

formative evaluation is done; and the product and cost-benefit analysis stages equate to a summative evaluation stage.

Use of the discrepancy model begins with a meeting to work on an evaluator's program description. All levels of program staff are invited and the large group is divided into smaller, workable groups. These groups respond to questions developed by the evaluator to elicit their ideas on how their program is designed. The resulting description of the design is then compared to design standards (devised by the sponsor or drawn from standards for the field or from some other source). The discrepancies that are observed, if any, between the standards and the developed design are communicated back to the staff for review and action. Now the evaluator can use the assessed design in the installation (or implementation) stage as the standard with which to compare the program's operation. The evaluator looks—with standards in mind—at staff and at clients and how they move through the program. A discrepancy evaluator's role is to determine the differences between what is and what should be. Again, this information is communicated back to the staff for any midcourse corrections.

In the process stage, there is a comparison between what is being accomplished (by clients, staff, and others) and the interim products that were anticipated. Here the evaluator communicates the degree to which these interim products have or will be achieved. In the product stage, the evaluator compares the degree to which the end products (for example, student learning, behavior change, and increased productivity) are in line with what were identified as anticipated end products in the original design.

In the final stage—cost evaluation—the evaluator compares the cost of similar programs having the same or a similar end product. Using the conclusions from this stage (and perhaps from the product stage as well), sponsors can make a policy decision to continue or end the program. Usually, this final stage is referred to as return on investment or cost-benefit analysis.

The discrepancy model is useful to a program staff that is interested in and able to have an evaluator working with them from the

outset of program operation. The strength of this model is in having the staff involved in determining and using the evaluation criteria and standards.

Goal-Free Model

In the goal-free evaluation model developed by Michael Scriven (Popham, 1974), the evaluation looks at a program's actual effect on identified needs. In other words, program goals are not the criteria on which the evaluation is based. Instead, the evaluation examines how and what the program is doing to address needs in the client population. With this model, you observe without a checklist, but record all data accurately and determine their importance and quality. Categories naturally emerge from your observations. This model of evaluation can use all forms of obtrusive methods (subject is aware of them—for example, tests) as well as unobtrusive ones (subject is not aware of them—for example, a hidden camera) to gather data. The evaluator has no preconceived notions regarding the outcome of the program (that is, goals). The staff should not contaminate the evaluator's method with goal statements. The evaluator is trying to form a description of the program, identify processes accurately, and determine their importance to the program. As the evaluator, you are gathering data on things that are actually happening and evaluating their importance in meeting the needs of the client population.

A good example of this model is the process followed by the Consumer's Union (producers of *Consumer's Report*) in which the manufacturer's intent for the product is irrelevant to the actual usefulness to the consumer.

The goal-free model is the most difficult to use, especially when the evaluator is part of the program or project; yet it is a popular method because it can be used within a program that has many different projects occurring simultaneously. In such a situation the same client population participates in a number of activities, and it is difficult to separate the results of two projects' activities. In fact,

program results might come from the interactions between two or more projects' activities.

For example, an evaluator might be asked to evaluate the effectiveness of an adult basic education (ABE) project housed within the program of a local adult learning center (ALC). Also housed in that program are workplace literacy, welfare to work, and adult computer literacy projects. Clients of the adult learning center may partake in any or all of these programs. Thus it would be difficult if not impossible to isolate the results of just one project's activities. A goal-free evaluation would examine the overall results for the clients of the ALC program, which would be more meaningful than individual evaluations of each project.

The person who performs the goal-free evaluation of the ABE project may have no subject-matter expertise in the field of adult education. This is a topic of debate among many experts. Some say the evaluator should have expertise in the field being evaluated; others say no expertise is better (Rossi and Freeman, 1993). The issue, of course, is preconceived notions. Some scholars say that an evaluator who is not familiar with the nuances, ideologies, and standards of a particular professional area will presumably not be biased when observing and collecting data on the activities of a program in that area. They maintain, for example, that a person who is evaluating a program to train dental assistants should not be a person trained in the dental profession. But other scholars allege that a person not aware of the nuances, ideologies, and standards of the dental profession may miss a good deal of what is important to the evaluation. Both sides agree that the evaluator must attempt to be an unbiased observer and be adept at observation and capable of using multiple data collection methods (Wholey, Hatry, and Newcomer, 1994).

Once the data have been collected, the evaluator attempts to draw some conclusions about the impact of the program on addressing client needs. This information is then delivered to parties interested in the evaluation results. Again, the evaluator using this

model makes a deliberate attempt not to know about program goals, written proposals, or brochures that exist. He or she simply studies the outcomes and reports on them.

The goal-free model works best for qualitative evaluation because the evaluator is looking at actual effects rather than anticipated effects for which quantitative tools have been designed.

Interestingly, Scriven suggests using two goal-free evaluators, each working independently. In this way, the evaluation does not rely solely on the observations and interpretations of one person.

Transaction Model

The transaction model developed by R. E. Stake in 1975 (Madaus, Scriven, and Stufflebeam, 1983) affords a concentration of activity between you, as both evaluator and participant, and the project staff. The main beneficiaries of an evaluation using this model are the clients and practitioners.

This model combines monitoring with process evaluation through a constant back-and-forth between evaluator and staff. The evaluator is an active participant, giving constant feedback. In effect, the evaluator is or acts as one of the project staff members.

The evaluator uses a variety of observational and interview techniques to obtain information from the program staff and clients. This model may have a goal-free or a goal-based orientation. Instead of trying to achieve objectivity as in the previous models, the evaluator uses subjectivity in the transaction model.

Using the previous example of the adult learning center, the transaction evaluator might be one of the teachers of the ABE project who is assigned to follow a group of clients through the other projects in an attempt to distinguish any measurable results coming from a single project. The evaluator is one of the staff of the ALC, participating in and providing project activities. The findings are shared with the staff of all the projects to improve both individual projects and the overall program.

Decision-Making Model

The decision-making model developed by Daniel Stufflebeam (Madaus, Scriven, and Stufflebeam, 1983) is employed to make decisions regarding the future use of the program. In this case, you are not as concerned with how the program is going presently. Instead you are concerned with its long-range effects, such as the number of cancer patients who survive in a five-year trial, or the number who survive from this program as compared to another program with a different approach. The focus is on decisions that need to be made in the future.

For example, an adult education program might have three different commercial packages for teaching people with a low literacy level to read English. In previous evaluations all three packages have proven effective in teaching reading; however, the sponsors of the program need to cut funds and a decision needs to be made to discontinue the use of one or more of the packages. This is a decision-based situation that requires focus not on the client, the staff, or the activity but on how best to cut operating expenses.

This model is wide open in the methodology you use to collect data. Both quantitative methods (such as tests and records) and qualitative methods (such as interviews, observations, and surveys) might be employed. This choice depends on what the sponsor wants to know in order to make the decision; it is a totally summative evaluation.

Goal-Based Model

The goal-based model, also called the objective attainment model, is the easiest to use and therefore the most often used. The evaluation may be based on stated objectives or goals found, for example, in a proposal, brochure, or other description of the program. This model is not concerned with ancillary items, variables, or occurrences that might be spin-off products of the program activities, just stated objectives. The wording of the objectives would usually iden-

tify the tests and standards for the evaluation. The evaluator is looking to measure specified outcome variables, using quantitative or qualitative methods. This model can become the most research-like, especially if you can convince the stakeholders to use a control or comparison group, as discussed in Chapter Eight.

Other Models

Other models to consider are the systems analysis model developed by Rivlin, the art criticism model developed by Eisner, and the adversary model developed by Owens (Madaus, Scriven, and Stufflebeam, 1983).

In the systems analysis model, the evaluator looks at the program in a systematic manner, studying the input, throughput, and output. Input is elements that come into the system—for example, clients, staff, facility, resources—as they are prior to encountering the program. Throughput consists of things that occur as the program operates—for example, activities, client performance, staff performance, resource availability, and the adequacy of resources such as money, people, and space. Output is the results of the program—for example, client change, staff effectiveness, adequacy of activities. The evaluator examines the program's efficiency in light of these categories. This model might be employed to determine whether a program is getting people into the program and out of the program in an efficient manner, as well as achieving its goals.

In the art criticism model, the evaluator is a qualified expert in the nuances of the program and becomes an expert judge of the program's operation. The effectiveness of the model relies heavily on the evaluator's ability to judge objectively. This model might be employed when a program wishes to have a critical review of its operation prior to applying for funding or accreditation of some sort. In the private sector, for example, this model could be used to prepare for accreditation by the International Organization for Standardization (ISO). In educational circles, likewise, the accreditation process could prompt the use of an expert judgment.

In the adversary model, the evaluator facilitates a jury that hears evidence from individuals (sometimes adversaries) on particular program aspects. The jury then uses multiple criteria to “judge” evidence and decide what is actually happening. This model can be used when there are different views of what is actually happening in a program. These might be differences among clients, staff, community members, or sponsors.

Choosing the Right Model

The logical question at this point is, “How do I know which model suits my particular situation and needs?” This question is answered by looking back at your answer to the original question, “Why evaluate this program?” If it was to meet some mandate of a funding source or management, you might want to employ the goal-based model. If it is to learn something about your program so that staff can improve service delivery, you might employ the systems analysis or goal-free models. If your intent is to critically examine certain aspects of your program for reduction or promotion, you might use the decision-making, art criticism, or transaction models. The adversary model would be used when the purpose of the evaluation is to settle differences of opinion between stakeholders.

Table 5.1 will help you to decide which model to use.

Evaluation Design Format

Now that you have chosen a model to use, you can return to the larger picture that we introduced at the end of Chapter One, the overall evaluation design format. This format introduces the components that may occur in any evaluation: evaluation questions, program objectives, activities observed, data sources, population samples, data collection design, responsibility, data analysis, and audience. Not all components appear in each of the models—for example, the program objectives component is not used in the goal-free model—but most are common to all.

TABLE 5.1 Choosing a Model.

<i>Model</i>	<i>Intended Outcome</i>	<i>Evaluator's Tasks</i>	<i>Sample Evaluation Questions</i>
Adversary	Resolution of differences of opinion	Facilitation	What are arguments for and against program components?
Art Criticism	Critical reflection, improved standards	Expert judgment	Would a professional approve of program activities?
Decision-Making	Effectiveness, impact, quality	Data collection, analysis, interpretation	Was the program effective? What aspects of the program were effective?
Discrepancy	Compliance with standards	Facilitation, monitoring, data collection, analysis, interpretation	How did the program perform compared to standards?
Goal-Based	Efficiency, effectiveness, impact	Data collection, analysis, interpretation	Did the clients change (grow, learn)?
Goal-Free	Usefulness, impact	Data collection, analysis, interpretation	What happened in the program?

TABLE 5.1 Choosing a Model, *continued*.

Model	Intended Outcome	Evaluator's Tasks	Sample Evaluation Questions
Systems Analysis	Efficiency, effectiveness	Monitoring, data collection, analysis, interpretation	Were the expected outcomes achieved? Were the expected effects achieved efficiently?
Transactional	Program understanding	Participation, data collection, analysis, interpretation	What does the program look like from different vantage points?

Evaluation questions are central to all the models and all evaluations. As described in Chapter Four, these are questions that the evaluator or program staff or both develop to ensure that the evaluation results will address meaningful questions and lead to program improvement or promotion. These are much larger questions than those presented by program objectives.

Program objectives may be added at this point. These are the statements of intent that the program developers created to communicate what would be accomplished if the plan were implemented. These statements are the evaluator's friend because, if stated correctly, they will contain what is to occur (activity), to whom (client), the expected outcome (criterion), and how you will know (measurement). With this information, the evaluator knows what the staff will perform, what results they expect, and how they will measure program performance.

Activities are those specific activities that program staff will con-

duct for clients. It is important for the evaluator to know the activities and to which objectives they relate in order to identify a cause-and-effect relationship.

Data source is the instrument for data collection and recording. Depending on whether you will be employing qualitative or quantitative methods, your data source may be surveys, interviews, observation protocols, tests, or calibrated measuring devices. Your data source may collect new data from clients or contain already existing data on clients.

Sample refers to the individuals from whom you will collect data, specifically to the proportion of individuals in a program who are anticipated to participate in the evaluation. The term *population* refers to all the individuals who might be eligible to participate, and the term *sample* refers to those whom you will target to participate.

Data collection design is the schedule on which you will collect data. Depending on the model you select and the purpose of your evaluation, you may need to collect data before the clients interact with your program (pre) and after they have partaken of your activities (post). On occasion, you might collect data while they are partaking (interim) so that you can monitor change. The data collection design also communicates whether only those partaking in the activity will be included in the data collection or whether you will select a group of individuals who will not partake in the activity (control or comparison group).

Responsibility refers to listing, as clearly as possible, what evaluation activities you will perform as the evaluator. Those evaluation activities that you will not perform need to be identified, along with who from the program staff will be responsible.

Data analysis explains what statistical manipulations (if any) will be performed on the data collected. Each data set from each data source will probably have its own data analysis procedure.

Audience is the group for whom the report is intended. In some cases, the audience may be different for different evaluation

questions. Sponsors, for example, may be more interested in evaluation questions directed at program efficiency and impact, whereas program staff would be more interested in evaluation questions on program effectiveness.

For example, if your task is to help sponsors or administrators decide the future of the program, you may elect to use the decision-making model. This makes the focus of the evaluation summative; you're not interested in how it's going at the moment but rather in the end results or effects. You need to report the program's effects to those who authorized or pay for both the program and the evaluation. So you can certainly say that you know the audience who will be interested in the evaluation you are doing.

Other Considerations

The eight models discussed in this chapter are by no means the only evaluation models available to you. Others offer different structures, foci, methods, and so forth, and hybrid models contain aspects of two or more of the models.

In Chapter One, we presented two definitions of evaluation that are commonly accepted in the field:

- Evaluation is the systematic process of collecting and analyzing data in order to determine whether and to what degree objectives have been or are being achieved.
- Evaluation is the systematic process of collecting and analyzing data in order to make a decision.

Just by examining these two definitions, you can begin to identify which evaluation models might best suit an evaluator who is operating within one of these philosophical frameworks. Yet circumstances might arise in a particular evaluation that would require you to use whatever works for both the program and you.

Putting It All Together

In this chapter, we have presented the “meat” of evaluations an evaluation methodology. The models discussed provide a variety of ways to look at a program that is either clearly designed or not clearly designed. The range of models allows an evaluator to be rather dogmatic in assessing a program against a preconceived set of standards and objectives, or fairly free to evaluate a program's worth based on what it produces—not what it said it would produce.

Organizations and the programs and projects associated with them have an abundance of people besides the evaluator who are assets to an evaluation. Because they have first-hand knowledge of the program or project, their contribution to the selection of model may be extremely valuable (Gray, 1998). The evaluator should choose the model only if those being evaluated desire that to happen. Often those involved with the program see evaluation as one of those highly technical or complicated functions in which they “should” have no say, but that view is incorrect. All stakeholders should be involved in the decisions surrounding an evaluation, with the evaluator acting as the facilitator of the evaluation.

Using the chosen model, you create the evaluation format which outlines the components of the evaluation matrix. The parts tend to flow in a systematic order, but they all stem from the evaluation questions. These questions represent the real interests of the stakeholders—what they want to know about the program in order to improve it.

Questions and Exercise

Now that you have read Chapter Five, return to the questions that you were asked to keep in mind at the beginning of the chapter.

1. What are some of the more popular evaluation models?
2. What are the component parts of an evaluation format?