

Introduction to Evaluation

Evaluation is a methodological area that is closely related to, but distinguishable from more traditional social research. Evaluation utilizes many of the same methodologies used in traditional social research, but because evaluation takes place within a political and organizational context, it requires group skills, management ability, political dexterity, sensitivity to multiple stakeholders and other skills that social research in general does not rely on as much. Here we introduce the idea of evaluation and some of the major terms and issues in the field.

Definitions of Evaluation

Probably the most frequently given definition is:

Evaluation is the systematic assessment of the worth or merit of some object

This definition is hardly perfect. There are many types of evaluations that do not *necessarily* result in an assessment of worth or merit -- descriptive studies, implementation analyses, and formative evaluations, to name a few. Better perhaps is a definition that emphasizes the information-processing and feedback functions of evaluation. For instance, one might say:

Evaluation is the systematic acquisition and assessment of information to provide useful feedback about some object

Both definitions agree that evaluation is a *systematic* endeavor and both use the deliberately ambiguous term 'object' which could refer to a program, policy, technology, person, need, activity, and so on. The latter definition emphasizes *acquiring and assessing information* rather than *assessing worth or merit* because all evaluation work involves collecting and sifting through data, making judgements about the validity of the information and of inferences we derive from it, whether or not an assessment of worth or merit results.

The Goals of Evaluation

The generic goal of most evaluations is to provide "useful feedback" to a variety of audiences including sponsors, donors, client-groups, administrators, staff, and other relevant constituencies. Most often, feedback is perceived as "useful" if it aids in decision-making. But the relationship between an evaluation and its impact is not a simple one -- studies that seem critical sometimes fail to influence short-term decisions, and studies that initially seem to have no influence can have a delayed impact when more congenial conditions arise. Despite this, there is broad consensus that the major goal of evaluation should be to influence decision-making or policy formulation through the provision of empirically-driven feedback.

Evaluation Strategies

'Evaluation strategies' means broad, overarching perspectives on evaluation. They encompass the most general groups or "camps" of evaluators; although, at its best, evaluation work borrows eclectically from the perspectives of all these camps. Four major groups of evaluation strategies are discussed here.

Scientific-experimental models are probably the most historically dominant evaluation strategies. Taking their values and methods from the sciences -- especially the social sciences -- they prioritize on the desirability of impartiality, accuracy, objectivity and the validity of the information generated. Included under scientific-experimental models would be: the tradition of experimental and quasi-experimental designs; objectives-based research that comes from education; econometrically-oriented perspectives including cost-effectiveness and cost-benefit analysis; and the recent articulation of theory-driven evaluation.

The second class of strategies are *management-oriented systems models*. Two of the most common of these are **PERT**, the **P**rogram **E**valuation and **R**eview **T**echnique, and **CPM**, the **C**ritical **P**ath **M**ethod. Both have been widely used in business and government in this country. It would also be legitimate to include the Logical Framework or "Logframe" model developed at U.S. Agency for International Development and general systems theory and operations research approaches in this category. Two management-oriented systems models were originated by evaluators: the **UTOS** model where **U** stands for Units, **T** for Treatments, **O** for Observing Observations and **S** for Settings; and the **CIPP** model where the **C** stands for Context, the **I** for Input, the first **P** for Process and the second **P** for Product. These management-oriented systems models emphasize comprehensiveness in evaluation, placing evaluation within a larger framework of organizational activities.

The third class of strategies are the *qualitative/anthropological models*. They emphasize the importance of observation, the need to retain the phenomenological quality of the evaluation context, and the value of subjective human interpretation in the evaluation process. Included in this category are the approaches known in evaluation as naturalistic or 'Fourth Generation' evaluation; the various qualitative schools; critical theory and art criticism approaches; and, the 'grounded theory' approach of Glaser and Strauss among others.

Finally, a fourth class of strategies is termed *participant-oriented models*. As the term suggests, they emphasize the central importance of the evaluation participants, especially clients and users of the program or technology. Client-centered and stakeholder approaches are examples of participant-oriented models, as are consumer-oriented evaluation systems.

With all of these strategies to choose from, how to decide? Debates that rage within the evaluation profession -- and they do rage -- are generally battles between these different strategists, with each claiming the superiority of their position. In reality, most good evaluators are familiar with all four categories and borrow from each as the need arises. There is no inherent incompatibility between these broad strategies -- each of them brings something valuable to the evaluation table. In fact, in recent years attention has increasingly turned to how one might integrate results from evaluations that use different strategies, carried out from different perspectives, and using different methods. Clearly, there are no simple answers here. The problems are complex and the methodologies needed will and should be varied.

Types of Evaluation

There are many different types of evaluations depending on the object being evaluated and the purpose of the evaluation. Perhaps the most important basic distinction in evaluation types is that between *formative* and *summative* evaluation. Formative evaluations strengthen or improve the object being evaluated -- they help form it by examining the delivery of the program or technology, the quality of its implementation, and the assessment of the organizational context, personnel, procedures, inputs, and so on. Summative evaluations, in contrast, examine the effects or outcomes of some object -- they summarize it by describing what happens subsequent to delivery of the program or technology; assessing whether the object can be said to have caused the outcome; determining the overall impact of the causal factor beyond only the immediate target outcomes; and, estimating the relative costs associated with the object.

Formative evaluation includes several evaluation types:

- *needs assessment* determines who needs the program, how great the need is, and what might work to meet the need
- *evaluability assessment* determines whether an evaluation is feasible and how stakeholders can help shape its usefulness
- *structured conceptualization* helps stakeholders define the program or technology, the target population, and the possible outcomes
- *implementation evaluation* monitors the fidelity of the program or technology delivery
- *process evaluation* investigates the process of delivering the program or technology, including alternative delivery procedures

Summative evaluation can also be subdivided:

- *outcome evaluations* investigate whether the program or technology caused demonstrable effects on specifically defined target outcomes
- *impact evaluation* is broader and assesses the overall or net effects -- intended or unintended -- of the program or technology as a whole
- *cost-effectiveness and cost-benefit analysis* address questions of efficiency by standardizing outcomes in terms of their dollar costs and values
- *secondary analysis* reexamines existing data to address new questions or use methods not previously employed
- *meta-analysis* integrates the outcome estimates from multiple studies to arrive at an overall or summary judgement on an evaluation question

Evaluation Questions and Methods

Evaluators ask many different kinds of questions and use a variety of methods to address them. These are considered within the framework of formative and summative evaluation as presented above.

In formative research the major questions and methodologies are:

What is the definition and scope of the problem or issue, or what's the question?

Formulating and conceptualizing methods might be used including brainstorming, focus groups, nominal group techniques, Delphi methods, brainwriting, stakeholder analysis, synectics, lateral thinking, input-output analysis, and concept mapping.

Where is the problem and how big or serious is it?

The most common method used here is "needs assessment" which can include: analysis of existing data sources, and the use of sample surveys, interviews of constituent populations, qualitative research, expert testimony, and focus groups.

How should the program or technology be delivered to address the problem?

Some of the methods already listed apply here, as do detailing methodologies like simulation techniques, or multivariate methods like multiattribute utility theory or exploratory causal modeling; decision-making methods; and project planning and implementation methods like flow charting, PERT/CPM, and project scheduling.

How well is the program or technology delivered?

Qualitative and quantitative monitoring techniques, the use of management information systems, and implementation assessment would be appropriate methodologies here.

The questions and methods addressed under summative evaluation include:

What type of evaluation is feasible?

Evaluability assessment can be used here, as well as standard approaches for selecting an appropriate evaluation design.

What was the effectiveness of the program or technology?

One would choose from observational and correlational methods for demonstrating whether desired effects occurred, and quasi-experimental and experimental designs for determining whether observed effects can reasonably be attributed to the intervention and not to other sources.

What is the net impact of the program?

Econometric methods for assessing cost effectiveness and cost/benefits would apply here, along with qualitative methods that enable us to summarize the full range of intended and unintended impacts.