2
Planning an Evaluation

The initiation of a program evaluation may arise in several different ways. The program personnel themselves may initiate the request. The program sponsors (such as the funding agency or the head of the institution in which the program functions) might order an evaluation as a condition for beginning or continuing a program. The evaluation or quality assurance department might seize the opportunity for a productive project and submit a preliminary proposal to either the program sponsors or the program director.

Whatever the source of the proposal, evaluators need to address a number of issues and make certain decisions before the collection of data begins. Before committing their resources, they need to learn about the goals and mechanics of the program, about the people who sponsor the program, about the program personnel, and about groups that may resist the evaluation of the program. After obtaining this information, evaluators must then decide whether an evaluation can be done. If the program is planned in a way that permits evaluation, then a decision needs to be made as to whether an evaluation should be done immediately, whether it should be done in the way in which it is proposed, or whether it should be done at all.

In the first part of this chapter, we provide an overview of the order and manner in which the above issues can be approached, identifying the steps to be taken between the time of the initial proposal and the beginning of data collection. The time and effort devoted to each step will vary, depending on the complexity of the program, the relationship of the evaluator to the program sponsors and personnel, and the urgency of time constraints. Some steps, such as selecting or developing measures, are very complex, and Chapter 3 will explain them in detail. In the second part of this chapter, we deal with potential sources of resistance to program evaluation.

SIX STEPS IN PLANNING AN EVALUATION

Step 1: Identify relevant people.

The first thing effective evaluators do is identify relevant people; that is, those who have a serious interest in the program and whose lives may potentially be affected by the evaluation. Relevant people are those who are heavily personally involved in the program, who derive some or all of their income from the program, whose future status or career might be affected by the quality of the program, or who are the clients or potential recipients of the program services. Such groups have been called stakeholders (Bryk, 1985), since they have a stake in the program and its evaluation.

Program personnel should be considered first. They will usually be more personally involved in the program than either the sponsors or the clients. The program director will be the key person to whom evaluators will relate during the entire project. It will help to learn as much about the background, interest, attitudes, and reputation of the program director as possible. In addition, other people involved in the delivery of program services must not be overlooked. It is also important to involve them in the planning stage so that they will assume ownership of the project and provide maximum support during the data collection stage. If possible, evaluators should learn about the relationship between the director and other program personnel, so that they can be dealt with more effectively during the planning meetings described below. For example, if the director tends to be a dictatorial person, it will be more difficult for evaluators to draw out the ideas of other personnel and determine when a consensus exists or when a more power decision has been made by the director.

Program sponsors should be considered next. At times program personnel are the sponsors. In other cases sponsors will be funding agencies, governmental bodies, or administrators of the institution in which the program exists. Often there will be a single person to whom the evaluators are to relate. The funding agency may delegate one or two representatives to handle the commissioning and supervising of an evaluation project. In an institution such as a hospital, the program sponsor might be the vice-president or administrator to whom the program director reports and who is ultimately responsible for the management of the program. Contact with the program sponsor will be especially important during the initial stages.
of planning and at the end of the evaluation. At the beginning it is important that the sponsor fully supports the proposed evaluation and that all the program personnel are aware of that support. At the end of the evaluation, it is important that the report be presented to sponsors in such a way that it will be fully utilized in decision making. The presentation of the report and effective utilization of findings will be discussed more fully in Chapters 12 and 13.

Finally, the clients or recipients of the program services need to be identified. The amount and type of contact with the clients will depend on the nature of the program and of the evaluation. For example, if a cancer care unit is to be evaluated, are the program recipients only the cancer patients, or are the relatives of the patients also included? Are all cancer patients included, or only those with an established diagnosis of cancer, or with certain types of cancer?

**Step 2: Arrange preliminary meetings.**

Before a final decision is made to undertake an evaluation, and before the writing of a detailed proposal, it is advisable to meet with relevant people to gather background information on five questions: (1) Who wants the evaluation? (2) What type of evaluation is desired? (3) Why do they want it? (4) When do they want it? and (5) What resources are available?

**Who wants the evaluation?** The ideal situation occurs when both the program sponsor and the program personnel desire to have the program evaluated. In this instance the evaluators will usually be interacting with competent people who are secure in their professional expertise, who are open to suggestions for improvement, and who welcome the opportunity to have documentation for what they feel is a successful program.

If the program sponsors initiate the evaluation proposal—either without the knowledge of, or against the desires of, the program personnel—the evaluators are faced with making the program personnel comfortable with the goals and methodology of the evaluation before data collection begins. If evaluators do not succeed in doing this, they face the possibility of open opposition or a lack of essential cooperation on the part of the program personnel. On the other hand, when the program personnel see the evaluators as “allies” rather than as “the enemy,” they are more likely to give the evaluators both much-needed assistance in the difficult job of data collection and valuable insights into the interpretation of the data.

If the program personnel initiate the evaluation proposal—either without the knowledge of, or against the desires of, the program sponsors—the evaluators need to convince the sponsors of the usefulness of the evaluation, or it will be very difficult to have any worthwhile changes effected. Sponsors who do not assume ownership of the evaluation project during the planning stage are inclined to allow the finished reports to lie unused in their offices.

**What type of evaluation is desired?** Early in the evaluators’ meetings with program sponsors and personnel, it will be clear that the term program evaluation does not have the same meaning for everyone. More often than not, program personnel will be thinking in terms of a formative evaluation that will help them modify and improve the program. On the other hand, program sponsors may desire a summative evaluation; they may be under pressure to divert resources to another program and must decide whether or not to continue the present program. Finally, some program personnel and sponsors will have little awareness of the whole concept of program evaluation and will therefore be expecting individual performance appraisal.

The task of evaluators at this point is to clarify the pertinent concepts for those who do not understand them and to help the relevant people decide what type of evaluation best meets their desires, needs, and resources. The choice is seldom between one type of evaluation and another. Often it is possible to incorporate some elements of various types of evaluations into the total scope of the project, depending on the complexity of the program goals and the resources available for the project. For example, in the evaluation of “Sesame Street” conducted by Cook et al. (1975), various types of assessment were included in the overall evaluation. He assessed (1) whether the program reached its target audience; (2) effectiveness by analysis of learning scores; (3) the magnitude of effects relative to the magnitude of the need that gave rise to the problem; (4) the ratio of benefits to cost; (5) the aspects of the complex home viewing situations that led to learning, and the conditions of TV viewing that most promote learning; and (6) the value of program objectives—that is, the importance of stimulating the growth of all social groups of preschool children versus that of narrowing the academic achievement gap between economically advantaged and disadvantaged children.

**Why is the evaluation desired?** Closely tied to the previous question is the issue of why evaluation is desired. The commissioning of an evaluation is rarely the product of an inquiring scientific spirit. More often it meets the needs of political forces. Evaluation furnishes ammunition for institutional combatants in fluid lines of battle and in transient alliances. Program sponsors often need to satisfy constituents and want to keep politically advantageous programs alive. Effective evaluators will put a high priority on identifying the reasons why the evaluation is desired. Were there some groups in the organization who objected to the evaluation? What were their motives? Is there real commitment among the program personnel and pro-
example, an administrator may use program evaluation as a ploy to avoid making a decision. Administrators who are pressed about the viability of a given program can buy time by saying that the program is being evaluated; by giving as little support as possible to the evaluators, they ensure that the evaluation will take a long time to be completed. If there is a need to buy still more time when the evaluation is completed, administrators may appoint a committee to study the evaluation report. Evaluation is also inappropriate when administrators know what decisions they will make but wish to go through the charade of program evaluation to give their decision legitimacy.

When is the evaluation desired? The scope of an evaluation project can be limited by situational factors. If the evaluators are given a deadline, they may be forced to make choices that are not ideal. For example, measures specifically designed for a program may have to be foregone in favor of existing questionnaires or instruments, although these existing instruments may not have the desired reliability, validity, or applicability. The number of respondents or participants in the study may have to be limited, thereby decreasing the probability of detecting the true effect of the program. Adequate control groups may not be immediately available, although they might be available at a later date. Evaluators must decide whether they should proceed under less than ideal circumstances, delay the evaluation, or refuse to do the evaluation in order to avoid the possibility of generating misleading results. For example, Carey and Posavac (1979) were invited to evaluate a new cancer care center five months before it was to be introduced within an acute care hospital. A five-month lead time was not long enough to develop new measures for all the program objectives, nor to collect data from a large sample of cancer patients before the institution of the center. To save time, the evaluators measured patient moods with an existing instrument used in previous research with cancer patients.

Even when there is no deadline, program personnel may be eager for results and push evaluators for an estimated completion date. It is usually better not to commit oneself to a deadline that can be met only if everything occurs on schedule. Murphy's law ("If anything can go wrong, it will") is especially applicable in program evaluation: Data collectors become ill; respondents are unavailable; the mail is slow; computers break down when they are most needed.

When an estimated date for the completion of the report is required, extra time should be allotted. If evaluators finish behind schedule, they are suspected of being disorganized at best, or incompetent at worst. On the other hand, when they finish ahead of schedule, they build a reputation for speed and efficiency. Nevertheless, it is important to remember that evaluators work in settings in which time is important—more important than in the setting in which university-based research is usually conducted.
Planning an Evaluation

dressed or the program intervention being made; (2) a lack of clear logic—
that is, the connection between the program intervention and the immediate
outcome is not understood clearly enough to permit testing; and (3) a
lack of management—that is, those in charge of the program lack the au-
thority or ability to follow through on evaluation results. Cronbach (1982)
suggests that evaluations should only be done when the evaluation and its
recommendations are likely to get a fair hearing.

It is necessary to point out that at times evaluators do not have the
freedom to decide not to conduct an evaluation. In-house evaluators may
be assigned to conduct an evaluation. Consultants may accept a contract in
order to maintain a staff, even if conditions are not perfect for carrying out
the best evaluation. The main consideration is to make a conscious
decision—not merely to drift into a project just because it is there. Chapter
4 suggests that when faced with an unevaluable program, an evaluator may
be able to help program sponsors redefine an evaluation project.

Step 4: Examine the literature.

Evaluators often work on a wide variety of projects. An independent
evaluator, for example, may evaluate a school system, then a community
program for alcoholics, and then the organizational climate in a bank. Such
a researcher must become familiar with an enormous amount of literature.
Although this is difficult, some evaluators prefer such work; they enjoy
the variety and do not like to restrict their efforts to a single area of interest.

Evaluators who specialize in limited areas have a less formidable task
in drawing up a proposal. If they have been exclusively evaluating educa-
tional systems, mental health programs, or personnel problems in organi-
zations, they develop an extensive knowledge of evaluation in their spe-
cialty. They can also build more easily on their own past research and
contribute to the development of theory in that area.

However, when evaluators work in an area that is new for them, it is
important to make a careful search of the literature before designing or
developing new instruments. Evaluators should learn from the successes
and failures of others and get a picture of the methodological, political, and
practical difficulties that must be overcome.

Depending on the type of evaluation, computerized search services
might be a good place to begin a search. Data on Program Evaluation
(DOPE), MED-LINE, Psychological Abstracts Search and Retrieval
(PASAR), and Educational Resource Information Center (ERIC) are four
such useful systems. A search of professional journals recommended by
program personnel is also helpful. After a few useful articles have been
identified, the bibliographies of these articles will provide additional
references.

While reading the articles, evaluators should keep these key questions
in mind: Has any evaluation been done on this type of program? What de-
signs were used? Were new measures developed? How reliable and valid were the measures? What type of statistical analysis was used? Was it appropriate? Is there a consensus among the findings of various studies? If there is conflict, is this due to sampling procedures, design, or interpretation of findings? What issues were not addressed or investigated?

**Step 5: Determine the methodology.**

After reviewing the literature, the evaluators are ready to make some methodological decisions regarding strategy and design, population and sampling procedures, control or comparison groups, operational measures, data collection, and statistical analysis. Separate chapters are devoted to some of these issues, but it is helpful to examine the main issues now so that the reader can better understand their interrelationships and how they fit into the planning stage of an evaluation.

**Strategy and design.** Strategy and design will largely be determined by the type of evaluation needed. Will the project be a need evaluation? A process evaluation? An outcome evaluation? An evaluation of cost effectiveness? Perhaps the project will include more than one kind of evaluation.

Strategy and design will also be influenced by whether the program is already in operation or whether it is still in the planning stage. If a new program is to be initiated, its effectiveness can be evaluated by gathering data on one or more occasions prior to the introduction of the program and comparing these data to information collected after its introduction. If the program can be introduced to some segments of a target population on the program can be introduced to some segments of a target population on the program at staggered basis, this is more desirable than introducing the program to the entire population at the same time. With this strategy evaluators can make multiple comparisons of the same group over time, and between groups.

**Population and sampling.** Once the target population (participants, respondents) has been identified, evaluators need to decide whether to include the entire population or to take a sample from that population. There are at least two arguments for including the entire population. The first is political or psychological—namely, people may be offended if they are not included. For example, the organizational attitudes of nurses at a given hospital could be accurately measured by taking a stratified random sample from the various nursing services. However, experience has shown that the hospital climate survey itself has a more positive impact on nurses' job satisfaction when all nurses are given the opportunity to express their views through the survey.

The second reason for including the entire population is methodological. If a sample is too small, there is a great danger of being unable to detect program effects even when the program is effective (a Type II error). For example, if a random sample of ten program participants is compared to a sample of ten who did not participate in the program, the program would have to have quite sizable effects before the results would be statistically significant. A naive use of statistical significance testing has been responsible for overlooking true program effects in some evaluations (Cohen, 1982).

There are two main arguments for using a random sample of participants rather than the entire population when the populations are large—namely, time and money. Program evaluation must be done within a budget, and usually there are some time restraints. When Andrew Greeley (1979) studied the effects of the Vatican Council on Catholic priests in the United States, he quite properly used a 10 percent stratified random sample of the 55,000 Catholic priests. Diocesan and religious order priests, priests of different age groups, priests from different geographical areas, and priests with different job categories (pastors, associate pastors, and those in special work) were represented in the sample. Greeley was able to get more than an 80 percent return from his sample and had sufficient numbers for studying subsets of the population. Because he chose to work with a sample rather than with the entire population, he kept within his budget and had the project completed in a reasonable time. Whether one chooses to include the entire population in the evaluation or to do a random sampling, a very important factor must be kept in mind—the probability that some people will drop out of the study. It is important to identify the characteristics of nonparticipants and those who drop out at various points of time. Both the rate of dropouts and the reasons for dropping out have a bearing on the conclusions to be made from the data collected.

If more than 50 percent of the target population or random sample are lost, the extrapolation of the findings to the entire population cannot be done with any confidence. Therefore, with a population of 1,000 respondents, evaluators will generally be able to make more valid interpretations if they take a 10 percent sample and obtain a 75 percent response rate (N = 75) than if they try to reach the entire population and obtain only a 30 percent response rate (N = 300).

The composition of the participant group will have an effect on subject dropout rates. For example, in a hospital setting it is more difficult to obtain a high percentage of returns from housekeeping personnel than from nursing personnel on a written employee questionnaire. This occurs because many housekeeping personnel have difficulty in reading English. In contacting patients in the aforementioned cancer care center study, it was difficult to obtain an adequate response rate because of the serious physical conditions of the target population. Therefore, the questionnaires had to be as short and simple as possible in order to obtain a response rate of 50 percent or better.
Control and comparison groups. When participants are randomly assigned to a group where they do not receive the services of the program, they constitute a true control group. When participants in a program are evaluated, they constitute a true control group. When participants in a program are compared to a group receiving the services of the program but not randomly assigned to a group, we speak of a comparison group (or a nonequivalent control group). If the goal of the evaluation is to determine whether the program caused a change in the participants, the ideal evaluation involves random assignment of participants to the treatment and control groups. (We will point out in later chapters that establishing causal evidence is often not the objective of program evaluations. See also Cronbach, 1982.)

Random assignment is frequently impossible; however, there may be ways to achieve the equivalent of this, even when the evaluators do not have total control of the situation. For example, in an evaluation of primary nursing in a hospital, the evaluators did not have control over which patients were assigned to the units with primary nursing or to those with team nursing. However, investigation showed that all physicians allowed their patients to be assigned to medical units on the basis of bed availability. This meant that although the evaluators had no control over the assignment of patients, situational factors resulted in random assignment. Demographic records of the patients in each group demonstrated that this system of random assignment resulted in two groups that were reasonably comparable.

One way of strengthening the interpretation of findings when one cannot have true control groups is to have more than one comparison group. For example, this was of great value in the cancer care center evaluation previously mentioned. On some criteria patients in the special cancer unit reported attitudes significantly more favorable than did one comparison group and significantly less favorable than did the other comparison group. On other criteria the patients in the special unit scored higher than both comparison groups. Therefore, with the use of two comparison groups, the evaluators were able to identify the advantages of the program with greater precision.

Nevertheless, one can question whether meaningful evaluation must always be comparative. Are there not some absolute standards of performance or quality against which results can be compared? For example, if one is evaluating a newly introduced drive-in banking service and 80 percent of banking customers are pleased, does one need a control or comparison group? It would not seem to be necessary if the program sponsors (the bank) had set an 80 percent satisfaction rate as their goal. Nevertheless, it

Primary nursing refers to a situation in which an individual nurse is assigned to give complete care to a few patients in a unit. It is distinguished from team nursing, in which a group of nurses works under a head nurse to provide care as needed to an entire unit of patients, perhaps alternating their responsibilities from day to day. Primary nurses work more closely with physicians in developing care plans and in discharge planning.

Selection of measures. The key idea with respect to measurement is to plan for multiple measures from multiple sources. Obviously, one of the main sources of data in program evaluation is the client. However, clients can provide data in more than one way. Clients can provide self-report measures by responding to direct questions about their perceptions and satisfaction with the program. They can also take tests designed to measure attitudes or moods that might have been affected by the program. Some such tests are described in the next chapter.

The behavior of clients can also be a source of data. For example, satisfaction with a new job-training program might be measured by the percentage of applicants who drop out of the new program as compared to other programs. Or the emotional adjustment of various groups of widows might be measured by the percentage of each group who remarried.

Another important source of data is significant others, that is, those people who are in close contact with the clients—for example, spouses, relatives, or close friends. Under some circumstances their perceptions may have less danger of bias than those of the clients themselves. For example, Ellsworth (1979) developed a multiple-scale measure of the progress of mental health patients to be taken by the spouse or main relative with whom the patient lived after discharge from the hospital.

Finally, program personnel are a source of data. Because program people are often skilled professionals, their perceptions in some ways can have greater validity than the perceptions of either the clients or significant others. On the other hand, they may tend to be biased by their personal involvement in the program.

In addition to collecting data from various sources, evaluators should consider whether multiple measures from each source are necessary or desirable. For example, instead of using an instrument that measures only anxiety, it might be preferable to use an instrument that also provides a measure of depression, fatigue, or confusion (moods sometimes closely associated with anxiety). Or (for example), if evaluators are measuring the effectiveness of a program for the rehabilitation of stroke patients, it might be important to have separate measures for activities of daily living (such as walking, feeding oneself, dressing oneself, and so forth) and for speech skills. Right and left hemiplegics experience qualitatively different damage and can be expected to respond differently to speech therapy than they do to physical therapy. In the evaluation of the rehabilitation program mentioned above, Carey and Posavac (1978) used two scales to measure progress within the hospital and three additional scales to measure progress at home after discharge from the hospital.
Data collection. Who will handle the day-to-day mechanics of data collection? This will usually involve an on-site coordinator who will keep the program evaluators abreast of the whereabouts of the program clients, personnel, and other sources of data, so that they can be contacted at appropriate times by the evaluation team. This is ordinarily a thankless and tedious task, and the evaluators need a reliable person to handle it.

Data collectors need to be sensitive to the issue of confidentiality. First, confidential information must be kept confidential. If information was obtained with a previous understanding of confidentiality, it should not be released until explicit approval is obtained. Second, the obligation of others to keep information confidential must be respected. Evaluators do not have a right to all information, even when it might have a bearing on the evaluation. Other values may at times conflict with the needs of the evaluators. Care must be taken not to ask information from people in a manner they find to be an invasion of privacy.

Choice of statistics. In any type of evaluation, appropriate statistics that will demonstrate both the level of statistical significance and the magnitude of effects are needed. It is preferable to use statistical procedures that are simple, because the findings will be presented to program sponsors and personnel who often do not have a great deal of mathematical expertise. Nonevaluators should ideally be able to understand and be convinced by the interpretation of the results; they should not merely be impressed by statistical sophistication. At times multiple-regression analysis and analysis of covariance will be the statistics of choice. However, it is good to keep in mind that nonevaluators will usually be more comfortable with less-complicated statistics. One approach is to use statistical analyses that are as powerful as necessary, but to illustrate the conclusions drawn using percentages or—e even better—graphs.

Final report. Finally, some thought should be given during the planning stage to the format of the final report. Graphs and bar charts are usually preferable to tables of statistics. Where tables of statistics are necessary, they should be kept to a minimum and put in an appendix. It is advisable to rough out a few tables or charts and use some estimated figures to see whether or not the type and amount of data are appropriate for the planned layout. Often a consideration of the style of the report will suggest changes in the plans for data collection and analysis. Chapter 12 is devoted to more complete discussion of evaluation reports.

Step 6: Present a written proposal

After reviewing the literature and thinking through the various methodological considerations outlined in Step 5, the evaluators are ready to prepare a written proposal for presentation to program personnel. The overall purpose is to be certain that the evaluators and program personnel agree on the nature and goals of the program, the type of evaluation desired, the operational measures of the program goals, and the readiness of the program for evaluation. It is psychologically important for the program personnel to fully understand the evaluation process, feel comfortable with it, and (if possible) be enthusiastic about it.

Some issues that were previously discussed during the initial meetings may have to be worked through once again. Program personnel may now see procedural problems that were not at first apparent. It may be necessary to delay the beginning of data collection either because they feel the program is not sufficiently operational to undergo evaluation or because the temporary presence of extraneous factors might interfere with the interpretation of data. For example, if the proposal calls for the collection of data from hospital patients on their reactions to resident physicians who make rounds with attending physicians, it would be advisable to collect this at a time when medical students are not also making rounds with the resident and attending physicians. Otherwise, the patients might confuse the residents with medical students who have not graduated from medical school and are not providing care to patients. The problem can be handled by delaying data collection until the months when the medical students are not present on the floor.

Potential Sources of Resistance to Program Evaluation

Political and psychological factors can undermine an evaluation project. Some factors are based on genuine concerns, while others are based on misunderstandings of evaluation. Effective evaluators will identify these factors during the planning stage, bring them to the surface promptly, and resolve them gently and directly.

Expectations of a “Slam-Bang” Effect

Program personnel are generally enthusiastic and confident about the potential effects of their program. Because they are enthusiastic and put forth a maximum effort, they expect their new program to have dramatic results—a “slam-bang” effect. They will have a tendency to feel betrayed if evaluators are able to demonstrate only a moderate improvement over the old program. The difficulty can arise when the program that was replaced was already achieving its goals reasonably well, and the new program is expected to improve on the replaced program.
Fear That Evaluations Will Inhibit Innovation

Personnel in human service organizations may worry that evaluation will interfere with innovation by inhibiting them from experimenting with new techniques. In both process and outcome evaluations, the staff may feel that evaluation permits absolutely no variation in the program during the period of data collection. This is partially true insofar as there cannot be major structural changes in the program that would alter the essential goals or nature of the program. However, the need for retaining program identity does not mean that clinicians or program personnel cannot be flexible in the day-to-day operation of the program within broad structural boundaries. Every program will have variability built into it. Evaluation will not limit this. However, it is wise not to attempt to evaluate a program that is just getting started; major changes can occur as staff become clearer about their objectives.

Fear That The Program Will Be Terminated

Although it is seldom true that a negative evaluation will lead to a program's termination, it is possible that an evaluation could result in the curtailment or elimination of a program when results demonstrate that a given approach is not working out as expected. In fact, cost-effectiveness evaluation is specifically concerned with cutting losses in time, effort, and money. However, before sponsors can completely eliminate a program designed to meet a specific problem, they are ordinarily under some pressure to decide what to put in its place. Therefore, it is more likely that an evaluation will result in the reassessment of a program rather than its elimination.

Early in the planning stage, effective evaluators will try to have program personnel view them as valuable and needed associates. Evaluators can often achieve this goal by describing their work in terms of "documenting success." Program sponsors are inclined to demand accountability from program personnel as a condition for funding. Evaluators are there to assist program personnel in fulfilling this obligation to the sponsors.

It is not always possible to dispel the threat of evaluation completely. The reputation of the evaluators for trustworthiness and competence is a valuable asset. One practice that will allay some anxiety is to promise the program personnel that they will see the final draft of a report and be asked for their suggestions and clarifications before it is sent to the sponsors. Nevertheless, evaluators will not be able to eliminate anxiety completely in people who are basically insecure or who have doubts about their own competence. In a good program these persons will be in the minority.

Self-Styled Experts in Evaluation

Frequently program personnel have had some experience working on a research project. Because of this experience, they may feel attached to a specific evaluation design. For example, they may insist on testing sub-jects prior to a treatment or new program, while the evaluators feel that a treatment group and the comparison group. For example, if a group of treatments or new program, while the evaluators feel that a treatment group and the comparison group. For example, if a group of psychosomatic patients treated to a special program in a hospital are compared to psychosomatic patients treated privately by physicians, then the benefit of the special program, both groups should improve—because both groups are receiving professional treatment. Often the value of such a both groups are receiving professional treatment.
Planning an Evaluation

many persons with exactly the same diagnosis. The severity of the problem often varies widely. Individual goals differ greatly from patient to patient. Behavior considered to be progress for one patient cannot be rated as progress for another. Similar comments could be made about children in educational programs and adults in job-training programs.

There is a good deal of merit in the charge of insensitivity. Evaluators will gain the confidence of managers and staff not only by being aware of this problem but also by articulating this awareness in such a manner that program staff members are reassured that the intricacies of human services have been appropriately addressed. Early in the planning stage, the program personnel can be assured that the evaluation will not begin until they have had the opportunity to carefully review the evaluation proposal and feel comfortable that their concerns in this area have been addressed properly.

Evaluation Drains Program
Money

The seven sources of resistance described thus far are focused on various aspects of evaluation but not on the concept of evaluation itself. Two objections to evaluation strike at the very concept of program evaluation. The first objection is that program evaluation drains money that could be spent on direct service. As the statement stands, it is true. However, the main question is whether evaluation can improve service. The alternative to spending money on evaluation is to risk spending money on services that are of doubtful value. Those who present this objection fail to face the reality that the day of accountability has arrived. Today it is hard to find a program funded either by government agencies or by private foundations that does not carry the stipulation that a certain amount of money in the grant be spent on evaluation.

Those who are not convinced by the accountability argument may be convinced by a more pragmatic one: Evaluation research, if done well, may help spread a good idea and may result in attracting more money and resources to a program.

Evaluation Has Little Impact

Some critics of evaluation point out that frequently evaluation has very little impact on programs. There is a good deal of validity to this objection. Evaluators have often been frustrated by seeing their reports set aside and disregarded.

However, evaluators should reflect on the hard reality that evaluation research is not a benign social science activity but rather a political decision-making tool. If relevant, evaluation results will be included among the other factors behind decisions. Because evaluation researchers work in a

Fear That Information Will Be Abused

In addition to the fear that a program may be terminated, there may exist some concern that information gained about the performance of staff may be abused. Even competent clinicians are rightly concerned about merit reviews, future promotion, and career advancement. Past experience may have taught them to be wary of evaluators who have overstepped the boundaries of program evaluation or who have been careless about maintaining confidences. Once again, being trustworthy is critical in allaying these suspicions. The reputation of evaluators for integrity and for being responsible is the most valuable asset for defusing this suspicion. It is easy to lose trust and hard to regain it once it has been lost. Effective evaluators will not only explicitly try to convey the idea that program evaluation is distinct from individual performance appraisal, but they will carefully avoid speaking or acting in such a way that might even give the appearance that they are engaged in such an activity.

Fear That Qualitative Methods May Be Supplanted

Clinicians, teachers, probation officers, and other service personnel rightly feel that their day-to-day observations are a valuable source of input both for improving the functioning of a program and for evaluating its effect. They may feel that the evaluators' questionnaires, complicated research designs, and statistical techniques are less sensitive than are their own personal observations and evaluations. At times they are right.

However, the point to be made with program personnel is that their input is one very valuable source of evaluation data. Nevertheless, their subjective evaluations can be biased and can be strengthened by both quantitative and qualitative data gathered from other sources. Their subjective observations will be of greatest importance when the data are being interpreted. The ideal is not to eliminate either the quantitative or qualitative approaches but rather to integrate and blend the findings from both methodologies.

Accusations That Evaluation Methods Are Insensitive

Closely allied to the objection that quantitative evaluation methodology devalues subjective observations is the accusation that evaluation designs are insensitive to the intricacies of human services. Program objectives are often varied and complex. Progress is often slow and irregular. For example, a hospital program for psychiatric patients presents a very difficult evaluation project. This is true because it is very difficult to find
political context, the results of their work must be timely and relevant to decision making (Cronbach, 1982). Well-designed and carefully executed studies are valuable only when they speak to issues that are important to the organization. When evaluators fail to show how the evaluation is relevant, they have not completed their work.

In spite of the apparent validity of complaints about lack of impact, there is also some evidence that these complaints are overstated. First, evaluation studies have had a cumulative effect in certain areas. For example, the common finding that class size has marginal effects on learning may at first incline one to set aside the factor of class size as being of minimal importance. However, research done on schools in other countries has now shown that the marginal effect of class size holds only within the range of class sizes customarily found in American schools (Berk and Rossi, 1976). Therefore, it sometimes takes time before the significance of evaluation research can be put into perspective.

Second, evaluation research has some long-term value insofar as it challenges conventional wisdom. For example, those who sanctioned and conducted the New Jersey negative income tax experiment boldly tested one of the oldest, most central economic assumptions. Most economists and politicians predicted that even a limited amount of guaranteed income would lead to massive malingering and a loss of motivation to work. The study yielded mixed results, some significantly at odds with expectations. To the extent that the results of the experiment can be generalized to the national low-income population, they indicate that a national program of guaranteed income at the benefit levels considered in this experiment would have only relatively small effects on the labor supply of male family heads (Havemann and Watts, 1976). The potential of evaluation research for challenging conventional wisdom is also illustrated by the Coleman Report (Coleman et al., 1966). This evaluation showed that schools make smaller contributions to the learning of children than does family background.

An unexpected benefit of the apparently limited impact of evaluation on programs has been the shift from summative evaluation to a greater emphasis on formative evaluation. This shift has helped to increase the involvement of evaluators in the planning stages of programs. This new emphasis and orientation of program evaluation is reflected in the subject matter and style of this text. Chapter 5 develops the role of the evaluator in program planning in greater detail.

**SUMMARY**

Following the six steps in planning an evaluation project serves to get the evaluation off to a good start. Note the steps are suggested to help the evaluator to be responsive to the needs of the people most concerned about the evaluation. Responsive evaluators will have fewer problems with the resistances outlined in the second part of the chapter, compared to evaluators who seem less concerned about stakeholder needs.

**STUDY QUESTIONS**

1. As an outside consultant you are invited by a large manufacturing plant to evaluate a new training program for managers, called “Management Contact,” which you are told applies the principles of transactional analysis to employer-employee relations. Explain the steps you would take and what you would need to know in approaching this job. What would be the major difficulties you would see?

2. You are part of an evaluation team retained by a major university. The chairman of the department of psychology instructs you to evaluate a new section to be established called “Community Psychology.” How would you approach this task? What would be the major pitfalls to avoid?

**FURTHER READING**