6

Assessing and Monitoring Program Process

Chapter Outline

To be effective in bringing about the desired improvements in social conditions, a program needs more than a good plan of attack. Most important, the program must implement its plan; that is, it must actually carry out its intended functions in the intended way.

Although implementing a program concept may seem straightforward, in practice it is often very difficult. Social programs typically must contend with many adverse influences that can compromise even well-intentioned attempts to conduct program business appropriately. The result can easily be substantial discrepancies between the program as intended and the program as actually implemented.

The implementation of a program is reflected in concrete form in the program processes that it puts in place. An important evaluation function, therefore, is to assess the adequacy of program process: the program activities that actually take place and the services that are actually delivered in routine program operation. This chapter introduces the procedures evaluators use to investigate these issues.

After signing a new bill, President John F. Kennedy is reputed to have said to his aides, "Now that this bill is the law of the land, let's hope we can get our government to carry it out." Both those in high places and those on the front lines are often justified in being skeptical about the chances that a social program will be appropriately implemented. Many steps are required to take a program from concept to full operation, and much effort is needed to keep it true to its original design and purposes. Thus, whether any program is fully carried out as envisioned by its sponsors and managers is always problematic.

Ascertaining how well a program is operating, therefore, is an important and useful form of evaluation, known as *program process evaluation.* (A widely used alternative label is *implementation evaluation.*) It does not represent a single distinct evaluation procedure but, rather, a family of approaches, concepts, and methods. The defining theme of program process evaluation (or simply process evaluation) is a focus on the enacted program itself—its operations, activities, functions, performance, component parts, resources, and so forth. When process evaluation involves an ongoing effort to measure and record information about the program's operation, we will refer to it as *program process monitoring.*

6.1 What Is Program Process Evaluation and Monitoring?

As was suggested in **Chapter 2 (ch02.html)**, evaluators often distinguish between process (or implementation) evaluation and outcome (or impact) evaluation. Process evaluation, in Scheirer's (1994) words, "verifies what the program is and whether or not it is delivered as intended to the targeted recipients. "It does not, however, attempt to assess the effects of the program on those recipients. Such assessment is the province of impact evaluation, which we consider in later chapters.

Where process evaluation is an ongoing function involving repeated measurements over time, it is referred to as program monitoring. Corresponding to the distinction between process and outcome evaluation, **program process monitoring** is the systematic and continual documentation of key aspects of program performance that assesses whether the program is operating as intended or according to some appropriate standard, whereas **outcome monitoring** is the continual measurement of intended outcomes of the program, usually of the social conditions it is intended to improve. We discuss outcome monitoring in conjunction with impact evaluations later in this book.

Program process evaluation generally involves assessments of program performance in the domains of service utilization and program organization. Assessing service utilization consists of examining the extent to which the intended target population receives the intended services. Assessing program organization requires comparing the plan for what the program should be doing with what is actually done, especially with regard to providing services. Usually, program process evaluation is directed at one or both of two key questions: (1) whether a program is reaching the appropriate target population and (2) whether its service delivery and support functions are consistent with program design specifications or other appropriate standards. Process evaluation may also examine what resources are being or have been expended in the conduct of the program.

More specifically, program process evaluation schemes are designed to answer such evaluation questions as these:

- How many persons are receiving services?
- Are those receiving services the intended targets?
- Are they receiving the proper amount, type, and quality of services?
- Are there targets who are not receiving services or subgroups within the target population who are underrepresented among those receiving services?
- Are members of the target population aware of the program?
- Are necessary program functions being performed adequately?
- Is staffing sufficient in numbers and competencies for the functions that must be performed?
- Is the program well organized? Do staff work well with each other?
- Does the program coordinate effectively with the other programs and agencies with which it must interact?
- Are resources, facilities, and funding adequate to support important program functions?
- Are resources used effectively and efficiently?
- Is the program in compliance with requirements imposed by its governing board, funding agencies, and higher-level administration?
- Is the program in compliance with applicable professional and legal standards?
- Is performance at some program sites or locales significantly better or poorer than at others?
- Are participants satisfied with their interactions with program personnel and procedures?
- Are participants satisfied with the services they receive?
- Do participants engage in appropriate follow-up behavior after service?

Setting Criteria for Judging Program Process

It is important to recognize the evaluative themes in process evaluation questions such as those listed above. Virtually all involve words such as *appropriate, adequate, sufficient, satisfactory, reasonable, intended,* and other phrasing indicating that an evaluative judgment is required. To answer these questions, therefore, the evaluator or other responsible parties must not only describe the program's performance but also assess whether it is satisfactory. This, in turn, requires that there be some bases for making judgments, that is, some defensible criteria or standards to apply. Where such criteria are not already articulated and endorsed, the evaluator may find that establishing workable criteria is as difficult as determining program performance on the pertinent dimensions.

There are several approaches to the matter of setting criteria for program performance. Moreover, different approaches will likely apply to different dimensions of program performance because the considerations that go into defining, say, what constitutes an appropriate number of clients served are quite different from those pertinent to deciding what constitutes an adequate level of resources. This said, the approach to the criterion issue that has the broadest scope and most general utility in program process evaluation is the application of program theory as described in **Chapter 5 (ch05.html)**.

Recall that program theory, as we presented it, is divided into program process theory and program impact theory. Program process theory is formulated to describe the program as intended in a form that virtually constitutes a plan or blueprint for what the program is expected to do and how. As such, it is particularly relevant to program process evaluation. Recall also that program theory builds on needs assessment (whether systematic or informal) and thus connects the program design with the social conditions the program is intended to ameliorate. And, of course, the process through which theory is derived and adopted usually involves input from the major stakeholders and, ultimately, their endorsement. Program theory thus has a certain authority in delineating what a program "should" be doing and, correspondingly, what constitutes adequate performance.

Program process evaluation, therefore, can be built on the scaffolding of program process theory. Process theory identifies the aspects of program performance that are most important to describe and also provides some indication of what level of performance is intended, thereby providing the basis for assessing whether actual performance measures up. **Exhibit 5-E (http://content.thuzelearning.com/books/Rossi.1553.18.1/sections/nav_34#exhibit5-E)** in the previous chapter, for instance, illustrates the service utilization component of the program process theory for an aftercare program for released psychiatric patients. This flowchart depicts, step by step, the interactions and experiences patients released from the hospital are supposed to have as a result of program service. A thorough monitoring procedure would systematically document what actually happened at each step. In particular, it would report how many patients were released from the hospital each month, what proportion were visited by a social worker, how many were referred to services and which services, how many actually received those services, and so forth.

If the program processes that are supposed to happen do not happen, then we would judge the program's performance to be poor. In actuality, of course, the situation is rarely so simple. Most often, critical events will not occur in an all-or-none fashion but will be attained to some higher or lower degree. Thus, some, but not all, of the released patients will receive visits from social workers, some will be referred to services, and so forth. Moreover, there may be important quality dimensions. For instance, it would not represent good program performance if a released patient was referred to several community services, but these services were inappropriate to the patient's needs. To determine how much must be done, or how well, we need additional criteria that parallel the information the monitoring procedure provides. If the monitoring procedure reports that 63% of the released patients are visited by a social worker within two weeks of release, we cannot evaluate that performance without some standard that tells us what percentage is "good." Is 63% a poor performance, given that we might expect 100% to be desirable, or is it a very impressive performance with a clientele that is difficult to locate and serve?

The most common and widely applicable criteria for such situations are simply **administrative standards** or objectives, that is, stipulated achievement levels set by program administrators or other responsible parties. For example, the director and staff of a job training program may commit to attaining 80% completion rates for the training or to having 60% of the participants permanently employed six months after receiving training. For the aftercare program above, the administrative target might be to have 75% of the patients visited within two weeks of release from the hospital. By this standard, the 63% found with program monitoring is a subpar performance that, nonetheless, is not too far below the mark.

Administrative standards and objectives for program process performance may be set on the basis of past experience, the performance of comparable programs, or simply the professional judgment of program managers or advisers. If they are reasonably justified, they can provide meaningful standards against which to assess observed program performance. In a related vein, some aspects of program performance may fall under applicable legal, ethical, or professional standards. The "standards of care" adopted in medical practice for treating common ailments, for instance, provide a set of criteria against which to assess program performance in health care settings. Similarly, state children's protective services almost always have legal requirements to meet concerning handling cases of possible child abuse or neglect.

In practice, the assessment of particular dimensions of program process performance is often not based on specific, predetermined criteria but represents an after-the-fact judgment call. This is the "I'll know it when I see it" school of thought on what constitutes good program performance. An evaluator who collects process data on, say, the proportion of high-risk adolescents who recall seeing program-sponsored antidrug media messages may find program staff and other key stakeholders resistant to stating what an acceptable proportion would be. If the results come in at 50%, however, a consensus may arise that this is rather good considering the nature of the population, even though some stakeholders might have reported much higher expectations prior to seeing the data. Other findings, such as 40% or 60%, might also be considered rather good. Only extreme findings, say 10%, might strike all stakeholders as distressingly low. In short, without specific before-measurement criteria a wide range of performance might be regarded as acceptable. Of course, assessment procedures that are too flexible and that lead to a "pass" for all tend to be useless.

Very similar considerations apply to the organizational component of the process theory. A depiction of the organizational plan for the aftercare program was presented in **Exhibit 5-F (http://content.thuzelearning.com/books/Rossi.1553.18.1/sections/nav_34#exhibit5-F)** in **Chapter 5 (ch05.html)** . Looking back at it will reveal that it, too, identifies dimensions of program performance that can be described and assessed against appropriate standards. Under that plan, for instance, case managers are expected to interview clients and families, assess service needs, make referrals to services, and so forth. A program process evaluation would document and assess what was done under each of those categories.

Common Forms of Program Process Evaluations

Description and assessment of program process are quite common in program evaluation, but the approaches used are varied, as is the terminology they employ. Such assessments may be conducted as a one-shot endeavor or may be continuous so that information is produced regularly over an extended period of time, as in program process monitoring. They may be conducted by "outside" or "inside" evaluators or be set up as management tools with little involvement by professional evaluators. Moreover, their purpose may be to provide feedback for managerial purposes, to demonstrate accountability to sponsors and decisionmakers, to provide a freestanding process evaluation, or to augment an impact evaluation. Amid this variety, we distinguish two principal forms of program process studies, process or implementation evaluation and continuous program monitoring.

Process or Implementation Evaluation

Process or implementation evaluation is typically conducted by evaluation specialists as a separate project that may involve program personnel but is not integrated into their daily routine. When completed and, often, while under way, process evaluation generally provides information about program performance to program managers and other stakeholders, but is not a regular and continuing part of a program's operation. **Exhibit 6-A (http://content.thuzelearning.com/books/Rossi.1553.18.1/sections/nav_40#exhibit6-A)** describes a process evaluation of an integrated services program for children.

As an evaluation approach, process evaluation plays two major roles. First, it can stand alone as an evaluation of a program in circumstances where the only questions at issue are about the integrity of program operations, service delivery, and other such matters. There are several kinds of situations that fit this description. A stand-alone process evaluation might be appropriate for a relatively new program, for instance, to answer questions about how well it has established its intended operations and services. Program process is often the focus of formative evaluation designed to provide useful feedback to managers and sponsors of new programs. In the case of a more established program, a process evaluation might be called for when questions arise about how well the program is organized, the quality of its services, or the success with which it is reaching the target population. A process evaluation may also constitute the major evaluation approach to a program charged with delivering a service known or presumed effective, so that the most significant performance issue is whether that service is being delivered properly. In a managed care environment, for instance, process evaluation may be employed to assess whether the prescribed medical treatment protocols are being followed for patients in different diagnostic categories.

EXHIBIT 6-A

Process Evaluation to Assess Integrated Services for Children

Many analysts have observed that the traditional system of categorical funding for children's services, with funds allocated to respond to specific problems under strict rules regarding eligibility and expenditures, has not served children's needs well. The critics argue that this system fragments services and inhibits collaboration between programs that might otherwise lead to more effective services. In 1991, the Robert Wood Johnson Foundation launched the Child Health Initiative to test the feasibility of achieving systemic changes through the integration of children's services and finances. Specifically, the initiative called for the development of the following components:

- A decategorization mechanism that would pool existing categorical program funds and create a single children's health fund
- A care coordination procedure using case management that would use the pooled funds to provide comprehensive and continuous care for needy children
- A monitoring system that would identify the health and related needs of children in the community and the gaps in existing services

Nine sites across the country were selected to launch demonstration programs. The Institute for Health Policy Studies, University of California, San Francisco, conducted an evaluation of these programs with two major goals: (1) to gauge the degree to which the implementation of the projects was consistent with the original planning objectives (fidelity to the model) and (2) to assess the extent to which each of the major program components was implemented. In the first year, the evaluation focused on the political, organizational, and design phase of program development. During subsequent years, the focus turned to implementation and preliminary outcomes. A combination of methods was used, including site visits, written surveys completed by the program managers, in-depth interviews of key participants, focus groups of service providers and clients, and reviews of project-related documents.

The evaluation found that most of the nine sites experienced some degree of success in implementing the monitoring and care coordination components, but none was able to implement decategorization. The general findings for each component were as follows:

- Decategorization: Several sites successfully created small pools of flexible funds, but these were from sources other than categorical program funds. No site was able to fully implement decategorization under the definitions originally adopted.
- Care coordination: This was implemented successfully by most of the sites at the client level through case management, but there was generally less coordination at the system level.
- Monitoring: The sites encountered a number of barriers in successfully completing this task, but most instituted some appropriate process.

SOURCE: Adapted from Claire Brindis, Dana C. Hughes, Neal Halfon, and Paul W. Newacheck, "The Use of Formative Evaluation to Assess Integrated Services for Children," *Evaluation & the Health Professions,* 1998, 21(1):66-90.

The second major role of process or implementation evaluation is as a complement to an impact evaluation. Indeed, it is generally not advisable to conduct an impact evaluation without including at least a minimal process evaluation. Because maintaining an operational program and delivering appropriate services on an ongoing basis are formidable challenges, it is not generally wise to take program implementation for granted. A full impact evaluation, therefore, includes a process component to determine what quality and quantity of services the program provides so that this information can be integrated with findings on what impact those services have.

Continuous Program Process Evaluation (Monitoring) and Management Information Systems

The second broad form of program process evaluation consists of continuous monitoring of indicators of selected aspects of program process. Such process monitoring can be a useful tool for facilitating effective management of social programs by providing regular feedback about how well the program is performing its critical functions. This type of feedback allows managers to take corrective action when problems arise and can also provide stakeholders with regular assessments of program performance. For these reasons, a form of process assessment is often integrated into the routine information systems of social programs so that appropriate data are obtained, compiled, and periodically summarized. In such cases, process evaluation becomes coextensive with the **management information system (MIS)** in a human service program. **Exhibit 6-B (http://content.thuzelearning.com/books/Rossi.1553.18.1/sections/nav_40#exhibit6-B)** describes an MIS that was developed for a marital and family counseling program.

MISs routinely provide information on a client-by-client basis about services provided, staff providing the services, diagnosis or reasons for program participation, sociodemographic data, treatments and their costs, outcome status, and so on. Some of the systems bill clients (or funders), issue payments for services, and store other information, such as a client's treatment history and current participation in other programs. MISs have become the major data source in many instances because much of the information that otherwise would have to be gathered in data collection for process monitoring is available in the program's MIS. Even when a program's MIS is not configured to completely fulfill the requirements of a thoroughgoing process evaluation, it may nonetheless provide a large portion of the information an evaluator needs for such purposes. MISs can thus supply data that can be used by both managers and evaluators.

EXHIBIT 6-B

An Integrated Information System for a Family and Marriage Counseling Agency in Israel

The Marital and Family Counselling Agency is run under the joint auspices of the Tel Aviv Welfare Department and the School of Social Work at Tel Aviv University. The agency provides marital and family counseling and community services for the Jewish, Muslim, and Christian residents of one of the poorest sections of Tel Aviv.

The integrated information system developed for the agency is designed to follow up clients from the moment they request help to the end of treatment. It is intended to serve the agency and the individual counselors by monitoring the process and outcomes of treatment and providing the data needed to make organizational and clinical decisions. To accomplish this, data are collected on three forms and then programmed into the computerized information system. The data elements include the following:

- Background data provided by the client, for example, sociodemographic characteristics, medical and psychological treatment history, the problems for which they are seeking help, the urgency of those problems, their expectations from treatment, and how they found out about the clinic.
- The McMaster Clinical Rating Scale, a standardized scale that monitors families on the basis of six dimensions of family functioning and overall family health; the counselors fill out this form once a month for each client.
- Retrospective evaluation forms filled out after treatment is completed, one by the counselors and another by the clients. This includes, for example, factual questions about the treatment such as its duration, the problems dealt with, the degree to which the client and counselor agreed on the problems, whether there were issues not addressed and why. It also includes retrospective assessments of the process, evaluations of improvement in the presented problems and the McMaster areas of functioning, and client and counselor satisfaction with the process and outcomes.

The counselors can enter and retrieve data from this system whenever they wish and are given a graph of each client's status every three months to support clinical decisions. Also, reports are generated for the clinic's management. For example, a report of the distribution of clients by ethnic group led to the development of a program located within Arab community centers to better reach that population. Other management reports describe the ways and times at which treatment is terminated, the problems that brought clients to the agency, and the percentage of people who applied for treatment but did not show up for the first session. The information system has also been used for research purposes. For example, studies were conducted on the predictors of treatment success, the comparative perceptions by clients and counselors of the treatment process and outcomes, and gender differences in presenting problems.

SOURCE: Adapted from Rivka Savaya, "The Potential and Utilization of an Integrated Information System at a Family and Marriage Counselling Agency in Israel," *Evaluation and Program Planning,* 1998, 21(1):11-20.

6.2 Perspectives on Program Process Monitoring

There is and should be considerable overlap in the purposes of process monitoring whether they are driven by the information needs of evaluators, program managers and staff, or policymakers, sponsors, and stakeholders. Ideally, the monitoring activities undertaken as part of evaluation should meet the information needs of all these groups. In practice, however, limitations on time and resources may require giving priority to one set of information needs over another. Although there are many exceptions, the perspectives of the three key "consumer groups" on the purposes of program monitoring typically vary. These differences in perspective apply generally to outcome monitoring as well.

Process Monitoring From the Evaluator's Perspective

A number of practical considerations underlie the need for evaluation researchers to monitor program process. All too often a program's impact is sharply diminished and, indeed, sometimes reduced to zero because the appropriate intervention was not delivered, was not delivered to the right targets, or both. We believe that more program failures are due to such implementation problems than to lack of potentially effective services. Process monitoring studies, therefore, are essential to understanding and interpreting impact findings. Knowing what took place is a prerequisite for explaining or hypothesizing why a program did or did not work. Without process monitoring, the evaluator is engaged in "black box" research with no basis for deciding whether a larger dose of the program or a different means of delivering the intervention would have changed the impact results.

Process Monitoring From an Accountability Perspective

Process monitoring information is also critical for those who sponsor and fund programs. Program managers have a responsibility to inform their sponsors and funders of the activities undertaken, the degree of implementation of programs, the problems encountered, and what the future holds (see **Exhibit 6-C (http://content.thuzelearning.com/books/Rossi.1553.18.1/sections/nav_41#exhibit6-C)** for one perspective on this matter). However, evaluators frequently are mandated to provide the same or similar information. Indeed, in some cases the sponsors and funders of programs perceive program evaluators as "their eyes and ears," as a second line of information on what is going on in a particular program.

EXHIBIT 6-C
Program and Service Utilization Studies

> Any service organization, especially in an era of shrinking resources, needs to evaluate its services and activities. Through these evaluative activities, an organization can develop and maintain the flexibility needed to respond to an ever-changing environment. It has been suggested that, even in an ideal world, an organization needs to be self-evaluating. Self-evaluation requires an organization to continually review its own activities and goals and to use the results to modify, if necessary, its programs, goals, and directions.
>
> Within an agency, the essential function of evaluation is to provide data on goal achievement and program effectiveness to a primary audience consisting of administration, middle management, and the governing board. This primary audience, especially the administration and board, is frequently confronted with inquiries from important sources in the external environment, such as legislators and funding agencies. These inquiries often focus on issues of client utilization, accessibility, continuity, comprehension, outcome or effectiveness, and cost. The building block of this information is the patterns of use or client utilization study. The patterns of use study, whether it consists of simple inquiries or highly detailed, sophisticated investigations, is basically a description. It describes who uses services and how. It becomes evaluative when it is related to the requirements or purposes of the organization.

SOURCE: Adapted from G. Landsberg, "Program Utilization and Service Utilization Studies: A Key Tool for Evaluation," *New Directions for Program Evaluation*, no. 20 (San Francisco: Jossey-Bass, December 1983), pp. 93-103.

Government sponsors and funding groups, including Congress, operate in the glare of the mass media. Their actions are also visible to the legislative groups who authorize programs and to government "watchdog" organizations. For example, at the federal level, the Office of Management and Budget, part of the executive branch, wields considerable authority over program development, funding, and expenditures. The U.S. General Accounting Office, an arm of Congress, advises members of the House and Senate on the utility of programs and in some cases conducts evaluations. Both state governments and those of large cities have analogous oversight groups. No social program that receives outside funding, whether public or private, can expect to avoid scrutiny and escape demand for **accountability.**

In addition to funders and sponsors, other stakeholders may press for program accountability. In the face of taxpayers' reservations about spending for social programs, together with the increased competition for resources often resulting from cuts in available funding, all stakeholders are scrutinizing both the programs they support and those they do not. Concerned parties use process monitoring information to lobby for the expansion of programs they advocate or find congenial with their self-interests and the curtailment or abandonment of those programs they disdain. Stakeholders, it should be noted, include the targets themselves. A dramatic illustration of their perspective occurred when President Ronald Reagan telephoned an artificial heart recipient patient to wish him well and, with all of the country listening, the patient complained about not receiving his Social Security check.

Clearly, social programs operate in a political world. It could hardly be otherwise, given the stakes involved. The human and social service industry is not only huge in dollar volume and number of persons employed but is also laden with ideological and emotional baggage. Programs are often supported or opposed by armies of vocal community members; indeed, the social program sector is comparable only to the defense industry in its lobbying efforts, and the stands that politicians take with respect to particular programs often determine their fates in elections. Accountability information is a major weapon that stakeholders use in their battles as advocates and antagonists.

Process Monitoring From a Management Perspective

Management-oriented process monitoring (including use of MISs) often is concerned with the same questions as program process or accountability studies; the differences lie in the purposes to which the findings are to be put. Evaluators' interest in monitoring data generally centers on determining how a program's potential impact is related to its implementation. Accountability studies primarily provide information that decisionmakers, sponsors, and other stakeholders need to judge the appropriateness of program activities and to decide whether a program should be continued, expanded, or contracted. Such studies may use the same information base employed by program management staff, but they are usually conducted in a critical spirit. In contrast,

management-oriented monitoring activities are concerned less with making decisive judgments and more with incorporating corrective measures as a regular part of program operations.

Process monitoring from a management perspective is particularly vital during the implementation and pilot testing of new programs, especially innovative ones. No matter how well planned such programs may be, unexpected results and unwanted side effects often surface early in the course of implementation. Program designers and managers need to know rapidly and fully about these problems so that changes can be made as soon as possible in the program design. Suppose, for example, that a medical clinic intended to help working mothers is open only during daylight hours. Monitoring may disclose that, however great the demand is for clinic services, the clinic's hours of operation effectively screen out most of the target population. Or suppose that a program is predicated on the assumption that severe psychological problems are prevalent among children who act out in school. If it is found early on that most such children do not in fact have serious disorders, the program can be modified accordingly.

For programs that have moved beyond the development stage to actual operation, program process monitoring serves management needs by providing information on process and coverage (the extent to which a program reaches its intended targets), and hence feedback on whether the program is meeting specifications. Fine-tuning of the program may be necessary when monitoring information indicates that targets are not being reached, that the implementation of the program costs more than initially projected, or that staff workloads are either too heavy or too light. Managers who neglect to monitor a program fully and systematically risk the danger of administering a program that is markedly different from its mandate.

Where monitoring information is to be used for both managerial and evaluation purposes, some problems must be anticipated. How much information is sensible to collect and report, in what forms, at what frequency, with what reliability, and with what degree of confidentiality are among the major issues on which evaluators and managers may disagree. For example, the experienced manager of a nonprofit children's recreational program may feel that the highest priority is weekly information on attendance. The evaluator, however, may be comfortable with aggregating the data monthly or even quarterly, but may believe that before being reported they should be adjusted to take into account variations in the weather, occurrence of holidays, and so on—even though the necessary adjustments require the use of sophisticated statistical procedures.

A second concern is the matter of proprietary claims on the data. For the manager, monitoring data on, say, the results of a program innovation should be kept confidential until discussed with the research committee of the board of directors and presented at the board meeting. The evaluator may wish immediately to write a paper for publication in the *American Journal of Evaluation.* Or a serious drop in clients from a particular ethnic group may result in the administrator of a program immediately replacing the director of professional services, whereas the evaluator's reaction may be to do a study to determine why the drop occurred. As with all relations between program staff and evaluators in general, negotiation of these matters is essential.

A warning: There are many aspects of program management and administration (such as complying with tax regulations and employment laws or negotiating union contracts) that few evaluators have any special competence to assess. In fact, evaluators trained in social science disciplines and (especially) those primarily involved in academic careers may be unqualified to manage anything. It is wise to keep in mind that the evaluator's role, even when sharing information from an MIS, is not to join the administrators in the running of the organization.

In the remainder of this chapter, we concentrate on the concepts and methods pertinent to monitoring program process in the domains of service utilization and program organization. It is in this area that the competencies of persons trained in social research are most relevant.

6.3 Monitoring Service Utilization

A critical issue in program process monitoring is ascertaining the extent to which the intended targets actually receive program services. Target participation concerns both program managers and sponsors. Managing a project effectively requires that target participation be kept at an acceptable level and corrective action be taken if it falls below that level.

Monitoring of service utilization is particularly critical for interventions in which program participation is voluntary or in which participants must learn new procedures, change their habits, or take instruction. For example, community mental health centers designed to provide a broad range of services often fail to attract a significant proportion of those who could benefit from their services. Even homeless patients who had been recently discharged from psychiatric hospitals and encouraged to make use of the services of community mental health centers often failed to contact the centers (Rossi, Fisher, and Willis, 1986). Similarly, a program designed to provide information to prospective home buyers might find that few persons seek the services offered. Hence, program developers need to be concerned with how best to motivate potential targets to seek out the program and participate in it. Depending on the particular case, they might, for example, need to build outreach efforts into the program or pay special attention to the geographic placement of program sites (Boruch, Dennis, and Carter-Greer, 1988).

Coverage and Bias

Service utilization issues typically break down into questions about coverage and bias. Whereas **coverage** refers to the extent to which participation by the target population achieves the levels specified in the program design, **bias** is the degree to which some subgroups participate in greater proportions than others. Clearly, coverage and bias are related. A program that reaches all projected participants and no others is obviously not biased in its coverage. But because few social programs ever achieve total coverage, bias is typically an issue.

Bias can arise out of self-selection; that is, some subgroups may voluntarily participate more frequently than others. It can also derive from program actions. For instance, a program's personnel may react favorably to some clients while rejecting or discouraging others. One temptation commonly faced by programs is to select the most "success prone" targets. Such "creaming" frequently occurs because of the self-interests of one or more stakeholders (a dramatic example is described in **6-D (http://content.thuzelearning.com/books/Rossi.1553.18.1/sections/nav_42#exhibit6-D)**). Finally, bias may result from such unforeseen influences as the location of a program office, which may encourage greater participation by a subgroup that enjoys more convenient access to program activities.

Although there are many social programs, such as food stamps, that aspire to serve all or a very large proportion of a defined target population, typically programs do not have the resources to provide services to more than a portion of potential targets. In the latter case, the target definition established during the planning and development of the program frequently is not specific enough. Program staff and sponsors can correct this problem by defining the characteristics of the target population more sharply and by using resources more effectively. For example, establishing a health center to provide medical services to persons in a defined community who do not have regular sources of care may result in such an overwhelming demand that many of those who want services cannot be accommodated. The solution might be to add eligibility criteria that weight such factors as severity of the health problem, family size, age, and income to reduce the size of the target population to manageable proportions while still serving the neediest persons. In some programs, such as WIC (Women's, Infants and Children Nutrition Program) or housing vouchers for the poor, undercoverage is a systemic problem; Congress has never provided sufficient funding to cover all who were eligible, perhaps hoping that budgets could be expanded in the future.

The opposite effect, overcoverage, also occurs. For instance, the TV program *Sesame Street* has consistently captured audiences far exceeding the intended targets (disadvantaged preschoolers), including children who are not at all disadvantaged and even adults. Because these additional audiences are reached at no additional cost, this overcoverage is not a financial drain. It does, however, thwart one of *Sesame Street's* original goals, which was to lessen the gap in learning between advantaged and disadvantaged children.

In other instances, overcoverage can be costly and problematic. Bilingual programs in schools, for instance, have often been found to include many students whose primary language is English. Some school systems whose funding from the program depends on the number of children enrolled in bilingual classes have inflated attendance figures by registering inappropriate students. In other cases, schools have used assignment to bilingual instruction as a means of ridding classes of "problem children," thus saturating bilingual classes with disciplinary cases.

EXHIBIT 6-D

"Creaming" the Unemployed

When administrators who provide public services choose to provide a disproportionate share of program benefits to the most advantaged segment of the population they serve, they provide grist for the mill of service utilization research. The U.S. Employment Service (USES) offers a clear and significant example of creaming, a practice that has survived half a century of USES expansion, contraction, and reorganization. The USES has as its major aim to provide employers with workers, downplaying the purpose of providing workers with work. This leads the USES to send out the best prospects among the unemployed and to slight the less promising.

It is hardly surprising that USES administrators, a generation after the establishment of the program, stressed the necessity rather than the desirability of an employer-centered service. Its success, by design, depended on serving employers, not the "hard-core" unemployed. As President Lyndon Johnson's task force on urban employment problems noted some two weeks before the 1965 Watts riots, "We have yet to make any significant progress in reaching and helping the truly 'hard-core' disadvantaged."

SOURCE: Adapted from David B. Robertson, "Program Implementation Versus Program Design," *Policy Study Review,* 1984, 3:391-405.

The most common coverage problem in social interventions, however, is the failure to achieve high target participation, either because of bias in the way targets are recruited or retained or because potential clients are unaware of the program, are unable to use it, or reject it. For example, in most employment training programs only small minorities of those eligible by reason of unemployment ever attempt to participate. Similar situations occur in mental health, substance abuse, and numerous other programs (see **Exhibit 6-E (http://content.thuzelearning.com/books/Rossi.1553.18.1/sections/nav_42#exhibit6-E)**).We turn now to the question of how program coverage and bias might be measured as a part of program process monitoring.

Measuring and Monitoring Coverage

Program managers and sponsors alike need to be concerned with both undercoverage and overcoverage. Undercoverage is measured by the proportion of the targets in need of a program that actually participates in it. Overcoverage is often expressed as the number of program participants who are not in need, compared with the total number of participants in the program. Efficient use of program resources requires both maximizing the number served who are in need and minimizing the number served who are not in need.

Exhibit 6-E

The Coverage of the Food Stamp Program for the Homeless

Based upon a rigorously designed survey of homeless persons sampled from shelters and food kitchens in American cities with a population of 100,000 and over, Burt and Cohen gave some precise dimensions to what we know is true virtually by definition: The homeless live on food intakes that are inadequate both in quantity and in nutritional content. There is no way that a demographic group whose incomes hover slightly above zero can have adequate diets. That the homeless do not starve is largely a tribute to the food kitchens and shelters that provide them with meals at no cost.

Because most homeless persons are eligible by income for food stamps, their participation rates in that program should be high. But they are not: Burt and Cohen reported that only 18% of the persons sampled were receiving food stamps and almost half had never used them. This is largely because certification for food stamps requires passing a means test, a procedure that requires some documentation. This is not easy for many homeless, who may not have the required documents, an address to receive the stamps, or the capability to fill out the forms.

Moreover, the food stamp program is based on implicit assumptions that participants can readily acquire their foodstuffs in a local food store, prepare servings on a stove, and store food supplies in their dwellings. These assumptions do not apply to the homeless. Of course, food stores do sell some food items that can be consumed without preparation and, with some ingenuity, a full meal of such foods can be assembled. So some benefit can be obtained by the homeless from food stamps, but for most homeless persons food stamps are relatively useless.

Legislation passed in 1986 allowed homeless persons to exchange food stamps for meals offered by nonprofit organizations and made shelter residents in places where meals were served eligible for food stamps. By surveying food providers, shelters, and food kitchens, however, Burt and Cohen found that few meal providers had applied for certification as receivers of food stamps. Of the roughly 3,000 food providers in the sample, only 40 had become authorized.

Furthermore, among those authorized to receive food stamps, the majority had never started to collect food stamps or had started and then abandoned the practice. It made little sense to collect food stamps as payment for meals that otherwise were provided free so that, on the same food lines, food stamp participants were asked to pay for their food with stamps while nonparticipants paid nothing. The only food provider who was able to use the system was one that required either cash payment or labor for meals; for this program, food stamps became a substitute for these payments.

SOURCE: Based on Martha Burt and Barbara Cohen, *Feeding the Homeless: Does the Prepared Meals Provision Help?* Report to Congress on the Prepared Meal Provision, vols. I and II (Washington, DC: Urban Institute, 1988). Reprinted with permission.

The problem in measuring coverage is almost always the inability to specify the number in need, that is, the magnitude of the target population. The needs assessment procedures described in **Chapter 4 (ch04.html)** ,if carried out as an integral part of program planning, usually minimize this problem. In addition, three sources of information can be used to assess the extent to which a program is serving the appropriate target population: program records, surveys of program participants, and community surveys.

Program Records

Almost all programs keep records on targets served. Data from well-maintained record systems—particularly from MISs—can often be used to estimate program bias or overcoverage. For instance, information on the various screening criteria for program intake may be tabulated to determine whether the units served are the ones specified in the program's design. Suppose the targets of a family planning program are women less than 50 years of age who have been residents of the community for at least six months and who have two or more children under age ten. Records of program participants can be examined to see whether the women actually served are within the eligibility limits and the degree to which particular age or parity groups are under- or overrepresented. Such an analysis might also disclose bias in program participation in terms of the eligibility characteristics or combinations of them. Another example, involving public shelter utilization by the homeless, is described in **6-F (http://content.thuzelearning.com/books/Rossi.1553.18.1/sections/nav_42#exhibit6-F)** .

However, programs differ widely in the quality and extensiveness of their records and in the sophistication involved in storing and maintaining them. Moreover, the feasibility of maintaining complete, ongoing record systems for all program participants varies with the nature of the intervention and the available resources. In the case of medical and mental health systems, for example, sophisticated, computerized management and client information systems have been developed for managed care purposes that would be impractical for many other types of programs.

In measuring target participation, the main concerns are that the data are accurate and reliable. It should be noted that all record systems are subject to some degree of error. Some records will contain incorrect or outdated information, and others will be incomplete. The extent to which unreliable records can be used for decision making depends on the kind and degree of their unreliability and the nature of the decisions in question. Clearly, critical decisions involving significant outcomes require better records than do less weighty decisions. Whereas a decision on whether to continue a project should not be made on the basis of data derived from partly unreliable records, data from the same records may suffice for a decision to change an administrative procedure.

Exhibit 6-F

Public Shelter Utilization Among Homeless Adults in New York and Philadelphia

The cities of Philadelphia and New York have standardized admission procedures for persons requesting services from city-funded or -operated shelters. All persons admitted to the public shelter system must provide intake information for a computerized registry that includes the client's name, race, date of birth, and gender, and must also be assessed for substance abuse and mental health problems, medical conditions, and disabilities. A service utilization study conducted by researchers from the University of Pennsylvania analyzed data from this registry for New York City for 1987-1994 (110,604 men and 26,053 women) and Philadelphia for 1991-1994 (12,843 men and 3,592 women).

They found three predominant types of users: (1) the chronically homeless, characterized by very few shelter episodes, but episodes that might last as long as several years; (2) the episodically homeless, characterized by multiple, increasingly shorter stays over a long period; and (3) the transitionally homeless, who had one or two stays of short duration within a relatively brief period of time.

The most notable finding was the size and relative resource consumption of the chronically homeless. In New York, for instance, 18% of the shelter users stayed 180 days or more in their first year, consuming 53% of the total number of system days for first-time shelter users, triple the days for their proportionate representation in the shelter population. These long-stay users tended to be older people and to have mental health, substance abuse, and, in some cases, medical problems.

SOURCE: Adapted by permission from Dennis P. Culhane and Randall Kuhn, "Patterns and Determinants of Public Shelter Utilization Among Homeless Adults in New York City and Philadelphia," *Journal of Policy Analysis and Management,* 1998, 17(1):23-43. Copyright © 1998, John Wiley & Sons, Inc.

If program records are to serve an important role in decision making on far-reaching issues, it is usually desirable to conduct regular audits of the records. Such audits are similar in intent to those that outside accountants conduct on fiscal records. For example, records might be sampled to determine whether each target has a record, whether records are complete, and whether rules for completing them have been followed.

Surveys

An alternative to using program records to assess target participation is to conduct special surveys of program participants. Sample surveys may be desirable when the required data cannot be obtained as a routine part of program activities or when the size of the target group is large and it is more economical and efficient to undertake a sample survey than to obtain data on all the participants.

For example, a special tutoring project conducted primarily by parents may be set up in only a few schools in a community. Children in all schools may be referred, but the project staff may not have the time or the training to administer appropriate educational skills tests and other such instruments that would document the characteristics of the children referred and enrolled. Lacking such complete records, an evaluation group could administer tests on a sampling basis to estimate the appropriateness of the selection procedures and assess whether the project is serving the designated target population.

When projects are not limited to selected, narrowly defined groups of individuals but instead take in entire communities, the most efficient and sometimes the only way to examine whether the presumed population at need is being reached is to conduct a community survey. Various types of health, educational, recreational, and other human service programs are often community-wide, although their intended target populations may be selected groups, such as delinquent youths, the aged, or women of child-bearing age. In such cases, surveys are the major means of assessing whether targets have been reached.

The evaluation of the *Feeling Good* television program illustrates the use of surveys to provide data on a project with a national audience. The program, an experimental production of the Children's Television Workshop (the producer of *Sesame Street*), was designed to motivate adults to engage in preventive health practices. Although it was accessible to homes of all income levels, its primary purpose was to motivate low-income families to improve their health practices. The Gallup organization conducted four national surveys, each of approximately 1,500 adults, at different times during the weeks *Feeling Good* was televised. The data provided estimates of the size of the viewing audiences as well as of the viewers' demographic, socioeconomic, and attitudinal characteristics (Mielke and Swinehart, 1976). The major finding was that the program largely failed to reach the target group, and the program was discontinued.

To measure coverage of Department of Labor programs, such as training and public employment, the department started a periodic national sample survey. The Survey of Income and Program Participation is now carried out by the Bureau of the Census and measures participation in social programs conducted by many federal departments. This large survey, now a three-year panel covering 21,000 households, ascertains through personal interviews whether each adult member of the sampled households has ever participated or is currently participating in any of a number of federal programs. By contrasting program participants with nonparticipants, the survey provides information on the programs' biases in coverage. In addition, it generates information on the uncovered but eligible target populations.

Assessing Bias: Program Users, Eligibles, and Dropouts

An assessment of bias in program participation can be undertaken by examining differences between individuals who participate in a program and either those who drop out or those who are eligible but do not participate at all. In part, the drop-out rate, or attrition, from a project may be an indicator of clients' dissatisfaction with intervention activities. It also may indicate conditions in the community that militate against full participation. For example, in certain areas lack of adequate transportation may prevent those who are otherwise willing and eligible from participating in a program.

It is important to be able to identify the particular subgroups within the target population who either do not participate at all or do not follow through to full participation. Such information not only is valuable in judging the worth of the effort but also is needed to develop hypotheses about how a project can be modified to attract and retain a larger proportion of the target population. Thus, the qualitative aspects of participation may be important not only for monitoring purposes but also for subsequent program planning.

Data about dropouts may come either from service records or from surveys designed to find nonparticipants. However, community surveys usually are the only feasible means of identifying eligible persons who have not participated in a program. The exception, of course, is when adequate information is available about the entire eligible population prior to the implementation of a project (as in the case of data from a census or screening interview). Comparisons with either data gathered for project planning purposes or community surveys undertaken during and subsequent to the intervention may employ a variety of analytical approaches, from purely descriptive methods to highly complex models.

In **Chapter 11 (ch11.html)**, we describe methods of analyzing the costs and benefits of programs to arrive at measures of economic efficiency. Clearly, for calculating costs it is important to have estimates of the size of populations at need or risk, the groups who start a program but drop out, and the ones who participate to completion. The same data may also be used in estimating benefits. In addition, they are highly useful in judging whether a project should be continued and whether it should be expanded in either the same community or other locations. Furthermore, project staff require this kind of information to meet their managerial and accountability responsibilities. Although data on project participation cannot substitute for knowledge of impact in judging either the efficiency or the effectiveness of projects, there is little point in moving ahead with an impact analysis without an adequate description of the extent of participation by the target population.

6.4 Monitoring Organizational Functions

Monitoring of the critical organizational functions and activities of a program focuses on whether the program is performing well in managing its efforts and using its resources to accomplish its essential tasks. Chief among those tasks, of course, is delivering the intended services to the target population. In addition, programs have various support functions that must be carried out to maintain the viability and effectiveness of the organization, for example, fund-raising, promotion advocacy, and governance and management. Program process monitoring seeks to determine whether a program's actual activities and arrangements sufficiently approximate the intended ones.

Once again, program process theory as described in **Chapter 5 (ch05.html)** is a useful tool in designing monitoring procedures. In this instance, what we called the organizational plan is the relevant component. A fully articulated process theory will identify the major program functions, activities, and outputs and show how they are related to each other and to the organizational structures, staffing patterns, and resources of the program. This depiction provides a map to guide the evaluator in identifying the significant program functions and the preconditions for accomplishing them. Program process monitoring then becomes a matter of identifying and measuring those activities and conditions most essential to a program's effective performance of its duties.

Service Delivery Is Fundamental

As mentioned earlier in this chapter, for many programs that fail to show impacts, the problem is a failure to deliver the interventions specified in the program design, a problem generally known as *implementation failure.* There are three kinds of implementation failures: First, no intervention, or not enough, is delivered; second, the wrong intervention is delivered; and third, the intervention is unstandardized or uncontrolled and varies excessively across the target population.

"Nonprograms" and Incomplete Intervention

Consider first the problem of the "nonprogram" (Rossi, 1978). McLaughlin (1975) reviewed the evidence on the implementation of Title I of the Elementary and Secondary Education Act, which allocated billions of dollars yearly to aid local schools in overcoming students' poverty-associated educational deprivations. Even though schools had expended the funds, local school authorities were unable to describe their Title I activities in any detail, and few activities could even be identified as educational services delivered to schoolchildren. In short, little evidence could be found that school programs existed that were directed toward the goal of helping disadvantaged children.

The failure of numerous other programs to deliver services has been documented as well. Datta (1977), for example, reviewed the evaluations on career education programs and found that the designated targets rarely participated in the planned program activities. Similarly, an attempt to evaluate PUSH-EXCEL, a program designed to motivate disadvantaged high school students toward higher levels of academic achievement, disclosed that the program consisted mainly of the distribution of buttons and hortative literature and little else (Murray, 1980).

A delivery system may dilute the intervention so that an insufficient amount reaches the target population. Here the problem may be a lack of commitment on the part of a front-line delivery system, resulting in minimal delivery or "ritual compliance," to the point that the program does not exist. **6-G (http://content.thuzelearning.com/books/Rossi.1553.18.1/sections/nav_43#exhibit6-G)** , for instance, expands on an exhibit presented in **Chapter 2 (ch02.html)** to describe the implementation of welfare reform in which welfare workers communicated little to clients about the new policies.

Wrong Intervention

The second category of program failure—namely, delivery of the wrong intervention—can occur in several ways. One is that the mode of delivery negates the intervention. An example is the Performance Contracting experiment, in which private firms that contracted to teach mathematics and reading were paid in proportion to pupils' gains in achievement. The companies faced extensive difficulties in delivering the program at school sites. In some sites the school system sabotaged the experiments, and in others the companies were confronted with equipment failures and teacher hostility (Gramlich and Koshel, 1975).

Another way in which wrong intervention can result is when it requires a delivery system that is too sophisticated. There can be a considerable difference between pilot projects and full-scale implementation of sophisticated programs. Interventions that work well in the hands of highly motivated and trained deliverers may end up as failures when administered by staff of a mass delivery system whose training and motivation are less. The field of education again provides an illustration: Teaching methods such as computer-assisted learning or individualized instruction that have worked well within experimental development centers have not fared as well in ordinary school systems because teachers did not have sufficient computer skills.

The distinction made here between an intervention and its mode of delivery is not always clear-cut. The difference is quite clear in income maintenance programs, in which the "intervention" is the money given to beneficiaries and the delivery modes vary from automatic deposits in savings or checking accounts to hand delivery of cash to recipients. Here the intent of the program is to place money in the hands of recipients; the delivery, whether by electronic transfer or by hand, has little effect on the intervention. In contrast, a counseling program may be handled by retraining existing personnel, hiring counselors, or employing certified psychotherapists. In this case, the distinction between treatment and mode of delivery is fuzzy, because it is generally acknowledged that counseling treatments vary by counselor.

EXHIBIT 6-G

On the Front Lines: Are Welfare Workers Implementing Policy Reforms?

In the early 1990s, the state of California initiated the Work Pays demonstration project, which expanded the state job preparation program (JOBS) and modified Aid to Families with Dependent Children (AFDC) welfare policies to increase the incentives and support for finding employment. The Work Pays demonstration was designed to "substantially change the focus of the AFDC program to promote work over welfare and self-sufficiency over welfare dependence."

The workers in the local welfare offices were a vital link in the implementation of Work Pays. The intake and redetermination interviews they conducted represented virtually the only in-person contact that most clients had with the welfare system. This fact prompted a team of evaluators to study how welfare workers were communicating the Work Pays policies during their interactions with clients.

The evaluators reasoned that worker-client transactions appropriate to the policy would involve certain "information content" and "use of positive discretion." Information content refers to the explicit messages delivered to clients; it was expected that workers would notify clients about the new

program rules for work and earnings, explain opportunities to combine work and welfare to achieve greater self-sufficiency, and inform them about available training and supportive services. Positive discretion relates to the discretion workers have in teaching, socializing, and signaling clients about the expectations and opportunities associated with welfare receipt. Workers were expected to emphasize the new employment rules and benefits during client interviews and communicate the expectation that welfare should serve only as temporary assistance while recipients prepared for work.

To assess the welfare workers' implementation of the new policies, the evaluators observed and analyzed the content of 66 intake or redetermination interviews between workers and clients in four counties included in the Work Pays demonstration. A structured observation form was used to record the frequency with which various topics were discussed and to collect information about the characteristics of the case. These observations were coded on the two dimensions of interest: (1) information content and (2) positive discretion.

The results, in the words of the evaluators:

> In over 80% of intake and redetermination interviews workers did not provide and interpret information about welfare reforms. Most workers continued a pattern of instrumental transactions that emphasized workers' needs to collect and verify eligibility information. Some workers coped with the new demand by providing information about work-related policies, but routinizing the information and adding it to their standardized, scripted recitations of welfare rules. Others were coping by particularizing their interactions, giving some of their clients some information some of the time, on an ad hoc basis.

These findings suggest that welfare reforms were not fully implemented at the street level in these California counties. Worker-client transactions were consistent with the processing of welfare claims, the enforcement of eligibility rules, and the rationing of scarce resources such as JOBS services; they were poorly aligned with new program objectives emphasizing transitional assistance, work, and self-sufficiency outside the welfare system. (pp. 18-19)

SOURCE: Adapted by permission from Marcia K. Meyers, Bonnie Glaser, and Karin MacDonald, "On the Front Lines of Welfare Delivery: Are Workers Implementing Policy Reforms?" *Journal of Policy Analysis and Management,* 1998, 17(1):1-22. Copyright © 1998, John Wiley & Sons, Inc.

Unstandardized Intervention

The final category of implementation failures includes those that result from unstandardized or uncontrolled interventions. This problem can arise when the design of the program leaves too much discretion in implementation to the delivery system, so that the intervention can vary significantly across sites. Early programs of the Office of Economic Opportunity provide examples. The Community Action Program (CAP) gave local communities considerable discretion in choosing among a variety of actions, requiring only "maximum feasible participation" on the part of the poor. Because of the resulting disparities in the programs of different cities, it is almost impossible to document what CAP programs accomplished (Vanecko and Jacobs, 1970).

Similarly, Project Head Start gave local communities funds to set up preschool teaching projects for underprivileged children. In its first decade, Head Start centers varied by sponsoring agencies, coverage, content, staff qualifications, objectives, and a host of other characteristics (Cicirelli, Cooper, and Granger,1969).Because there was no specified Head Start design, it was not possible to conclude from an evaluation of a sample of projects whether the Head Start concept worked. The only generalization that could be made was that some projects were effective, some were ineffective, and, among the effective ones, some were more successful than others. Only in the past decade has a degree of standardization been achieved to the point that in 2001 it was possible to design and start an evaluation that promises to provide estimates of the program's effectiveness (Advisory Committee on Head Start Research and Evaluation, 1999).

The Delivery System

A program's delivery system can be thought of as a combination of pathways and actions undertaken to provide an intervention. It usually consists of a number of separate functions and relationships. As a general rule, it is wise to assess all the elements unless previous experience with certain aspects of the delivery system makes their assessment unnecessary. Two concepts are especially useful for monitoring the performance of a program's delivery system: specification of services and accessibility.

Specification of Services

A specification of services is desirable for both planning and monitoring purposes. This consists of specifying the actual services provided by the program in operational (measurable) terms. The first task is to define each kind of service in terms of the activities that take place and the providers who participate. When possible, it is best to separate the various aspects of a program into separate, distinct services. For example, if a project providing technical education for school dropouts includes literacy training, carpentry skills, and a period of on-the-job apprenticeship work, it is advisable to separate these into three services for monitoring purposes. Moreover, for estimating program costs in cost-benefit analyses and for fiscal accountability, it is often important to attach monetary values to different services. This step is important when the costs of several programs will be compared or when the programs receive reimbursement on the basis of the number of units of different services that are provided.

For program process monitoring, simple, specific services are easier to identify, count, and record. However, complex elements often are required to design an implementation that is consistent with a program's objectives. For example, a clinic for children may require a physical exam on admission, but the scope of the exam and the tests ordered may depend on the characteristics of each child. Thus, the item "exam" is a service but its components cannot be broken out further without creating a different definition of the service for each child examined. The strategic question is how to strike a balance, defining services so that distinct activities can be identified and counted reliably while, at the same time, the distinctions are meaningful in terms of the program's objectives.

In situations where the nature of the intervention allows a wide range of actions that might be performed, it may be possible to describe services primarily in terms of the general characteristics of the service providers and the time they spend in service activities. For example, if a project places master craftspersons in a low-income community to instruct community members in ways to improve their dwelling units, the craftspersons' specific activities will probably vary greatly from one household to another. They may advise one family on how to frame windows and another on how to shore up the foundation of a house. Any monitoring scheme attempting to document such services could describe the service activities only in general terms and by means of examples. It is possible, however, to specify the characteristics of the providers—for example, that they should have five years of experience in home construction and repair and knowledge of carpentry, electrical wiring, foundations, and exterior construction—and the amount of time they spend with each service recipient.

Indeed, services are often defined in terms of units of time, costs, procedures, or products. In a vocational training project, service units may refer to hours of counseling time provided; in a program to foster housing improvement, they may be defined in terms of amounts of building materials provided; in a cottage

industry project, service units may refer to activities, such as training sessions on how to operate sewing machines; and in an educational program, the units may be instances of the use of specific curricular materials in classrooms. All these examples require an explicit definition of what constitutes a service and, for that service, what units are appropriate for describing the amount of service.

Accessibility

Accessibility is the extent to which structural and organizational arrangements facilitate participation in the program. All programs have a strategy of some sort for providing services to the appropriate target population. In some instances, being accessible may simply mean opening an office and operating under the assumption that the designated participants will "naturally" come and make use of the services provided at the site. In other instances, however, ensuring accessibility requires outreach campaigns to recruit participants, transportation to bring persons to the intervention site, and efforts during the intervention to minimize dropouts. For example, in many large cities, special teams are sent out into the streets on very cold nights to persuade homeless persons sleeping in exposed places to spend the night in shelters.

A number of evaluation questions arise in connection with accessibility, some of which relate only to the delivery of services and some of which have parallels to the previously discussed topic of service utilization. The primary issue is whether program actions are consistent with the design and intent of the program with regard to facilitating access. For example, is there a Spanish-speaking staff member always available in a mental health center located in an area with a large Hispanic population?

Also, are potential targets matched with the appropriate services? It has been observed, for example, that community members who originally make use of emergency medical care for appropriate purposes may subsequently use them for general medical care. Such misuse of emergency services may be costly and reduce their availability to other community members. A related issue is whether the access strategy encourages differential use by targets from certain social, cultural, and ethnic groups, or whether there is equal access for all potential targets.

Program Support Functions

Although providing the intended services is presumed to be a program's main function, and one essential to monitor, most programs also perform important support functions that are critical to their ability to maintain themselves and continue to provide service. These functions are of interest to program administrators, of course, but often they are also relevant to monitoring by evaluators or outside decisionmakers. Vital support functions may include such activities as fund-raising; public relations to enhance the program's image with potential sponsors, decisionmakers, or the general public; staff training, including the training of the direct service staff; recruiting and retention of key personnel; developing and maintaining relationships with affiliated programs, referral sources, and the like; obtaining materials required for services; and general advocacy on behalf of the target population served.

Program process monitoring schemes can, and often should, incorporate indicators of vital program support functions along with indicators relating to service activities. In form, such indicators and the process for identifying them are no different than for program services. The critical activities first must be identified and described in specific, concrete terms resembling service units, for example, units of fund-raising activity and dollars raised, training sessions, advocacy events, and the like. Measures are then developed that are capable of differentiating good from poor performance and that can be regularly collected. These measures are then included in the program monitoring procedures along with those dealing with other aspects of program performance.

6.5 Analysis of Program Process Monitoring Data

Data, of course, are useful only when they have been appropriately analyzed. In general, the analysis of program process monitoring data addresses the following three issues: description of the program operation, comparison between sites, and conformity of the program to its design.

Description of the Program Operation

Assessing the extent to which a program as implemented resembles the program as designed depends on having a full and accurate description of how the program actually operates. A description derived from program process monitoring data would cover the following topics: estimates of coverage and bias in participation, the types of services delivered, the intensity of services given to participants of significant kinds, and the reactions of participants to the services delivered. Descriptive statements might take the form of narrative accounts, especially when monitoring data are derived from qualitative sources, or quantitative summaries in the form of tables, graphs, and the like.

Comparison Between Sites

When a program includes more than one site, a second question concerns differences in program implementation between the sites. Comparison of sites permits an understanding of the sources of diversity in program implementation and, ultimately, outcomes, such as differences in staff, administration, targets, or surrounding environments, and it also can facilitate efforts to achieve standardization.

Conformity of the Program to Its Design

The third issue is the one with which we began: the degree of conformity between a program's design and its implementation. Shortfalls may occur because the program is not performing functions it is expected to or because it is not performing them as well as expected. Such discrepancies may lead to efforts to move the implementation of a project closer to the original design or to a respecification of the design itself. Such analysis also provides an opportunity to judge the appropriateness of performing an impact evaluation and, if necessary, to opt for more formative evaluation to develop the desired convergence of design and implementation.

Summary

- Program process evaluation is a form of evaluation designed to describe how a program is operating and assess how well it performs its intended functions. It builds on program process theory, which identifies the critical components, functions, and relationships assumed necessary for the program to be effective. Where process evaluation is an ongoing function involving repeated measurements over time, it is referred to as program process monitoring.

- The criteria for assessing program process performance may include stipulations of the program theory, administrative standards, applicable legal, ethical, or professional standards, and after-the-fact judgment calls.

- The common forms of program process evaluation include process (or implementation) evaluation and program process monitoring.

- Process evaluation assesses whether the program is delivered as intended to the targeted recipients and is typically conducted as a separate project by evaluation specialists. It may constitute a stand-alone evaluation when the only questions are about implementation of program operations, service delivery, and other such matters. Process evaluation is also often carried out in conjunction with an impact evaluation to determine what services the program provides to complement findings about what impact those services have.

- When a program has a well-developed MIS, program process monitoring can be integrated into a program's routine information collection and reporting.

- Program process monitoring takes somewhat different forms and serves different purposes when undertaken from the perspectives of evaluation, accountability, and program management, but the types of data required and the data collection procedures used generally are the same or overlap considerably. In particular, program process monitoring generally involves one or both of two domains of program performance: service utilization and organizational functions.

- Service utilization issues typically break down into questions about coverage and bias. The sources of data useful for assessing coverage are program records, surveys of program participants, and community surveys. Bias in program coverage can be revealed through comparisons of program users, eligible nonparticipants, and dropouts.

- Monitoring of a program's organizational functions focuses on how well the program is organizing its efforts and using its resources to accomplish its essential tasks. Particular attention is given to identifying shortcomings in program implementation that prevent a program from delivering the intended services to the target population. Three sources of such implementation failures are incomplete interventions, delivery of the wrong intervention, and unstandardized or uncontrolled interventions.

- Monitoring of organizational functions also includes attention to the delivery system and program support functions.

- The analysis of monitoring data typically addresses such issues as description of program operations, comparison of sites, and conformity of a program to its design.

Key Concepts

Accessibility

The extent to which the structural and organizational arrangements facilitate participation in the program.

Accountability

The responsibility of program staff to provide evidence to stakeholders and sponsors that a program is effective and in conformity with its coverage, service, legal, and fiscal requirements.

Administrative standards

Stipulated achievement levels set by program administrators or other responsible parties, for example, intake for 90% of the referrals within one month. These levels may be set on the basis of past experience, the performance of comparable programs, or professional judgment.

Bias

As applied to program coverage, the extent to which subgroups of a target population are reached unequally by a program.

Coverage

The extent to which a program reaches its intended target population.

Management information system (MIS)

A data system, usually computerized, that routinely collects and reports information about the delivery of services to clients and, often, billing, costs, diagnostic and demographic information, and outcome status.

Outcome monitoring

The continual measurement and reporting of indicators of the status of the social conditions a program is accountable for improving.

Program process monitoring

Process evaluation that is done repeatedly over time.