Encyclopedia of Evaluation

Evaluability Assessment

Contributors: M. F. Smith
Editors: Sandra Mathison
Book Title: Encyclopedia of Evaluation
Chapter Title: "Evaluability Assessment"
Pub. Date: 2005
Access Date: April 06, 2015
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780761926092
Online ISBN: 9781412950558
DOI: http://dx.doi.org/10.4135/9781412950558.n177
Print pages: 137-140

http://dx.doi.org/10.4135/9781412950558.n177
Evaluability assessment was thrust on the evaluation scene in the 1970s and was initially thought to show great promise for improving programs and saving valuable evaluation resources that might have been wasted by evaluating programs that were not ready to be evaluated. After a short burst of interest and activity, the process appears to have lost much of its appeal among evaluators. This entry provides a definition of evaluability assessment and offers some conjectures as to why a tool with such demonstrated promise seems to have all but disappeared from the practice of evaluation—at least as the practice is described in published literature.

Evaluability Assessment: A Definition

Evaluability assessment (EA) is a systematic process for describing the structure of a program and for analyzing the plausibility and feasibility of achieving objectives; their suitability for in-depth evaluation; and their acceptability to program managers, policy makers, and program operators. This is accomplished by the following process:

1. Program intent is clarified from the points of view of key actors in and around the program.

2. Program reality is explored to clarify the plausibility of program objectives and the feasibility of program performance.

3. Opportunities to improve program performance are identified.

Two primary outcomes are expected from an EA:

1. *Definition of a program's theory.* This includes the underlying logic (cause and effect relationships) and functional aspects (activities and resources), with indications of types of evidence (performance indicators) for determining when planned activities are implemented and when intended and unintended outcomes are achieved.

[p. 137 ↓]

2. *Identification of stakeholder awareness of and interest in a program.* This means stakeholders' perceptions of what a program is meant to accomplish, their concerns or worries about a program's progress toward goal attainment, their perceptions of adequacy of program resources, and their interests in or needs for evaluative information on a program.

When an impact evaluation is anticipated, both of these outcomes should be attained before the evaluation is designed. When a program is being planned, or when improvement is the intent, only Outcome 1 may be pursued: Having a defined program framework increases the likelihood that program staff will manage their programs to achieve intended impacts, whether or not the impacts are to be measured. When the purpose is a preparatory step to further evaluation, these outcomes permit a clear indication of whether an intensive evaluation is warranted and, if so, what components or activities in the program can provide the most desirable data. In essence, they prevent evaluators from committing two types of error: Type III, measuring something that does not exist, and Type IV, measuring something that is of no interest to management or policy makers (Scanlon, Horst, Nay, Schmidt, & Waller, 1979).

Type III error exists when the program has not been implemented, when the program is not implemented as intended, or when there is no testable relationship between the program activity carried out and the program objectives being measured. Type IV occurs when the evaluator brings back information that policy makers and management have no need for or cannot act on. Both types of error are avoidable if an evaluability assessment is conducted. Type III errors may be avoided by defining the program and describing the extent of implementation; Type IV, by determining from the stakeholders what they consider important about the program and the evaluation.

Origin and Decline in Use

Evaluability assessment originated in the early 1970s for the purpose of improving summative program evaluations. Such evaluations, then and now, were often perceived by policy makers as expensive wastes of time that produced little in the way of timely, useful information. Evaluators, on the other hand, often found that programs had grandiose goals and few concrete objectives. This led them to produce evaluations

that angered policy makers by highlighting program deficiencies or else the evaluations were as muddled and vague as the programs. Joseph Wholey and his associates at the Urban Institute in Washington, DC, decided that an impasse had developed between stakeholders of programs and evaluators of those programs because of differences between rhetoric (i.e., claims about a program) and reality. They explored ways of aligning rhetoric with reality, and evaluability assessment was born.

For a few years the process flourished, but use dropped off dramatically after Joe Wholey left the U.S. Department of Health and Human Services. (He was the Deputy Assistant Secretary for Planning and Evaluation at the Department of Health, Education, and Welfare, the forerunner of the Department of Health and Human Services.) Debra Rog (1985) attributed this pattern of diffusion and decline in use to Wholey's advocacy: While he and the other associates who conceived the process were active in its implementation, its use increased; when they became less active, its use declined. However, it is more likely that the scarcity of concretely defined methodology had as much or more to do with the minimal adoption of EA as the advocacy of its creators. In other words, an evaluation process needs to meet the same implementation requirements as a program if it is to be successful. In this case, that would mean clearly defined outcomes for an EA and clearly defined and plausible activities (methods, steps, tasks) for reaching those outcomes—and neither of these (outcomes or methods) were products of the early work done by Wholey and his associates. They did discuss tasks that needed to be accomplished, such as document analysis, meetings of work groups, and site visits, but they did not identify specific guidelines on how to accomplish these tasks.

In 1984, Smith and her colleagues in the U.S. Department of Agriculture (USDA) initiated a project to define the evaluability assessment process in a methodological sense and to encourage adoption in the USDA's Cooperative Extension Services throughout the United States. Implementation proceeded iteratively, in a different major programming area, in five different states. After each iteration, procedures were analyzed and revised to clarify and, where possible, to simplify to make the process more usable and more “operator robust.” The outcome from the project was a set of guidelines for implementing evaluability assessments. Ten tasks were described for the production **[p. 138 ↓]** of a successful EA, and methods were defined for the accomplishment of those tasks.

Wholey and his associates had developed EA as a method for analyzing a program's structure to determine the extent to which it was suitable for effectiveness evaluation, and they later recognized it as a way for determining the extent to which a program was capable of being managed to obtain successful results. Smith (1989) demonstrated EA's contribution to program planning for developing programs capable of providing evidence of that achievement.

It is not known why the decline in published literature on the use of evaluability assessment occurred, but a number of factors may have contributed to the phenomenon:

1. *EA Unreported.* Evaluability assessments often go unreported. Most EAs lead to the conclusion that the program under review is not ready for an impact study, if the intent is to show that the program is making intended impacts (e.g., those mandated by Congress). Reports of such studies usually fail to meet the criteria for publication or are written by persons who are not interested in publishing them. Also, when the purpose is to improve a program, these types of studies are often done in house and are not reported or shared with those outside the program or organization.

2. *Methodology Unavailable.* Methodology for conducting evaluability assessments was not available early in EA's history and was not easily accessible later. As already discussed, the early promoters of EA provided scant methodology for the actual conduct of such a study, and what they did publish was not readily available (e.g., Schmidt, Scanlon, & Bell, 1979) or was hard to interpret into action steps (e.g., Nay & Kay, 1982). In addition to this early void of methodology, the first set of specific guidelines for how to conduct EA (Smith, 1989) received scant attention in the United States, partly because the publisher did not promote the text among evaluators in the United States and partly because the price of the book made it mostly unattainable ($80+ was a lot for an evaluation book in 1989). Also, the book may not have been held in high esteem because agriculture, the subject of the programs used for the methodological research, was not recognized as a leader in evaluation theory and methodology.

3. *Implementation Difficult.* Good evaluability assessments are difficult to implement effectively; the process requires much skill and experience. Nay and Kay discussed the fragility of negotiating relationships with and among policy makers, program

implementers, and other stakeholders to achieve the understandings necessary to determine what was expected and being implemented as a program, and they suggested that senior evaluators be the ones to carry out these tasks. To develop a plausible theory of program implementation, persons representing all levels and aspects of a program must be involved in describing program actions and resources, and their input must be in depth and thoughtful, revealing what is being done and why. Such descriptions are not easy to elicit because (a) they begin to reveal a person's beliefs and values (e.g., about the people they are trying to serve with the program), (b) they can reveal inadequacies of skills and work habits of staff (e.g., that something less than or different from expected is being done), and (c) many staff will think the exercise is not wise use of their time (i.e., it takes them away from their programming duties). Many hours, much savvy (program, political, and policy), and a wide repertoire of skills are needed by the evaluator to negotiate all the tasks necessary to navigate a successful EA, and not all evaluators possess the essential skills or have the interest in spending evaluation resources in this way.

4. *Name and Promotional Confusion.* The name given to the approach and its promotion during the height of interest in it suggested that EA was not an evaluation approach in its own right. The original developers of the evaluability approach promoted the process as a preliminary step to an impact evaluation, as something to do before conducting an evaluation. They described evaluability assessment as the beginning of a four-step evaluation process, to be followed by rapid feedback evaluation; performance monitoring; and, where there was sufficient program implementation, intensive study of program results. The usual pressure evaluators face for results pushes many to proceed directly to the intensive evaluation instead of doing the preliminary work.

5. *Objectivity Loss.* Evaluability assessment evaluators may lose program objectivity, in appearance or in reality. In each of the three purposes for implementing an evaluability assessment, the conductor becomes very involved in the program. In each case, involvement with the staff is intense and personal, especially as assumptions and values surface, as they **[p. 139 ↓]** must in theory definition. In later steps, conclusions of evaluability (when the purpose is summative) or recommendations for improvement (when the purpose is formative or summative) or conclusions of plausibility (when the purpose is program planning) all represent assessments of program value—all indicate that the program is worthy of investment of additional resources. Although these are

appropriate positions for evaluators to take, problems can occur when the same person who conducted the EA conducts a subsequent evaluation of program performance or impact. The problem can be a loss of credibility of evaluation results with those who are in positions to make decisions about the program.

6. *Assumption of Rationality.* Evaluability assessment is based on an underlying assumption of rationality; that is, that organizations and their programming efforts are tightly coupled and highly structured or will be at the conclusion of the assessment. It is based on a rational model of organizational decision making, with corresponding assumptions of evaluability very close to that of the problem-solving model. Further assumptions of rationality are that *the* decision makers can be identified and that programs will remain static long enough for some model of program behavior to be appropriate or measurable. In other words, programming is depicted as a deliberate process: First we think, then we act; first we formulate, then we implement. As much as some (evaluators and others) would like to see this orderly process in programs, in many practical situations the assumptions underlying the rational model do not hold; also, many evaluators refuse to accept rationality as an assumption of real-world program development or implementation.

7. The primary outputs for evaluability assessment are now being sought under separate evaluation rubrics, which are readily found in published literature. Using the metaphor of a corporate takeover, the two primary outputs of EA have been sold as program theory evaluations and stakeholder evaluations and, in so doing, have created a situation in which the whole is greater than the sum of its separate parts. The two processes practiced separately produce far less than when combined as a comprehensive evaluability assessment, in terms of (a) producing real knowledge about program intents and implementation, (b) building ownership and commitment among all levels of decision makers for creating a climate in which changes essential for program success can and will be made, (c) facilitating the ability to manage a program for success, and (d) clarifying and improving the criteria for exercising program accountability.

M. F.Smith

http://dx.doi.org/10.4135/9781412950558.n177

Further Reading

Nay, J. N., & Kay, P. (1982) Government oversight and evaluability assessment . Lexington, MA: Lexington Books.

Rog, D. (1985) A methodological analysis of evaluability assessment . Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN.

Scanlon, J. W., Horst, P., Nay, J. N., Schmidt, R. E., & Waller, J. D. (1979) Evaluability assessment: Avoiding Types III and IV errors . In G. R. Gilbert, ed. & P. J. Conklin (Eds.), Evaluation management: A selection of readings . Washington, DC: Office of Personnel Management, Federal Executive Institute.

Schmidt, R. E., Scanlon, J. W., & Bell, J. B. (1979) Evaluability assessment: Making public programs work better (Human Services Monograph Series, No. 14) . Washington, DC: Urban Institute.

Smith, M. F. (1989) Evaluability assessment: A practical approach . Boston: Kluwer.

Wholey, J. S. (1979) Evaluation: Promise and performance . Washington, DC: Urban Institute.

Wholey, J. S. (1987) Evaluability assessment: Developing agreement on goals, objectives and strategies for improving performance . In J. Wholey (Ed.), Organizational excellence: Stimulating quality and communicating value . Washington, DC: Heath.