Some Evaluation Questions

By William Shadish

This digest was adapted with permission from Shadish, W. (1998) Presidential address: Evaluation Theory is Who We Are. American Journal of Evaluation, 19, 1, 1-19.

Introduction

In question and answer format, this digest illustrates the variety of basic and theoretical issues in evaluation with which aspiring evaluators should be conversant in order to claim they know the knowledge base of their profession. lease note that none of these questions have a single correct answer and space limitations prevent providing the level of detailed discussion that each deserves. The questions vary considerably in difficulty and in how universally the issues involved would be recognized by most evaluators today. What follows, therefore, are outlines of the issues rather than correct "answers." For more extensive information on the topics discussed in these questions, please refer to the references found at the end of this digest, especially Shadish, Cook, and Leviton (1991).

What are the four steps in the logic of evaluation?

Scriven (1969, 1989) published a variety of writings on the topic of the logical sequence of concepts that defines how people try to connect data to value judgments that the evaluand is good or bad, better or worse, passing or failing, or the like. Scriven outlined the four steps in 1980:

1. selecting criteria of merit, those things the evaluand must do to be judged good
2. setting standards of performance on those criteria, comparative or absolute levels that must be exceeded to warrant the appelation "good"
3. gathering data pertaining to the evaluand's performance on the criteria relative to the standards
4. integrating the results into a final value judgment.

To the extent that evaluation really is about determining value, some version of this logic ought to be universally applicable to the practice of evaluation.

Are qualitative evaluations valid?

More qualitative theorists than not seem to both use the word, subject to validity criticism, and endorse some version of its applicability. However, some qualitative theorists reject both the term and any cognates who seem to garner attention disproportionate to their representation in their own field. From outside the qualitative camps, the answer also seems to be more uniformly "yes." Nevertheless, the subtleties required for an intelligent discussion of this question are extensive, of which the following few will illustrate but not exhaust. Even those who reject the concept of "validity" will acknowledge they are concerned in their work to "go to considerable pains not to get it all wrong" (Wolcott,1990, p. 127). Further, within and across those methods qualitative theorists often disagree among themselves. In addition, qualitative methods often aim to produce knowledge of a substantively different kind than other methods, so that particular validity criteria may be less pertinent to the interests of qualitative evaluations. Indeed, it would be wrong to assume all qualitative methods are alike. Different qualitative methods may have different aims that bring different validity criteria to bear. In the end, though, some version of validity as an effort to "go to considerable pains not to get it all wrong"(Wolcott, 1990, p. 127) probably underlies all methods used by all evaluators, quantitative and qualitative.

What difference does it make whether the program being evaluated is new or has existed for many years?

Rossi and Freeman (e.g., 1993) long made this distinction central to their approach to evaluation because it has several implications for evaluation practice. Brand-new programs have not yet had time to work out program conceptualization and implementation problems. Thus, a focus on those kinds of questions is likely to be more useful and more acceptable to program staff than a focus on, say, outcome questions. In addition, less background information and fewer past evaluations are likely to exist for new programs, so more work will have to be done "from scratch." Well-established programs may be more ready for outcome evaluation, and they may have a greater wealth of information already available on them. However, long-established programs may also have reached so many of the potential participants that outcome evaluations might be thwarted by difficulty finding appropriate control group participants if a controlled design is used.

Is there a difference between evaluating a large program, a local project within that program, or a small element within that project?

This distinction points to an interesting tradeoff between ease and frequency of short-term change on the one hand, and likely impact on the other (Cook, Leviton, & Shadish, 1985; Shadish, Cook, & Leviton, 1991). Small elements have natural turnover rates that are much more frequent than for local projects, which themselves turnover less often than large programs. Hence, the opportunity to change each of them by replacement is more frequent for smaller than larger entities. However, smaller entities are usually likely to have a smaller impact on the overall set of problems to which the program, project, or elements are aimed. All this has implications for the kinds of questions worth asking depending on what kind of use and impact is desired.

How can the chances of evaluation results being used in the short-term to make changes be increased?

The literature on this topic is extensive (e.g., Cousins &

Shuhla, 1997; Patton, 1997), and includes advice to locate a powerful user(s), identify questions of interest to the user(s), focus on things that the user has sufficient control over to change, discuss exactly what changes the user(s) would make given different kinds of answers that might result from the evaluation, provide interim findings at points when they might be useful, consider reporting results in both traditional and nontraditional formats, provide brief executive summaries of results, have continued personal contact after the evaluation ends, and lend support to subsequent efforts to foster use of evaluation results.

What are the disadvantages of focusing on short-term instrumental use?

There is a risk that the evaluation will focus on less important interventions or questions than might otherwise be the case, and lose the big picture or the long-term outlook about what is important. In part this reflects the tradeoffs discussed regarding the program-project-element distinction because instrumental use is more likely with smaller elements likely to have less impact. It also reflects the fact that the modern industrial societies where much evaluation takes place have often solved the easiest problems, so those that remain are often difficult to do anything about in the short-term. Those things that can be addressed in the short-term are rarely likely to fall into the set of most difficult problems. Finally, it is rare to find a user who can control options that promise truly powerful or fundamental changes.

What role does causal inference play in evaluation?

The most obvious version of this question concerns the role of outcome evaluation. From an early dependency on outcome evaluation as paradigmatic for the field, the field realized the value of asking a wide array of other questions depending on contingencies like those discussed previously regarding use, program size, and stage of program development So causal inference of the traditional sort assumed a smaller role in evaluation than in early years. Another version of this question appeals to the distinction between descriptive causal inferences and causal mediation; the latter has enjoyed some recent resurgence in some kinds of theory-driven evaluation.

Would the answer change if questions were asked about the role that causal inference played in making value judgements?

Most readers probably assumed "evaluation" in the previous question to mean the wide range of activities that fall under the rubric of professional evaluation practice. This question plays on limiting the meaning of the term "evaluation" to the activity of making a value judgment. Some readers might not realize that, even in this limited context, causal inference still plays an important role. Referring back to the answer to the first question about the logic of evaluation, implicit in it in most applications is that the thing being evaluated caused the observed performance on the criteria of merit (e.g., that the treatment met recipient needs). If that were not the case, it would be improper to attribute the merit or value to the evaluand; rather, it should be attributed to whatever else actually caused the improvement in the criteria of merit. Thus attributing merit or worth is frequently causal in substantial part.

When does a question have leverage?

Cronbach and his colleagues (Cronbach, Ambron, Dornbusch, Hess, Hornik, Phillips, Walker, & Weiner, 1980) used this term to describe questions they thought particularly worth asking because of their potential for high payoff. Such questions have little prior information available, they can be feasibly answered with the resources and in the time available, the answers will probably reduce uncertainty significantly, and the answers are of interest to the policy shaping community.

What is metaevaluation, and when should it be used?

Metaevaluation is the evaluation of evaluation (Cook & Gruder, 1978; Scriven, 1969), and recommendations vary from doing it for every evaluation to doing in periodically. The general prescription is that metaevaluation can be done using the same general logic (and sometimes methods) for doing the primary evaluation. One might apply the logic of evaluation from the first question, for example, asking what the evaluation would do well to be a good evaluation (e.g., would it be useful, true, important?), deciding how well would it do so (e.g., how useful? true by what standards?),measuring the performance of the evaluation in these regards, and then synthesizing results to reach a judgment about the merits of the evaluation. Metaevaluation can be applied at nearly any stage of an evaluation, from evaluating its planned questions and methods, to a mid-evaluation review, to evaluating the completed report by submitting it to independent consultants and critics.

References and Additional Reading

Campbell, D. T. (1971). *Methods for experimenting society.* Paper presented at the meeting of the Eastern Psychological Association, New York and at the meeting of the American Psychological Association, Washington, DC.

Cook, T. D. and Gruder, C. L. (1978). Metaevaluative research. *Evaluation Quarterly, 2,* 5-51.

Cook, T. D., Leviton, L. C., and Shadish, W. R. (1985). Program Evaluation. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (3rd ed., pp. 699-777). New York: Random House.

Cronach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D.C., Walker, D. F., and Weiner, S. S. (1980). *Toward Reform of Program Evaluation.* San Francisco: Jossey Bass.

Lincoln, Y. S. (1990). Program Review, Accreditation processes, and outcomes assessment: Pressures on institutions of higher education. *Evaluation Practice, 11*(1), 13-25.

Maxwell, J. A. (1992). Understanding validity in qualitative research. *Harvard Educational Review, 62*(3), 279-300.

Patton, M. Q. (1997). *Utilization-focused Evaluation: the New Century Text.* Thousand Oaks, CA: Sage Publications.

Rossi, P. H. & Freeman, H. E. (1993). *Evaluation: a Systemic Approach*. Newbury Park, CA: Sage Publications.

Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report, 2,* 36-38.

Scriven, M. (1980). *The Evaluation of College Teaching.* Syracuse, N.Y.: National Dissemination Center, Syracuse University School of Education.

Shadish, W. R., Cook, T. D., & Leviton, L. D. (1991). *Foundations of Program Evaluation: Theories of Practice.* Newbury Park, CA: Sage Publications.

Wolcott, H. F. (1990). *Writing up Qualitative Research.* Newbury Park, CA: Sage Publications.

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract RR93002002. The opinions expressed in this report do not necessarily reflect the positions or policies of OERI or the U.S. Department of Education. ***Permission is granted to copy and distribute this ERIC/AE Digest.***