theoretical issues in those areas.

Questions about the relationships between evaluation theory and practice have often arisen during metaevaluations. For example, metaevaluations of highly visible, national programs such as Headstart, Follow Through, Push/Excel, and Cities-in-Schools, have illustrated the problems of translating current evaluation theory into acceptable evaluation practice. But even the writing on metaevaluation has been largely methodological (e.g., Cook and Gruder, 1978), or with few analytic or comparative studies of actual metaevaluations (cf. Smith, 1981a, 1990). Although metaevaluations are an important source of information about the relationships between theory and practice, they tend to
be narrowly-focused analyses of the quality of specific evaluation studies, rather than examinations of evaluation practice across collections of related studies, for example, metaanalyses of metaevaluations. Evaluations of evaluation practice are needed, but so are research studies of evaluation practice.

There has been little response to the repeated calls over the past 20 years (see Worthen, 1990; Scriven, 1991; Smith, in press; for brief reviews) for increased empirical study of evaluation practice to describe the nature of actual practice; to compare the feasibility and effectiveness of alternative models, methods, and theories; to provide a basis for the development of descriptive evaluation theories; and to assess the utility of prescriptive theories. Much of the current empirical justification of evaluation theories is from self-reported cases of what works. Shadish, Cook, and Leviton (1991) summarize the problem as follows:

To the extent that evaluation theory is still conceptual, a move toward citation of cases to support and illustrate hypotheses is welcomed if for no other reason than to provide these preliminary benefits. In the long run, however, evaluation will be better served by increasing the more systematic empirical content of its theories. . . (p. 483). Such efforts have always been relatively rare in evaluation because so little effort is generally put into developing empirically testable hypotheses based in evaluation theory, and because so few evaluators are both interested in the topic and in a position to undertake such studies (p. 484).

In spite of the continued scarcity of empirical studies of evaluation practice, such studies are so crucial to the future development of evaluation theory that yet another call for such studies seems warranted. Perhaps a few examples will illustrate the utility of such studies.

ILLUSTRATIVE EMPIRICAL STUDIES OF PRACTICE

Alkin (1991) studied the factors that influence evaluation theorists to change their conceptions of their own models or approaches. He reports that two primary factors are (a) increased experience in actually conducting evaluations, and (b) accumulation of personal research on evaluation practice. Consider the following examples of studies of evaluation practice and how their findings are relevant to the revision of evaluation theories.

- In the United States in the 1960s, the initial impetus and financial support for evaluation came from the federal government. For several years, many evaluation theorists seemed to assume that local level evaluations were in fact only conducted in response to federal mandates and with federal flow-through funds. In 1978, however, Lyon et al. (1978) conducted a study of local level evaluation which showed that fully 65 percent of local evaluations were locally initiated and supported.

- In 1980, a study by Boruch and Cordray (1980) showed that only 34 percent of the evaluations conducted by the U.S. Office of Education in 1977-1979 focused primarily on estimating the effects of programs on major target groups, in spite of the widespread advocacy and apparent agreement on the necessity of outcome evaluation at that time.
Patton's (Patton et al., 1977; Patton, 1986) study of the actual use of evaluation recommendations by decisionmakers in federal agencies appears to have had a significant influence on the development of enlightenment, as contrasted with instrumental, theories of evaluation use.

Smith (1985) reviewed every application of the adversarial/judicial/committee hearings approach to evaluation over a fifteen-year period, finding common problems and limitations across the various studies. There appears to have been little use of these approaches in recent years, perhaps due to general recognition of the practical difficulties and limitations of these methods.

In a 16-state study of accreditation evaluations, Smith (1980a) found that, contrary to the assumptions of stakeholder-oriented theorists, many stakeholders did not want to be involved in evaluation planning or implementation, and did not feel that their participation would improve the quality or utility of evaluation. These findings addressed an ongoing discussion between Smith (1980a, 1980b) and Patton (1978, 1986, 1987) over the proper role of stakeholder involvement in evaluation. Subsequently, Smith (1983) published a report of four empirical studies of evaluation practice which investigated theoretical assumptions held about the use of citizen judgements in program evaluation.

The articulation of a theoretical position into actual practice can reveal hidden problems in the theory. For example, Smith (1990) studied Stake's use of the naturalistic case study approach in conducting a metaevaluation and illustrated how the approach is susceptible to a subtle form of evaluator bias. Such bias may be a prevalent problem in other interpretative evaluation approaches which have not developed the bias control mechanisms of such interpretive inquiries as, for example, psychotherapy.

How practitioners view their own practice is certainly relevant to the development of more realistic and applicable theory. Useful descriptive studies of how practitioners and theorists view their own practice have been provided by Shadish and Epstein (1987) and Williams (1989), respectively. Such information greatly facilitates the interpretation of subtle points of theory and practice. While informative, these studies share the limitation of several previously cited studies that rely on self-report measures of evaluation practice.

Finally, Smith, and Mukherjee (1990) analyzed the questions addressed in all health and education evaluations published in the 12 volume Evaluation Study Review Annual. They illustrate how evaluators in the two fields of evaluation seem to be addressing generically different types of evaluation questions, and how evaluators appear to shape policy questions into causal/comparative questions.

This is neither an exhaustive nor exemplary list of empirical studies of evaluation practice, there are other examples. These are merely illustrative examples with which I am most familiar.

FROM STUDIES OF PRACTICE TO IMPROVEMENT OF THEORY

Through studies of evaluation practice we can accumulate accurate knowledge to replace widespread supposition and overgeneralization of local, context-specific knowledge (see
Smith, 1979). Such studies are needed to acquire knowledge of what works and how to make it happen, how to avoid bad practice, how local contexts differ and what works across different contexts, and where problems or inadequacies of evaluation practice could be ameliorated by improved evaluation theory.

Theorists present and advocate theories largely in abstract conceptual terms, seldom in concrete terms based on how the theories would be applied in practice. We need to know how practitioners articulate or operationalize various models or theories, or whether, in fact, they actually do so. Indeed, it is not clear what is meant when an evaluator claims to be using a particular theoretical approach.

Although evaluators sometimes speak of designing evaluations which follow a particular model, their language generally refers not to instrumental application of procedural specifics, but to the selection of an overall orientation or approach. Because evaluation models are not procedurally prescriptive, are subject to varied interpretations, are mute on many of the details required to implement an evaluation, and must be operationalized within the demands of a specific context of application, many decisions are left to the evaluator's professional judgment in spite of the prior selection of a given model. No one study can thus be argued to be the epitome of a given model, and many quite different studies are arguably appropriate versions of the same model (Smith, in press, p. 4).

If evaluation theories cannot be uniquely operationalized, then empirical tests of their utility become increasingly difficult. If alternative theories give rise to similar practices, then theoretical differences may not be practically significant. There is a need to identify which theoretical claims in fact presuppose testable empirical facts.

Studies of practice can provide the empirical basis on which to develop descriptive theories of evaluation practice; i.e., theories that describe what evaluators do, why they do it (in terms of both their rationale and the contextual forces which shape their behavior), and what use, impact, or change results from their actions. Studies of practice can also be used to assess prescriptive theories of evaluation when the prescriptions are based on claims or assumptions about the results of doing an evaluation the prescribed way (e.g., utilization of results will be increased), as opposed to prescriptive theories derived from a formal metatheory or from normative claims about proper evaluation practice. (For example, Scriven (1991) argues that evaluation necessarily involves the assessment of value. Whether empirical studies of evaluation practice confirm that evaluators actually do assess value is irrelevant to Scriven's metatheoretical claim that they must do so in order logically to be engaged in the act of evaluating.) Shadish, Cook, and Leviton (1992) describe their five component program evaluation theory (see Shadish, Cook, & Leviton, 1991) as having both descriptive and prescriptive elements, and Shadish (1992) has identified a lengthy set of empirical questions that could be studied in order to test, refine, and support each of the five components of their theory: social programming, knowledge construction, valuing, knowledge use, and evaluation practice.

Empirical knowledge of practice is necessary for further development of what Shadish, Cook, and Leviton (1991) consider to be the strongest type of evaluation approaches, contingency theories. Contingency theories, (in contrast with focusing theories that prescribe practice that does not vary over evaluands, contexts, or evaluators) seek to specify the conditions under which different evaluation practices are effective. Shadish et al. (1991) thus judge the contingency theories of Rossi and Cronbach as stronger because they are based on accumulated information about effective practice. An example of this kind of
contingency information is seen in Smith's (1980a) study of stakeholder involvement mentioned earlier in which administrators from small school districts had greater time and desire to participate in accreditation evaluations, whereas administrators from large school districts had little time, saw their role as an oversight function, and did not wish to participate in such evaluations. Finally, information on the contingencies of effective practice would seem to be essential in the further development and application of the multiplist approaches of Scriven (1983), Cook (1985), and others, in which the creation of an evaluation study entails selection from among multiple approaches, designs, techniques, roles, etc.

If multiplist and contingency approaches reflect the next steps in the development of evaluation theory, then theoretical development must necessarily be more closely connected to empirical studies of practice.

A variety of types of studies should be considered in this research, including metaevaluations, descriptive studies of what is actually done in practice and its relationship to relevant theories, comparative analyses of theoretical and methodological alternatives, studies of fundamental assumptions underlying dominant theories and modes of practice, and so on. These various types of studies are illustrated in the examples provided earlier (also, see Smith, 1979, 1980b, 1981b, 1982, for more extended discussions). Whatever the form these empirical studies take, evaluation no longer has the luxury of a-empirical theoretical development.

Evaluation practitioners should continue to be publicly reflective about their own practice, but evaluation researchers need to focus increased effort on the independent, empirical study of the practice of evaluation.

NOTES

1. Additional evidence of the general lack of interest in empirical studies of evaluation practice can be seen in comments by reviewers when this article (Smith & Mukherjee, 1990) was submitted for publication. A reviewer for Evaluation Review suggested that the types of questions evaluators address is not of particular importance, and a reviewer for Educational Evaluation and Policy Analysis commented that few readers of that journal are interested in research on evaluation.

2. See Shadish, Cook, and Leviton (1991, esp. pp. 477-484) for a similar argument to the one I make here including citation of other empirical studies of evaluation practice and a summary of important theoretical issues needing empirical study.

REFERENCES


