The real challenge for evaluation research and policy analysis is to develop methods of assessment that emphasize learning and adaptation rather than expressing summary judgments of pass or fail.

—Giandomenico Majone (1988)

This chapter discusses the purposes, acknowledged and unacknowledged, for which people decide to undertake program evaluation. I suggest that the evaluator find out what the program community really seeks from the study and how they expect to use the results. With this knowledge, she1 can most effectively tailor the evaluation to provide the kinds of information that people are interested in. The chapter also discusses the procedures used to initiate and contract for an evaluation. The location of the evaluation unit—where it fits in the organizational structure—can make a difference in whether the study has sufficient latitude to be useful.

Who Wants Evaluation?

Many people in many locations have an interest in evaluating programs and policies. An organization that funds programs, such as a philanthropic foundation or a state legislature, wants to know what the local operating agencies are doing with the money they receive and how well they are serving their clients. In the field of international development, for example, it is common for donor agencies to seek evaluations of the work done with the funds they provide to developing nations.

An organization that oversees programs, such as the federal Head Start office or the national Girl Scouts of the USA, wants to find out what is going on in the field and how clients are faring in the local units. Situated far from the action, they need a mechanism that will provide unbiased information about what is happening on the ground.

Managers of a local project want evidence on the short- and long-term effects of the activities they run. Although they are close enough to see what is going on, they usually do not know the consequences of the project and whether the clients they serve benefit in expected ways. They also do not generally know whether variations in activities produce differential outcomes. They want to learn, for example, whether some modalities of service are more effective than others, whether staff with certain qualifications have better results than others, whether increases in intensity or length of service would improve outcomes, or whether they could achieve the same results for less money.

In some agencies evaluators themselves have stimulated evaluation. When an evaluation unit exists in the agency, the evaluator can press for periodic study of issues that arise during the conduct of projects. This happens at federal, state, and local levels of government and in direct-service organizations. Evaluators who are not on the staff of the agency (i.e., who work in outside research, evaluation, and consulting firms, institutes, or universities) do not generally have an avenue to promote the use of their evaluation services—at least not directly. But they can use many indirect means to make the conduct of evaluations look like an attractive and responsible thing to do.

For whatever reason an evaluation is originated, note that the people who launch it are likely to be situated at the top of the hierarchy. Funders, directors, managers—they come to the evaluator with a particular set of questions and concerns. An evaluator’s initial impulse is to accept the concerns that they articulate as the key questions for the evaluation. After all, they initiated the study; they are paying the bills; they have authority to use evaluation results to make changes in the program.

So far, so good. It is important to know what is on the minds of people in authority. But evaluators should recognize from the start that funders and managers do not represent the whole range of questions that are floating around about the program. Nor, however high their rank, are they the only ones who will make decisions about the program. Many other people will have a say about what happens, not least the staff whose day-to-day actions determine what the program actually is and the clients whose decisions to attend or not attend and how seriously to pay attention set limits to what the program can do. Evaluators are often wise not to restrict their attention to questions of those in the upper echelons but, where possible, to consider the questions of other people affected by the program.

Overt and Covert Purposes

In olden times evaluators believed that a study that responded to the questions of the study sponsor would almost automatically be used. They assumed that people funded a study because they needed information in order to make better decisions. In the past decades we have learned that this is not always true. Of course, many people do want to know what evaluation can tell them, and many more can be per-
sueded that evaluation can tell them things that they want to know. But it is equally true that people decide to have a program evaluated for many different reasons, from the eminently rational to the patently political. Before taking the mission at face value, the evaluator should consider the possibility that people are turning to evaluation for less legitimate reasons.

**Evaluation as Subterfuge**

**Postponement** People in positions of authority may be looking for ways to delay a decision. Instead of resorting to the usual ploy of appointing a committee and waiting for its report, they can commission an evaluation study, which takes even longer.

**Ducking Responsibility** Sometimes one faction in the program organization is espousing one course of action and another faction is opposing it. The administrators look to evaluation to get them off the hook by producing dispassionate evidence that will make the decision for them.

**Window Dressing** There are cases in which administrators know what the decision will be even before they call in the evaluators, but they want to cloak it in the trappings of research. They use evaluation to provide legitimacy.

**Public Relations** Occasionally, evaluation is seen as a way of self-gloration. The manager believes that he has a highly successful program and looks for a way to make it visible. A good study will fill the bill. Copies of the report, favorable of course, can be sent to boards of trustees, members of legislative committees, executives of philanthropic foundations who give large sums to successful programs, and other influential people. Suchman (1967) suggests two related purposes: eyewash and whitewash. In an eyewash evaluation, an attempt is made to justify a weak program by selecting for evaluation only those aspects that look good on the surface. A whitewash attempts to cover up program failure by avoiding any objective appraisal.

The program administrator's motives are not, of course, necessarily crooked or selfish. Often, there is a need to justify the program to the people who pay the bills, and he is seeking support for a concept and a project in which he believes. Generating support for existing programs used to be a fairly common motive for embarking on evaluation, until managers found out that evaluations tend to find flaws in even well-regarded programs. In those cases where the general tenor of evaluation is positive, the study also surfaces the "warts and all." Now only the naive manager turns to evaluation when he seeks to shore up the agency's public relations, or the manager who so stringently limits the questions that are addressed that he can be sure of the findings.

**Fulfilling Grant Requirements**

Often the decision to evaluate stems from sources outside the program. Many federal agencies and foundations mandate evaluation for each demonstration project and innovative program they fund. An evaluation is conducted merely to fulfill the requirement. Some ongoing projects, long past the demonstration stage, are also required to conduct evaluation to satisfy conditions of the grant. For many years this was true of projects to improve the school performance of disadvantaged pupils funded under Title I of the Elementary and Secondary Education Act.

For policymakers who support large-scale multi-site programs, it is obvious that wise policy and proper management are served by a supply of systematic evidence about the program. For untried innovative projects, it makes sense for funders to require evidence on the extent to which the project is working. But to the operators of a project, the demands of starting up and running the project take priority. Plagued as they often are by immediate problems of staffing, budgets, logistics, community relations, and all the other trials of project life, they tend to look on evaluation as just another burden. They see it mainly as a ritual designed to placate the funding bodies, without any real usefulness to them.

For ongoing projects, routine often sets in. Administrators and staff who have been receiving federal funds for 20 years to run pretty much the same project are no longer experimenting with different techniques or materials. They are operating the project that they have learned how to operate. The demand for evaluation looks to them like another one of those pointless hoops they have to jump through to keep getting the money.

Evaluation, then, is a rational enterprise often undertaken for nonrational, or at least noninformational, reasons. We could continue the catalog of covert purposes (justifying a program to Congress, demonstrating that the administrator is acting in accordance with highly prized procedures of management), but the important point is that such motives have consequences for the evaluation that can be serious and bleak.

An evaluator who is asked to study a particular program usually assumes that she is there because people want answers about what the program is doing, how it is doing it, and what the consequences are. When this is not the case, she may in her naivete become a pawn in an intraorganizational power struggle, a means of delaying action, or the rallying point for one ideology or another. Some evaluators have found only after their study was done that they had unwittingly played a role in a larger political game. They found that nobody was particularly interested in learning from their results, but only in using them (or any quotable piece of them) as ammunition to destroy or to justify.

Lesson number one for the evaluator newly arrived on the scene is: Find out who initiated the idea of having an evaluation of the project/program and why. Were there other groups in the organization who questioned or objected to the evaluation? What were their motives? Is there real commitment among practitioners, administrators, and/or funders to paying attention to the results of the evaluation? If the real purposes for the evaluation are not oriented to increased understanding and there is little commitment to considering results, the chances for influence are likely to be small.

The saving grace is that evaluations often prove to be useful even when they were begun with no intention to put the results to use. When the report comes in, funders and managers may find that it gives them direction for improving the
way the program is run. They may find that the results give them grounds for doing things they were considering doing anyway. Further, so long as the evaluation report is made public (and all reports to federal agencies have to be made public), evaluation findings reach not only their intended audiences but also percolate into the awareness of other members of the program community. Through a variety of direct and indirect routes, people in many places hear about the findings. The potential for general enlightenment from evaluation helps to overcome some of the unfavorable signals emitted by uninterested study sponsors.

**Conditions Unfavorable for Evaluation**

Is evaluation always warranted? Should all programs if they are good little programs go out and get themselves evaluated? Heretical as it may seem coming from an evaluator, the answer is probably no. If evaluation results are not going to have any influence on the way people think about the program, it is probably an exercise in futility. Evaluation may not be worthwhile in four kinds of circumstances:

1. **When the program has few routines and little stability.** Program staff improve activities from day to day, based on little thought and less theory, and the program shifts and changes, wanders around and seeks direction. The program that an evaluator starts studying today will bear little resemblance to the program a year or two from now when the evaluation report is ready. At that point the report will be about a program that is ancient history! If evaluation found out what the effects were on participants, there would be no way of knowing what was the program that caused them.

   Even if the program initially adhered to a plan, staff may veer off in other directions, either as individuals or collectively. The initial plan is left behind without being replaced by a coherent alternative. Again it is not clear what the program is, and therefore it would not be clear what the evaluation means.

2. **When people involved in the program cannot agree on what the program is trying to achieve.** If there are vast discrepancies in perceived goals, staff are probably working at cross purposes. Again, the coherence of the program is in doubt. Furthermore, it is not clear what criteria the evaluator should use in evaluating program outcomes. A study of the inner workings (process) of the program might be helpful, but major disagreements about goals should signal a go-slow approach for outcome evaluation.

3. **When the sponsor of the evaluation or program manager sets stringent limits to what the evaluation can study, putting off limits many important issues.** The sponsor or manager wants eyewash or whitewash, and the evaluator is expected to give legitimacy to the enterprise.

4. **When there is not enough money or no staff sufficiently qualified to conduct the evaluation.** Evaluation is a demanding business, calling for time, money, imagination, tenacity, and skill. Without sufficient quantities of all of these, evaluation may produce information that is more misleading than informative.

**Purposes of Evaluation**

Of course, we can argue that even in such dismal circumstances, evaluation research can produce something of value, some glimmering of insight that will light a candle for the future. This is a fetching notion, and I sometimes succumb to it. But experience suggests that even good evaluation studies of well-defined programs often wind up as litter in the bureaucratic mill. When conditions are unfavorable, the prospective evaluator should be clear about what she is getting into.

**Evaluation for Decision Making**

Most program directors and project managers undertake an evaluation for reasons that are more auspicious than those we have just contemplated. They may want to know what is happening in the program so that they can take midcourse corrective action. For example, if the evaluation of program process reveals that staff is discouraging tough cases from persisting in the program, the managers may want to retrain or reorient staff to the importance of serving hard-to-serve people. Policymakers may want to know the consequences of the program for its clients so that they can decide whether or not to expand it to new sites or to new categories of participants. These are the kinds of evaluation-for-decision purposes that most people assume evaluation serves.

**Midcourse Corrections** An important reason for undertaking evaluation is to find out early in the game what is going on so that changes can be made. Sometimes evaluation is triggered by warning signs that all is not well. But in many other cases, a regular commitment to evaluation is based on the hope that ineffective practices can be identified early before they do lasting harm. Evaluations that focus on program process can yield data on what is happening during the program’s course and identify points at which improvement is called for. Perhaps eligibility criteria for clients have to be redefined, or rules have to be enforced more stringently when clients fail to follow directions, or new categories of staff, such as outreach workers, have to be added. Evaluation of program processes can send up early signals that attention is needed and identify the kinds of change that are warranted.

**Continuing, Expanding, or Institutionalizing the Program, or Cutting, Ending, or Abandoning It** Evaluation can also be dedicated to finding out the extent to which a program is achieving its goals. Information on this score can be used to decide whether the program should be continued and spread to further locations, or whether investments in it should be cut. Past experience shows that evaluation results are not the only basis on which decisions about the future of the program will be made—and not necessarily the major one. Because programs serve many functions, political and symbolic as well as instrumental, evidence of outcomes is only one of the considerations that are taken into account. Nevertheless, evaluation provides data on what the program accomplishes and
fails to accomplish, and for whom, and thus clarifies the trade-offs that decision makers have to make.

**Testing a New Program Idea**  When an agency comes up with a new programmatic approach to carry out its charge, it can try out the approach in a series of demonstration projects and evaluate effects. For example, a philanthropic foundation can conclude that providing services to poor people individually is not enough to lift them out of poverty. Instead, agencies should work at rebuilding the communities in which poor people live. The foundation then sets up demonstrations of the new approach. If the projects prove successful in lively and important ways, those who make decisions about future programming can adopt the community approach as a basic strategy. Or managers of other programs may incorporate some of the key practices into ongoing programs elsewhere. If the projects have poor or mediocre results, modifications can be made to improve their working. In the end, if evaluation shows that they do less well than activities already in practice, the idea may be abandoned.

**Choosing the Best of Several Alternatives**  Sometimes several versions of a program are run, such as two or more math curricula or several modalities of health care cost containment. The intent of evaluation is to find out which of the options has the best outcomes, with the expectation that the best one will be adopted. Experience suggests that no one alternative is likely to be dominant on all dimensions. Each version of a program is apt to do better in some ways, worse in others. Still, clarifying which each alternative gains and loses can provide significant learning. Those who make decisions then have to consider which dimensions are most salient to the varied publics concerned.

**Deciding Whether to Continue Funding**  The funder of a set of projects wants to know which of the agencies currently receiving program funds should continue receiving them. This is a legitimate purpose for evaluation, but it is the kind of purpose that tends to make evaluated projects nervous and hostile. They know that they are being judged and that their continued livelihood may depend on how well they do. To borrow a term from the testing field, these are high-stakes evaluations. Under such circumstances, projects may try to influence, subvert, or even sabotage the evaluation. Many evaluators are uneasy about conducting studies that make judgments about individual projects, but experienced evaluators learn how to avoid invidious comparisons and keep the focus on things that matter.

### Evaluation as Organizational Learning

The purposes listed so far have all dealt with instrumental action. Evaluation is expected to contribute to wiser choices and improved programming. There are other worthy purposes for evaluation that are less oriented to program decisions, and it is to these I now turn.

**Record of Program History**  Someone in the program or project seeks evaluation as a record of program history, perhaps from a belief that epochal things will take place that ought to be documented, perhaps with a vague sense that he or others can learn lessons from history. An evaluator under such circumstances will have some difficulty narrowing down the open-ended charge to record what happens and find an agreed-upon focus for the study, but there are opportunities here to contribute to program understanding and, perhaps, action.

**Feedback to Practitioners**  Rather than providing information to program managers, sometimes evaluation is expected to feed back information to practitioners as the program goes along. The evaluator in this case is expected to be a sensitive observer who keeps track of events and makes visible to staff what they are doing, how participants are responding, what external conditions are impinging on the program, and so on. When the evaluator and the staff share common assumptions about the program and a common language in which to communicate, this can be a highly useful role. Sometimes the evaluator is expected to go further and give staff advice. If this is the case, it is important that the evaluator be something of an expert about programs of this sort, so that practitioners have confidence that the advice is realistic and competent. Where there is a disjunction in viewpoint between evaluator and staff, extended negotiations may be necessary before the feedback is accepted and stands a chance of having practical effects.

A related purpose for evaluation is to give practitioners a tinge of skepticism about what they are doing. Most staff engaged in program service take for granted that their service is good and will have good effects. They believe in the value of their work. They almost have to believe, so that they can get up and go to work every day. As one program manager said to me long ago, the job of a practitioner is to believe; the job of an evaluator is to doubt. The whole raison d'être of evaluation is to ask whether things that the program assumed were good turn out to be good in fact. A sophisticated program manager may want to transmit to practitioners something of the evaluator's doubt. (He may also hope to give the evaluator a touch of the staff's belief.) It may be good for staff to step back a bit and reflect on the consequences of their work and ways in which they can improve their practice. Although this is rarely a direct purpose for doing evaluation, it is sometimes an outgrowth of the evaluation procedure. As such, a canny administrator may find the encouragement of reflective practice (Schon, 1983) to be a desirable byproduct.

**Highlighting Program Goals**  Another way in which program managers can use evaluation to influence practitioners' behavior is by using the occasion of evaluation to emphasize program goals. If the evaluator is encouraged to use program goals as the criteria against which outcomes are weighed, the study reinforces staff's concern with the goals that have official sanction. As one program manager told me, this turns the evaluation into a disciplining mechanism, a way to get staff to agree (again) on what they mean to accomplish and attend to objectives that might otherwise slide from the forefront of their minds. When they see the connections between
what they are trying to accomplish and their daily activities, they bring their work into line with program priorities.

**Accountability**

Given the current stringency of budgets, public agencies are urged, sometimes required, to give reports to the public detailing what they have accomplished with public funds. The press for accountability can lead agencies to undertake systematic evaluation. Although accountability reasons for evaluation may seem to be ritualistic, much like the “fulfilling requirements” purpose I sniffed at before, there may be more promise in this situation. When evaluation results are reported to policymakers, funders, and the public, they get a wider hearing. By virtue of being public, results may stimulate adaptive action. Knowing that other people will be looking may make practitioners more attentive to the public’s expectations.

**Understanding Social Intervention**

Few evaluations are initiated with the express purpose of acquiring basic knowledge about the kinds of things that work to change social conditions and human behavior. Still, evaluations provide a first-rate opportunity for studying the consequences of social interventions. The study of policies and programs in action presents a chance to study the stimuli that lead to change and the environmental conditions under which change is likely to occur. It provides a chance to develop theory about program initiatives. This kind of purpose for evaluation is sometimes termed as academic and remote from the practical everyday concerns that should be evaluators’ fare. But as the social psychologist Kurt Lewin said, “Nothing is so practical as a good theory.” If social scientists really understood what it took to alter people’s behavior, the nation would save an enormous amount of waste in trial-and-error programmatic efforts.

Because human beings are so various, and because conditions change at an accelerating rate, the development of theory about human action and reaction is a precarious enterprise. Still, evaluation is an opportune site for investigation. Because it goes on in a setting where somebody is trying to change clients’ knowledge, attitudes, or behavior, it provides an acid test for old theories of intervention and an opportunity to help construct new ones. The theories it contributes to are apt to be partial, time-bound, and generalizable to only a small subset of interventions, but even modest advances can help improve the efficacy of policy and program design.

The question for the evaluator to keep in mind is: Do the conditions that led to evaluation hold promise that people will give considered attention to the findings, either now or in the future? If there’s a chance of it, and the resources are adequate, let’s do it.

**Intended Uses for Evaluation**

Even when evaluation is undertaken for good reasons, people can have widely differing expectations of the kinds of answers that will be produced. If the evaluator is not to be caught unawares, it behooves her to know from the outset what questions her study is expected to answer.

---

**Who Expects What?**

Expectations for the evaluation generally vary with a person’s position in the system. Top policymakers tend to ask for the kind of information that will help them address broad issues: Should the program be continued or dropped, institutionalized throughout the system or limited to a pilot program, continued with the same strategies of operation or modified? Should more money be allocated to this program or to others? They therefore have an interest in the overall effectiveness of the program. That is the kind of knowledge that enriches their thinking and enables them to proceed with greater conviction. Foundation executives, for example, have the opportunity to move their resources from one field to another with expediency. Legislators can initiate new programs and policies and reform or terminate old ones. Although experience shows that evaluation findings alone rarely lead to the abandonment of unsuccessful programs or the expansion of successful ones (many other things have to fall into place), evaluation sponsors usually expect data on the program’s overall success.

Project managers tend to be less interested in global assessments of the program’s effectiveness. They are not generally engaged by issues of extension or termination or provision of more or fewer resources. They are not in charge of such decisions. They are not authorized to decide whether to move from treatment to prevention, from centralized to decentralized operation, or lower to higher expenditures. They are responsible for the current project. They have to work within the parameters that have been defined. Their concern is how to make the project in its present incarnation work as well as possible.

Overall assessments of program effectiveness are directly useful primarily when the data show substantial success. Such information helps project directors mobilize support for the program and ward off criticism. When overall outcome data show mixed effectiveness or a list toward the negative, the information is useful to them to the extent that it helps them figure out which procedures and techniques are more or less successful, which are achieving results most efficiently and economically, which features of the program are essential and which can be changed or dropped. In other words, they seek the kind of direction from evaluation that helps them to improve the program.

The distinction I draw here between policy-level concerns and program-level concerns is worth making. Policymakers care about overall effects; program managers are more interested in the differential effects of alternative strategies. But the lines between policy and program issues become blurred in operation. Policymakers often get involved in program details. For example, the Congress and many state legislatures have taken on responsibility for specific program and policy provisions. They have become micromanagers. Instead of writing broad prescriptions in 10-page bills, they sometimes write bills that cover hundreds of pages, prescribing minute details of a program’s functioning, from the membership on its local advisory committees to the square footage of space it shall occupy. Information on details used to be of interest mainly to staff in executive departments who wrote guidelines and regulations for the program and managers who actually ran the projects. Now such information often finds a ready audience in the halls of Congress.
Direct-service staff deal with individuals and small groups. They have practical day-to-day concerns about techniques. Should they spend more of the time in job training courses trying to develop good work habits and less time on teaching subject matter? Should they put more emphasis on group discussions and films or lectures? Should they accept more younger people (who are not already set in their ways) or more older people (who have greater responsibilities and more need)? Practitioners, who are accustomed to relying on their own experience and intuitive judgment, often challenge evaluation to come up with something practical on topics such as these.

Nor do these three sets of actors—policymakers, program directors, and practitioners—exhaust the list of those with a stake in the evaluation. The funders of evaluation research, particularly when they are outside the direct line of operations, may have an interest in adding to the pool of knowledge in the field. They may want answers less to operating questions than to questions of theory and method. Should efforts to help poor families be concentrated on education or job creation? Does increasing the available career opportunities for low-income youth result in lower rates of juvenile delinquency? If coordination among community health services is increased, will people receive better health care? Here is another purpose for evaluation—to test propositions about the utility of models or theories of service. The public, too, has a stake, as taxpayers, as parents of schoolchildren, as contributors to voluntary organizations. They are concerned that their money is wisely and efficiently spent.

Consumers of services have a particular stake in the program, and therefore in its evaluation. At the outset they want some gauge of a project’s effectiveness to help them select a particular school, alcoholism treatment center, or vocational rehabilitation project. Later on, if they are not receiving the kind of help they need and want, they would like the evaluation to reveal the extent of shortcomings and point the way to more beneficial services. Consumers may seek to use evaluation to ask client-eye questions about the program under study. Is the program serving the objectives that clients value? In some communities, opposition has arisen to traditional formulations of program criteria. Activists not only want to improve school achievement or health care, they are concerned with community participation or community control of programs and institutions as a mechanism to assure that clients’ values get priority attention. When such issues are paramount, evaluative questions derive from a radically different perspective.

A host of interested onlookers are the managers and staff of other similar projects who want to learn as much as they can about what works. The managers of other probation services or nutrition projects want to find out how the project under study was run and with what effects. They expect to learn from the evaluation how they can improve their own project, or if they have not yet started, whether its results are so good that they should undertake something similar.

Finally, there are the designers of policies and programs who look to evaluation studies for guidance about the directions that they should take, or avoid, as they go about designing new policies. There are also social scientists who want to draw conclusions about the whole field of social programming and about the ways that human beings respond to interventions in their social world.

**Formative and Summative Evaluation**

In discussing potential uses and prospective users, a useful set of terms was introduced by Scriven (1967). He was talking about the evaluation of educational curriculums, and he distinguished between formative and summative evaluation. Formative evaluation produces information that is fed back during the development of a curriculum to help improve it. It serves the needs of developers. Summative evaluation is done after the curriculum is finished. It provides information about the effectiveness of the curriculum to school decision makers who are considering adopting it in their schools.

Scriven (1991) later simplified the distinction. In a phrase he attributes to Robert Stake, he offers this definition:

> When the cook tastes the soup, that's formative evaluation; when the guest tastes it, that's summative evaluation. (p. 19)

This distinction is not limited to curriculum development but can be applied to other types of programs as well, with obvious advantages for the clarification of purpose. Formative evaluations are designed to assist those who are developing projects in the early phases. The emphasis is on feedback to developers with an eye to improving the final product. Summative evaluation is meant for decisions about whether to continue or end a program, extend it to other locations or cut it back.

However, many programs are never “finished” in the sense that a curriculum is finished when it is published and distributed. Programs continue to adapt and transmute in response to conditions inside and outside the program agency. For reasons good and not so good, program practitioners and managers modify what they do over time. Change may come about because they learn better strategies through practice, or because of demands from clients, loss of a charismatic director, hiring of staff with different skills, changing community sentiments, rising staff morale, or any of a dozen other shifts. The need for “formative” information continues—that is, information fed back to program staff on how to improve the program.\(^3\)

When other sites seek to adopt the program, they too need formative information. They never replicate the program exactly but adapt it to suit local conditions (Berman & McLaughlin, 1977, 1978). The perennial pattern of program modification is a reason it has proved so difficult to identify the “ideal program” and then work for its institutionalization systemwide. Programs change in operation, and today’s exemplary program is tomorrow’s mediocrity. Outside conditions differ, too. The program model that starred in “What Works” in Savannah becomes a dud in Detroit. Hopes that evaluation can be used to define the program that works and that agencies can “go to scale” with it have been repeatedly disappointed (Elmore, 1996).

In practice, evaluation is most often called on to help with decisions about

---

\(^3\)Data on outcomes as well as data on program process can be useful for formative purposes.
improving programs, projects, and components. Go/no-go, live-or-die decisions about programs are relatively rare (Kaufman, 1976), and reliance on evaluation data to make those decisions is rarer still. Even when evaluation results show the program to be a failure, the usual use of the information is to patch it up and try again. Rare, too, is the use of evaluation in theory-oriented tests of program approaches and models. Despite considerably more writing about program theory in recent years, only occasionally are evaluations funded by people interested enough in current happenings to take a cool theoretical look. It is the search for improvements in components and strategies that supports most evaluation activity.

Decision makers may start out with global questions (Is the program worth continuing?), but they often receive qualified results (“There are these good effects, but…”) that lead them to look for ways to modify present practice. They become interested in the likelihood of improved results with different components, a different mix of services, different client groups, different staffing patterns, different organizational structure, different procedures and mechanics. One of the ironies of evaluation practice is that its methods are powerful for assessment of overall impact, suited to the uncommon go/no-go decision, but less well developed for understanding how and why programs achieve the degree of impact observed and how they can be improved.

At many stages people turn to evaluation for information that will help them modify the program. They expect evaluation to point the way to constructive change.

**Formative-Summative and Process-Outcome: Different Constructs**

At first blush, there seems to be close correspondence between the formative-summative distinction and the process-outcome distinction that I made in Chapter 1. Process evaluations examine what goes on inside the program while it is in progress, with a focus on such things as participant enrollment, activities offered, actions taken, staff practices, and client actions. Outcome evaluations put the emphasis on what happens to clients after their participation in the program as a result of the intervention. Formative and process come in the early stages and seem to be dedicated to improving the program, whereas summative and outcome are about what happens to participants at the conclusion.

However, the two pairs of terms have quite different implications. Formative and summative relate to the intentions of the evaluator in undertaking the study—whether to help develop the program or to render judgment on it. Process and outcome have nothing to do with the evaluator’s role but rather relate to the phase of the program studied.

Evaluators’ intentions can change over the course of a study. Evaluation sponsors who originally requested summative evaluation can realize that they want the information to help improve the program (a formative purpose), or the evaluator may change her intentions when she analyzes what the data say. The

important point is that outcome data are often helpful for formative purposes, and process data can help the policy community realize why outcomes came out as they did and therefore understand how much summative guidance they provide for the future. Accordingly, the more useful distinction for many purposes is that between the study of program process (looking at what goes on during the program) and the study of outcomes (looking at consequences for participants at the end).

**Compatibility of Purposes**

With all the possible uses for evaluation to serve, the evaluator has to make choices. The all-purpose evaluation is a myth. Although a number of different types of questions can be considered within the bounds of a single study, this takes meticulous planning and design. Not even a well-planned study will provide information on all the questions from all the people involved.

In fact, some purposes for evaluation are incompatible with others. Consider the evaluation of a particular educational program for slow learners. The teaching staff wants to use evaluation results to improve the presentations and teaching methods of the course, session by session, in order to maximize student learning. The state college of education wants to know whether the instructional program, based on a particular theory of learning, improves pupil performance. In the first case, the evaluator will have to examine immediate short-term effects (e.g., what students learn after the morning lesson). The teachers are not concerned about generalizing the results to other populations, and so the study needs neither comparison groups nor sophisticated statistics. The teachers care about these students in the here and now. The evaluator will want to maximize feedback of results to the teachers so that they can modify their techniques as they go along.

On the other hand, when evaluation is testing the proposition that a program developed from certain theories of learning will be successful with slow learners, it is concerned with long-range effects. Does learning occur and does it endure? The evaluation requires rigorous design so that observed results can be attributed to the stimulus of the program and not to extraneous events. The results have to be generalizable beyond the specific group of students. The instructional program should be faithful to the principles embodied in the theory of learning from which it derives, and it should be insulated from alterations during its course in order to preserve the clarity of the program that led to the effects observed.

Ideally, it is possible to achieve both an assessment of overall program effectiveness and a test of the effectiveness of component strategies. Textbooks on the design of experiments present methods of factorial design that allow the experimenter to discover both total effect and the effects of each experimental treatment. A combination of qualitative evaluation of program strategies and quantitative-experimental evaluation of outcomes can also serve the dual purpose.

However, in practice, juggling multiple questions is a difficult task. The constraints of the field situation may hobble the evaluation—too few students or too homogeneous a group of students, inadequate funds to support the necessary inten-
sity and diversity of evaluation methods, contamination of the slow-learners groups by receipt of other services, dropouts from the program, lack of access to records and data, and so on. These difficulties can affict an evaluation addressing one question. When the evaluator is trying to attend to two or three very different orders of questions at the same time, it becomes harder to give the requisite care to each. And sometimes they interfere with one another. For example, putting effort into control groups and long-term follow-up will take away resources from the study of immediate effects of teachers’ strategies. Giving quick feedback to teachers may lead them to make adjustments in the program that depart from the original theory. Such adjustment runs counter to the need for program stability in the interest of answers about the long-term viability of the program’s theory.

With money, time, evaluation know-how, and determination to raise different orders of questions, the evaluator can satisfy the informational demands of multiple groups. But it is a difficult and time-consuming job. Even when evaluators turn to informal procedures and less-than-elegant designs to accomplish the task, they often encounter frustrations. It is useful for the evaluator to identify the key questions raised by different actors and negotiate priorities among them. Or she may be able to plan a series of studies to be carried out in succession, one building on the findings of the earlier ones, to address additional questions. Whatever the plan turns out to be, it remains useful for the evaluator to know the priority among the purposes for evaluation. If a crunch comes, she can jettison the extra baggage and fight for the essentials.

The decision about whose uses shall be served is sometimes beyond the evaluator’s control. The manner in which the study is commissioned may determine whose questions get top billing. Whether the evaluator is inside the program agency or outside in a research organization or consulting firm also influences her access to influence. Either inside or outside, her location in the organizational structure will give her more or less opportunity to shape the study. It is to these issues I now turn.

### How Evaluations Are Commissioned

There are three basic ways in which an agency that wants evaluation goes about getting a study done: (a) hiring an evaluator on to the staff of the program agency (or assigning the task to an evaluator already there); (b) hiring an outside research/evaluation or consulting organization to do the evaluation; and (c) opening up bidding for the study to all applicants, through a procedure that is known in government circles as request for proposals (RFP). In the latter case, the commissioning agency sets out its requirements for the study and asks evaluation organizations to submit proposals explaining how they would conduct the study and at what cost. Very occasionally, there is a fourth route to evaluation. A university investigator develops a program on the basis of a theory of intervention and applies for a grant to study the effectiveness of the program and, concurrently, the validity of the theory on which it is based (e.g., Olds, 1988). In such a case, the investigator initiates the evaluation and seeks funding from a research grants agency to conduct the study.

### Purposes of Evaluation

#### Evaluator on Staff

Many agencies maintain an evaluation staff. Large school districts, mental health centers, colleges, and hospitals are among the agencies that often have a permanent evaluator or evaluation team. One of the main advantages is that the evaluator is there, ready and able to take on the job whenever evaluation is wanted. Another advantage is that the evaluator understands the agency and its programs. There is no need to orient her to the realities of program life and the constraints under which the agency operates. A disadvantage is that cost-conscious agencies do not want to maintain idle evaluation capacity and so usually keep the evaluation staff small—sometimes too small to undertake the complex study they want to have done or to do a needed second study when one is already in progress. Staff evaluators can also be drawn off to do other tasks in the agency. For example, evaluators in school districts are often called upon to administer testing programs. The small number of available evaluators can limit the type of evaluation that can be done.

An agency can also hire an evaluation expert to do the work, as an extension of its own staff. The expert reports directly to the chief, and is for all intents and purposes a part-time adjunct staff member. Agencies also hire experts, often university faculty members, to serve as consultants to advise the staff evaluator or review her work periodically with suggestions for modification or further work.

#### Hiring an Evaluation Organization

When an agency doesn’t have its own evaluation staff, it often seeks an outside evaluation organization to do the study. Several types of organizations are available: for-profit firms, nonprofit organizations, institutes attached to universities, faculty members of a school or department of the university. The program agency may check into which groups are qualified to do the kind of evaluation it wants. Or it may go to one organization that it knows has a good reputation, or whose staff it knows, or whose price is right, or with which it satisfactorily worked on a previous evaluation. Or it can hold a limited competition, asking several organizations to explain how they would go about the task, and then select the one whose plans best fit its needs.

The hiring of an evaluation organization gives the program agency a wider choice than it would have if it were limited to its own staff. It can call on a range of skills, talents, and experience that may be lacking on staff, and it can contract for a larger study than it would be likely to undertake in-house.

#### Requests for Proposals

For government programs, the law limits the opportunity to choose one research/evaluation organization without competition. In order to prevent nepotism, cronyism, corruption, or sweetheart arrangements (where, e.g., the evaluator promises to find good results in order to get the job), the agency has to open the bidding for the contract to all evaluation organizations on an equal basis. The agency publishes

---

The government has two basic mechanisms for funding research: contracts and grants. With a contract, government sets the parameters of the study and awards the contract to the organization whose proposal to fulfill those parameters best suits its purposes. Contracts thus give the government major control over the content, methods, timing, and reporting of studies. Grants are awarded to researchers who initiate their own study and apply for funding. The applications are reviewed for merit, but there is no competition to do the same work. Most government evaluations are done under contract.
a notice about the availability of the RFP in the *Commerce Business Daily*, a publication that lists notices of all goods and services that the government plans to buy. (Yes, it really appears daily. The government purchases amazing numbers of things.) Any organization that is interested in an evaluation RFP can submit a proposal. A RFP is a request for proposals. A RFP asks bidders to submit a proposal, which includes a description of the proposed project and a plan for its execution. The RFP sets out the nature of the desired study. The specification can be extremely detailed in terms of sites, number of cases, design, measures, times for data collection, and so on, or it can be fairly general. It also specifies the criteria on which proposals will be judged and the date on which proposals are due.

Sometimes the agency holds a bidders' meeting to which all prospective applicants are invited. There, agency staff answer questions that are raised about the desired study. Again, the procedure is designed to create a level playing field. No one bidder gets inside information. Word about the agency's intentions is given to all at one time and one place.

The next step is for respondent organizations to submit proposals to do the work. Separately they submit budgets detailing the costs they expect to incur in completing the evaluation—and thus the price of the work. After the proposals have been received, the agency assembles a group of reviewers to judge their adequacy. Reviewers often include outside experts, such as university faculty, and consultants, as well as inside staff. The substantive proposals are usually reviewed without reference to the budget statement. Only after the top proposals have been selected as finalists are budgets examined. If organizations submit budgets that are orders of magnitude too high or too low, the staff may conclude that the applicants don't understand the task. In that case, they may jettison the substantive proposal.

Reviewers judge substantive proposals against the standards set by the agency. Criteria will include such things as understanding of the agency's needs, capability of staff who will conduct the evaluation, responsiveness of the proposal to the RFP, organizational resources to be devoted to the study. For large contracts, the top bidders can be invited to make oral presentations of their proposals and answer agency questions. After one or more rounds of review, the winning proposal is selected. Even at that point, the government can still negotiate changes in the organization's proposal (or budget) to make it better suit agency purposes.

The RFP procedure represents a fair contracting method, one that is used in government acquisition of thousands of items, from buttons to fighter aircraft. One advantage is that it opens the bidding to all comers. Another is that it tailors the evaluation to the factors that the agency sees as important. A disadvantage is that, when RFPs are very detailed, the mechanism limits the opportunity for evaluation organizations to exercise much creativity and initiative in the design of the study. Evaluators complain that acquiring evaluations is not like acquiring buttons; they are not standardized or interchangeable. When the government agency exercises overly tight control, it risks losing the quality and ingenuity that an outstanding evaluator can bring to the task. Nevertheless, staff in government evaluation offices are usually competent and experienced, so that requirements generally make sense and ingenious evaluators find ways to propose approaches and techniques that will make the study even more useful than the original RFP anticipated. Another disadvantage is that the RFP procedure is onerous and time-consuming for evaluation organizations. Small or new organizations don't have the resources to absorb the costs of preparing many losing bids.

At the levels of state and local government, simpler versions of the federal RFP procedure are usually followed. Foundations and nonprofit agencies sometimes choose to go the RFP route, too. However, they often invite only a few evaluation groups that they know and trust to understand their issues.

**Extent of Influence**

The mechanism by which the study is commissioned will affect how much influence the evaluator will have on defining the questions that the study addresses. An insider can have a good deal of influence, but not if she sits somewhere deep in the bowels of the agency. An evaluator called to the task because of her expertise and renown can help to determine which purposes the study will serve, but not if she is hamstrung by an elaborate and highly specific RFP. The evaluator in each situation has to learn how to maneuver within the constraints of setting and procedures.

**Inside versus Outside Evaluation**

Parallel to the issue of the commissioning mechanism is the issue of location. Whether the evaluator is a member of the agency staff or an outsider affects not only how much say she has in determining the purposes to which the study is directed; her location has other consequences as well. There is a long tradition of controversy about whether in-house or outside evaluation is preferable. The answer seems to be that neither has a monopoly on the advantages. Some of the factors to be considered are administrative confidence, objectivity, understanding of the program, potential for utilization, and autonomy.

**Administrative Confidence**

Administrators must have confidence in the professional skills of the evaluation staff. Sometimes agency personnel are impressed only by the credentials and reputations of academic researchers and assume that the research people it has on staff or can hire are second-raters. Conversely, it may view outside evaluators as too remote from the realities, too ivory tower and abstract, to produce information of practical value. Often it is important to ensure public confidence by engaging evaluators who have no stake in the program to be studied. Competence, of course, is a big factor in ensuring confidence and deserves priority consideration.

**Objectivity**

Objectivity requires that evaluators be insulated from the possibility of biasing their data or its interpretation by a desire to make the program look good (or bad). Points usually go to outsiders on this score. They have no obvious stake in the program and can afford to let the chips fall where they may. Furthermore, they are seen by others

---

'Recently a number of federal agencies have discontinued holding bidders' meetings on the grounds that they have become unproductive. Prospective applicants have reportedly become wary of asking questions at the meetings for fear that they would be disclosing their plans to competitors.'
as impartial. While fine evaluations have been done by staff evaluators of scrupulous integrity, there is a lingering suspicion that they will try to please agency superiors by presenting the program in the best possible light. A wish to get along with colleagues in the program part of the agency may exert a subtle press on the evaluator to downplay negative information. When I was a staff evaluator in a small agency, I didn’t want to antagonize my program colleagues and have to eat lunch alone.

But outsiders are not always immune from such pressures. They can get so engaged with practitioners and clients and so involved in the processes of the program that they become almost unwitting advocates for it. It even happens that an outside research firm will try to sweeten the interpretation of program results (by choice of respondents, by selection of data) in order to ingratiate itself with agency managers and be first in line for further contracts.

In all events, safeguarding the study against bias is important. While no researcher, inside or out, is totally objective (we all have our beliefs and values, which inevitably color our methodological choices and interpretations), evaluation sponsors should seek conditions that minimize biases for or against the program. A good test of whether they have succeeded is whether the final report is respected as a valid account by program stakeholders with conflicting interests.

**Understanding of the Program**

Knowledge of what is going on in the program is vital for an evaluation staff. They need to know both the real issues facing the agency and the real events that are taking place in the program if their evaluation is to be relevant. It is here that in-house staff chalk up points. Because they usually sit close to the action, they are likely to see and hear a great deal about what goes on. They hear the gossip, they know quickly about any cataclysmic event, and they see the faces of clients and staff. Of course, outsiders, too, can find out about program processes if they make the effort and are given access to sources of information.

**Potential for Utilization**

Utilization of results often requires that evaluators take an active role in moving from research data to interpretation of the results in a policy context. In-house staff, who report results and perhaps make recommendations on the basis of results, have many opportunities to bring them to attention of practitioners and managers. They attend staff meetings, go to conferences, meet informally with staff, and thus are often better able to get evaluation results a hearing. But sometimes it is outsiders, with their prestige and authority, who are able to induce the agency to pay attention to the evaluation. If need be, they can go around the chain of command and bring findings to the attention of top brass, oversight agencies, client groups, or the media.

**Autonomy**

Insiders generally take the program’s basic assumptions and organizational arrangements as a given and conduct their evaluation within the existing framework. The outsider is able to exercise more autonomy and take a wider perspective. While respecting the formulation of issues set by the program, she may be able to introduce alternatives that are a marked departure from the status quo. The implications she draws from evaluation data may be oriented less to tinkering and more to fundamental restructuring of the program. Such fundamental rethinking of program parameters may be especially valuable for unsuccessful, routinized, tradition-encrusted programs—to help them see a broader horizon and imagine an alternative future. However, evaluation interpretations that look toward major program change are probably not likely to be used in the near term. Old habits and interests keep an agency moored in familiar channels. But in the longer term, especially if reinforced by influential actors, the wider perspective may help to bring fresh air and fresh ideas into the tradition-bound agency.

Even if the evaluator does not challenge customary boundaries, the outsider has the chance to raise issues that would be uncomfortable for an inside evaluator to raise. Based on data and observation, she can question activities and ideas that are dear to program people’s hearts, such as the fact that service isn’t available on weekends or that staff are giving priority to people with less serious problems. She can proffer recommendations that go beyond the commonplace. She doesn’t have to abide by the prevailing conventions.

**Balance**

All these considerations have to be balanced against each other. Costs are a factor in some cases. Previous commitments may be relevant. There is no one best site for evaluation. The agency must weigh the factors afresh in each case and make an estimate of the side on which the benefits pile up.

**Level in the Structure**

Another factor that influences the nature of the evaluation performed and the uses to which it can be put is the location of evaluation within the organizational structure. Whoever actually does the evaluation, she is enabled or constrained by the organizational level to which she is assigned. The inside evaluator is responsible to some level in the hierarchy. The outside evaluator, too, receives her assignment and reports her progress to (and may get intermediate direction from) the holder of a particular organizational position.

The evaluation should be placed within the organizational structure at a level consonant with its mission. If it is directed at answering the policy questions (How good is the program overall?), evaluators should report to the policy level. To have the evaluation under the control of the people who run the program is to invite gross or subtle efforts to make the study come out “right.” Program people can put up roadblocks in countless ways, from directing evaluators to sheer away from questions of program effectiveness to cutting off their funds for computer paper and photocopying (a true case). If the evaluator nevertheless perseveres and produces a report that disparages the overall effectiveness of the program, managers are likely to stuff the report at the program level so that it never receives consideration in higher councils where something could be done.

Surely there are many program directors who are scrupulously fair and would not dream of interfering with an evaluation. And they can do good things, too—
such as ease of access to sites, organize conferences with stakeholders, and make plans to apply evaluation results to the improvement of the program. Still, it is unwise to put too much temptation in their way by giving them the opportunity to exercise full control. Furthermore, when the evaluation is under the control of program personnel, however nobly they behave, the study tends to lose credibility in the eyes of others.

When the basic shape of the program is unquestioned and the evaluation issue centers on processes and program components, the evaluator should probably be responsible to program managers. It is program managers who have authority to institute the kinds of internal changes in programming that would be suggested by the study. They may have the responsibility for selecting local projects and for supervising staff who deliver services in ways that evaluation may find need to be modified or reinforced. They can hire new staff or new categories of staff. So evaluation that is geared to questions about activities and operations fits comfortably at the program management level.

In deciding on appropriate location, two criteria apply. One is who can control what the evaluation does and says. The other is who can put the findings to use. If program managers oversee an evaluation, they may be wary of letting questions be asked or information released that might reflect poorly on their performance. On the other hand, they have almost daily opportunities to put results to use when they bear on the regular operations of the program. In contrast, policymakers, who often have less at stake if evaluation findings are negative, are unlikely to see the evaluation. But they have fewer opportunities to make the day-to-day modifications in program operation that effective use of evaluation results might involve.

Wherever the evaluation project sits in the structure, it should have the autonomy that all research requires to report objectively on the evidence and to pursue issues, criteria, and analysis beyond the limits set by people pursuing their own or their organization’s self-interest. The charge to the evaluator is to report as well and as accurately as she knows how, in order to understand and interpret the phenomena under study. If disputes arise between the evaluators and the evaluation sponsor, there should be a set procedure that specifies how the issue will be resolved. The procedure may involve appeal to a higher level in the organization or to an outside advisory committee, or a mediation process may be established to be staffed by managers from another program unit or a research and planning office. At the outset it may look like a bureaucratic frill to worry about procedures for dispute resolution, and the evaluator may be reluctant to talk about disputes before they happen. But disputes do come up, and by the time they have erupted, it is often too late for sound and sober judgment. Tempers are frayed. The time to set rules and regulations for resolution is at the beginning. It is important that the process is in place, so that if conflict comes up, it can be resolved in an orderly fashion.

Whose Use Shall Be Served?

This chapter considers a glittering array of possible users of evaluation results. Each of them has his own interests and ideas, his own questions and concerns, and a unique palette of opportunities for putting the results to use. The players in the drama include the following:

- Funding organization (government, private foundation)
- National agency (governmental, private)
- Local agency
- Directors of the specific project
- Direct-service staff
- Clients and prospective clients of the project
- Directors and managers of programs with similar purposes elsewhere
- Program designers
- Program consultants, often academics, who consult for and advise decision makers
- Scholars in the disciplines and professions who build knowledge and teach oncoming generations
- The public, which, in the form of public opinion, sets the general direction of, and limits to, policy and programming

The array of people with a possible stake in the evaluation are often called stakeholders. The term is appropriate, but inexact. Appropriate, because these actors do have a potential stake in what the evaluation does and what it finds out, but inexact, because there is little consensus on who is and is not among the stakeholding elect. A body of current lore holds that the evaluator should enlist stakeholders in the definition and conduct of the evaluation. But some of the groups on our list will not even know that an evaluation is in progress until the final results are reported. This is true, for example, of directors of similar programs elsewhere and scholars in the disciplines. Some groups will know about the evaluation but be uninterested in it, even though they recognize a passionate stake in the program. Such is often the case for client groups (Weiss, 1983a, 1983b). I return to the subject of stakeholders in Chapter 5.

Which purposes shall the evaluation serve and for whom? In some cases, the question is academic. The evaluator is on the staff of the organization—federal department, school system—and she does the job assigned to her. Still, even in such a case, when choices are constrained, the evaluator has options. She can make suggestions about central evaluation questions, design, measurement, and possible uses for the data. Since she is the expert on the subject of evaluation, her words will usually carry weight. If she has done a good bit of preparatory exploration, she will be able to offer cogent reasons for her suggestions and ways in which her plans serve the most important needs. But ultimately, if her superiors disagree, she will have to go along with what they choose to do—or leave the job.

When the evaluator works in a private research firm or an academic institute or department, she has wider latitude. In an outside research organization, she and her colleagues have the opportunity to elaborate on the RFP and write into their proposal emphases and methods that they believe are most appropriate. In for-profit
research firms and not-for-profit research institutes dependent on contract funds, evaluators often do not want to depart too drastically from the funder's specifications for the study. They want to win the contract and are afraid that too much independence of mind will jeopardize the award. In academic settings, evaluators may be willing to exercise more discretion in proposing the kind of study they believe is suitable. They can take bigger risks because their salary doesn't usually depend on winning the study contract and the academic value system prizes high-quality research and independent judgment. At the same time, academic evaluators may have less motivation to be relevant to the program agency's problems and useful to its purposes.

Wherever the evaluator is located, she can, if she tries, often negotiate some of the terms of the study. Program personnel and funders on occasion display a surprising lack of clarity about what uses the evaluation shall serve. When the issue is squarely raised, different parties in the program system may disagree among themselves. Such confusion offers the evaluator room for maneuver.

If she can help shape the basic focus of the study, the evaluator will consider a number of things. First is probably her own set of values. A summer program for inner-city youth can be evaluated for city officials to see if it cools out the kids and prevents riots. The evaluator may want to view the program from the youths' perspective as well and see if it has improved their job prospects, work skills, and life chances. It is important that the evaluator be able to live with the study, its uses, and her conscience at the same time.

Beyond this point, the paramount consideration is who is listening. If there is a decision coming up (e.g., about introducing Web sites for transmitting program information), and people who will be involved in that decision want to know what the evaluation can tell them, the evaluator will take their concerns seriously. In general, the evaluator scouts the terrain to see whether key decisions are pending, who will make them, when they will be made, and on what grounds they are likely to be made. If she finds that evaluation has a chance of influencing decisions, she will give special attention to questions that have potential decisional payoff.

Often there is no decision pending, or people who will have a voice in the decision are not interested in hearing the findings from evaluation. Still there are likely to be potential audiences who want to learn from the study and consider the results, and there will be others who do not. When local program managers are conscientiously seeking better ways to serve their clients while the policymakers at higher levels are looking primarily for program vindicators, the local managers' purposes may deserve more attention. On the other hand, if the locals want a whitewash and the higher levels want to understand how the program is using its resources, the evaluator should place more emphasis on the higher echelon's concerns.

However, even people who intend to use evaluation results for specific purposes do not make their decisions solely on the basis of the evaluation. Although the news is a blow to those who expect the world to run on rational principles, evaluation findings alone rarely determine decisions. Other considerations almost inevitably enter in. Each individual in the program community has a set of values and beliefs that influence his stand—for example, "Poor people deserve access to the same mental health services as everyone else," or "Voluntarism is a better way to go than government-provided services." He also often has interests—self-interest—in promoting one course over another. In addition, he has a stock of information—from his prior experience with the program, reports from staff, the grapevine, and from his professional experience and training. These three I's (ideology, interests, and information) are in place, and the new information that comes from evaluation has to contend with them. Essentially, it interacts with them in influencing the positions that people take (Weiss, 1983).

Moreover, there is a fourth I, the institution. Program decisions are made within an organizational context (Weiss, 1993). The organization has a history, a set of operating procedures, and a culture, and these institutional elements exert pressures and set constraints. Evaluation findings that imply a need for major change may run into organizational roadblocks, while findings that support the direction in which the organization is moving can be readily embraced.

Then there are all the practical issues that come up when program people seek to implement big changes based on evaluation. They have to consider such issues as the cost of the changes, support or opposition from important political figures, the interest of program staff in keeping the program going, the prevailing political climate, clients' pressure for continued attention, availability of appropriate staff, and so on. In short, when findings suggest fundamental redirection, evaluation findings will contribute to discussion and debate but rarely single-handedly carry the day.

Accordingly, in recent years evaluators have come to pay more attention to the prospects that evaluation can contribute to organizational learning. Organizational learning is often a slow incremental affair, punctuated with occasional spurts of change. There have been many cases where evaluation results have been ignored in the early months after a study's conclusion only to surface two or three years later as one of the rationales for changes in organization activity. In fact, experience has shown that evaluation findings often have significant influence. They provide new concepts and angles of vision, new ways of making sense of events, new possible directions. They puncture old myths. They show that some issues expected to be important are not important and that unexpected issues have become salient. They sometimes occasion a reordering of priorities and a rethinking of program directions. As they percolate into the consciousness of the program community, they shape the ways in which people think about problems and plan their responses.

This view of evaluation as provider of enlightenment is a far cry from old expectations that evaluation would point the way to big decisions about program and policy. The interpretation here may look like a retreat toward insignificance. On the contrary, it is optimistic to recognize that evaluation influences what people know and believe, and that these ideas have long-lasting importance for the kind of world they create. This view celebrates the importance of information. It avoids succumbing to the disillusionment that some evaluators face when the results of their studies are not immediately put into practice. It treasures the instances when the program and policy communities find a direct way to use evaluative results for improvement. It appreciates the complex processes of democratic decision making that do not give any person (however well informed) the authority to make decisions unilaterally but require the participation of many parties in developing new courses of action. On a practical note, it gives evaluators clues on how to move their studies toward more and
more responsible use. If the evaluator can identify people in the organization who are willing to listen, and keep up a dialog with them over an extended period of time, she can perhaps influence the scope and pace of the organization’s learning trajectory.

Finding attentive listeners and champions of change can be an effective strategy. But the evaluator should not place too much faith in her ability to locate the key decision maker. In fact, there are almost never one or two decision makers. Decisions in our system, particularly in public agencies, almost always involve a great many people in a variety of different offices. Sometimes the legislature and the judiciary get involved, too. Multiple actors have a say.

Therefore, from the start the evaluator should take account of the questions, complaints, and ideas of the variety of people who will take part in making decisions about the program. If the school principal wants an evaluation of a program for gifted and talented students, remember that any decisions that follow on the evaluation will not be his alone. The school superintendent will need to agree, and probably the school board as well. Nor will much happen on the ground (in the program, in the classrooms) unless the teachers go along with the decision.

So when we think about whose use should be served, let us not think of the single decision maker. As Cronbach and Associates (1980) have written, program decisions are made not in a context of command, where one person has all the power and his word is law, but in a context of accommodation, where many people negotiate to reach a decision. The key implication, then, is that the evaluation needs to consider not only the questions of the project manager or program director but also the concerns of other people who will take part in decision making. Beyond that, there are people who help set the climate of opinion around the program, people who help organize the funds to run the program, and people who join up for, or fail to participate in, the program. These widening circles of people who are affected in some way by the program—and potentially by the evaluation of the program—are the stakeholders, and the evaluation needs to be alert to the kinds of use that they may make of evaluation results.

The next task, then, is designing the evaluation to provide the answers that are needed. Based on her best estimate of likely use, the evaluator has to make decisions on the basic plan for the evaluation (Chapter 4), measures to be used (Chapter 6), sources of information (Chapter 7), its design (Chapters 8, 9, 10, and 11), and modes of analysis (Chapter 12). Finally she will have to write the report(s) and get the word out to the people who can use the results (Chapter 13).

Summary

Agencies undertake evaluations for a host of rational and nonrational (or at least, nondecisional) reasons. An evaluator newly arrived on the scene should try to understand the purposes that evaluation is expected to serve and the range of expectations (some of them conflicting or contradictory) held by people in different locations in the program community. Realistically she has to design the study in ways that serve the information needs of those who commission the evaluation, but at the same time she has to be aware of the needs and interests of many others in the environment whose actions will affect what happens during and after the study.

I suggest that the evaluator pay attention to the questions of those who face fairly immediate program decisions, those who are most receptive to learning from the evaluation and conscientiously considering its results, and those (like clients) who have few other avenues to make their interests known. Balancing the competing claims calls for a high order of political, as well as conceptual and methodological, skills. Sometimes it requires re-negotiating the original parameters of the study. Since program and policy decisions are influenced by people in many locations, an evaluation is likely to be most influential when it is responsive to questions from a variety of stakeholders.

The distinction between formative and summative evaluation brings a modicum of clarity to demands upon the evaluation. Formative evaluation is designed to help program managers, practitioners, and planners improve the design of the program in its developmental phases. Summative evaluation is designed to provide information at the end of (or at least one cycle of) the program about whether it should be continued, dismantled, or drastically overhauled. Formative/summative evaluation refers to the intention with which the evaluator works. Process/outcome evaluation refers to the phase in a program’s life at which evaluation is carried out: during its course or at its conclusion. The two sets of terms, while overlapping, refer to different constructs. I suggest that most programs are never “finalized” but continue to develop as they go on. Therefore, formative data continue to be needed to help in their improvement. Both process and outcome data can be helpful for formative purposes.

Evaluation can be performed by staff within the program agency or by outsiders. Each location has advantages. Inside evaluators understand the organization, its interests and needs, and the opportunities that exist for putting results to use. Insiders also have multiple opportunities to bring results to attention. Outside evaluators have greater autonomy, greater prestige, often a wider range of skills, and a chance to bypass the agency to get results a hearing.

Whether inside or out, evaluators should report to the level in the agency structure that allows them the most appropriate mix of autonomy and access to both information and influence. Reporting to policy levels of an organization will usually minimize attempts to bias the evaluation; reporting to working levels will allow the evaluator to understand what staff do and why they do it and to establish good working relationships with many of those who can later put findings to use.

Evaluation has traditionally stressed the provision of outcome data as a tool to help in decision making. Recent years have seen increased awareness that evaluation findings almost never determine program decisions alone. Policymakers and program managers have to take many other considerations into account. Evaluators should also be alert to the fact that, although some uses of the study are immediate and obvious, much use will be conceptual and difficult to discern with the naked eye. Organizations and members of the policymaking community do not always apply results directly to an immediate decision. More often they learn from evaluation, and their learning affects program decisions subtly and gradually. Nevertheless, the evaluator has to shape the evaluation to answer important questions, so that what the varied audiences learn will lead to policy and program improvements.