Programs have fuzzy borders, so that it is not always plain sailing to determine precisely what the program was that required evaluation.


We should employ metaphors of process, focused activity, marginal change, cumulative transformation, and above all think more imaginatively in terms of determined models of formative, generative processes.

—Frederik Barth (1990, p. 652)

People in the evaluation world talk about program evaluation, but the textbooks and journals put the emphasis on evaluation rather than on the program. Program is somebody else’s business. Evaluators come in to whatever program is on the scene, pull out their methodological bag of tricks, and proceed to do the evaluation.

This chapter recommends that evaluators pay close attention to the program. They should be familiar with the general field very early in the game, and they should learn a good deal about the specific happenings of the program under study as the evaluation moves along. They should also understand exactly how the program expects to bring about required change. There are a number of significant reasons for this advice.

**Why Program Knowledge Pays**

**To Develop a Good Sense of the Issues**

When the evaluator is well informed at the outset about the type of programming (e.g., family planning services or energy conservation), she has a better idea of the issues that evaluation should address. She knows the general line of development of programming and the controversies in the field. She can gear the study to issues that are current and pressing.

**To Formulate Questions**

Knowledge of the program helps the evaluator to make evaluative questions relevant and incisive. She can then collect the appropriate data to answer the questions.

**To Understand the Data**

As evaluation data start coming in, some of the information may look anomalous. Even well along in the study, unexpected and seemingly incongruous findings sometimes emerge. To make sense of them, it helps if the evaluator understands the way that the program is working and how practitioners interact with clients. Such knowledge enables the evaluator to interpret the information she receives and perhaps strike out into new areas of investigation.

**To Interpret the Evidence**

Only with a good sense of what the program is and how it has been working can the evaluator interpret the meaning of outcome data. If, for example, the program has switched direction in the middle (say, moving from didactic lectures to small-group work in class), knowing about the change is important for drawing conclusions about what worked and what didn’t in the program.

**To Make Sound Recommendations**

Policymakers and program directors expect good data from an evaluation, but often they want more. At the conclusion of a study they want sage counsel about what to do next. If the evaluator is to come out at the end and say something smart and useful about the program, it helps if she knows something about the program going in.

Too often in the past the evaluator’s recommendations were afterthoughts, poorly thought through, often just the opposite of whatever condition seemed to be associated with low effectiveness. (If the rehabilitation counselors in a low-effectiveness program had low levels of education, the recommendation might be to hire better educated staff—even though there was no evidence that the education of staff affected client rehabilitation.) The evaluator without program savvy has little way of knowing whether her potential recommendations are practical, feasible to implement, politically acceptable, financially affordable, or stand much chance of making things better. Program knowledge improves the odds.

**For Reporting**

However carefully the evaluator may specify that she is studying the Southwestern Lagos Water Development Project, readers of the report are likely to think of the study as an evaluation of water development projects in less developed countries generally and to draw conclusions that go far beyond the case at hand. The report should do more than name the project. It should specify what the project did, how it was run, what was in and not in the project, and even something about surrounding...
Meta-analysis is the systematic summary of the results from a number of different evaluations of the same kind of program. In recent years evaluators have developed statistical procedures for combining data from different studies in order to reach summary estimates of the extent of success of, say, preschool programs or juvenile delinquency prevention programs. In Chapter 10, I discuss methods of meta-analysis.

It is important to note that meta-analysts need a lot of information about what the program actually was and did. Particularly when they try to characterize programs as open classroom programs or mental health counseling programs, they need to understand the reality of the program. Often they have to decide whether a program fits their label, and to do that, they need to know whether it fulfills 5 or 10 or 20 of the criteria of open classrooms (Tyler, 1991/1942) or mental health counseling. More than that, meta-analysts are now trying to understand which specific features of a program are responsible for better or poorer showings (Cook et al., 1992). For this task, specification of the attributes of the program-in-operation are essential.

For all these reasons, evaluators are well advised to understand the larger field of programming and the specifics of the program or project they are studying. Most evaluators are not experts in a particular program field. Only after some years of working in the same domain do they acquire a degree of expertise. But right from the start they can take steps to advance their understanding.

Social programs are complex undertakings. They are an amalgam of dreams and personalities, rooms and theories, paper clips and organizational structure, clients and activities, budgets and photocopies, and great intentions. Evaluators of social programs look with something akin to jealousy at evaluators in agriculture who evaluate a new strain of wheat or evaluators in medicine who evaluate the effects of a new drug. These are physical things you can see, touch, and—above all—replicate. The same stimulus can be produced again, and other researchers can study its consequences—under the same or different conditions, with similar or different subjects, but with some assurance that they are looking at the effects of the same thing.

Social programs are not nearly so specific. They incorporate a range of components, styles, people, and procedures. It becomes difficult to describe what the program really is. In special cases, a program can be expressed in terms that are clear and reproducible: for example, a change in highway speed limits or a decrease in the size of probation officers’ caseloads. Here the program is the change in speed limits or caseloads, and people want to know how much difference the change in scale makes. Another kind of readily describable program is one that involves an increase or decrease in cash payments or charges. Examples would be an increase in unemployment benefits, a decrease in subsidies to peanut farmers, a change in the minimum wage, a rise in rents charged in public housing.

Fairly easy to characterize, too, are programs that involve a mechanism that is well understood or available for inspection. Mammography is mammography, and to understand a program that makes mammograms available, the evaluator probably needs to know only where and to whom they are offered. A new mathematics curriculum also looks fairly clear. However, teachers may implement the curriculum in vastly different ways, and the evaluator cannot assume that the written curriculum equals the program. She needs to find out how different teachers use the new material in class.

In most social programs, it takes effort to understand the content of the program, what actually goes on. Operations often differ markedly from day to day and from staff member to staff member. With programs as large and amorphous as regional economic development or federal technical assistance for state health planning, it takes a major effort just to describe and analyze what the program is. Process evaluation specializes in comprehending the operations of the program.

In outcome evaluation, an evaluator may question whether she needs to spend much time worrying about program content. If evaluation data show that the program yields the desired results, does it make any difference whether it is using rote drill, psychoanalysis, or black magic? A few evaluators still see the program as a black box, the contents of which do not concern them; they are charged with discovering effects. But if the evaluator has no idea of what the program really is, she may fail to ask the right questions. Perhaps she believes the inflated barrage of program propaganda and expects mountain-moving outcomes from what are really puny efforts. More likely, she looks for the wrong order of effects. She looks for the attainment of the types of outcomes that have been verbalized, when the main resources of the operating program have been invested in a different course of action.

Furthermore, unless there is some reasonably accurate and coherent definition of the program, the evaluator does not know to what to attribute the outcomes she observes. Decision makers need to know what it was that worked or didn’t work, what it is that should be adopted throughout the system or modified. In an extreme case, when a program is a smashing success and 40 communities want to adopt it forthwith, what is it that we tell them to adopt?

The evaluator has to discover the reality of the program rather than its illusion. If she accepts the description given in the application for funds or in publicity releases, she may evaluate a phantom program. The study will be attributing effects (or no effects) to a program that never took place at all, or one that operated at so low a level of competence or in such a different manner that it hardly deserves to be called by the program name. For example, a business school is supposed to be introducing consideration of ethics into its courses, but some faculty members do not know how to go about this kind of teaching and some do not see the new charge as their business. A recreation program for schoolchildren may be closed on weekends when children have free time; it may be closed erratically so that the children get discouraged from attending; it may be offering activities that fail to attract participants. It will hardly be necessary to collect data on the effects of these programs, because there is little or no program in operation.
Getting to Know the Program

Evaluators can take a number of steps to familiarize themselves with the program area and with the specific program they will be studying. The first and most obvious way is to read about the field. Read what people have written about traffic safety or vocational guidance or whatever the program area may be. Find a good overview book or journal article, perhaps with the help of a reference librarian, and pursue citations in that source that sound interesting. There may be an annual review in the field or an informative overview in an entry in a specialty encyclopedia.

A special resource is previous evaluations. Find as many reports on similar programs as possible, whether they are published or unpublished. The prospective evaluator will learn a good deal about the problematics of the program, and along the way will pick up tips about methodology—the kinds of methods that others have used and how well these methods worked to yield relevant information. Prior evaluations are full of information and object lessons, good and bad. It is worth taking full advantage of the program knowledge they provide. In fact, a required first step in conducting an evaluation should be to read previous evaluation reports on programs of the same and similar kind. Reviews of the literature may exist that summarize a spectrum of past results and identify key issues and problems.

Another source of program information about the generic field is people, particularly people who work in programs of the type the evaluator will study. Talking to program directors and practitioners will give the evaluator insights that “book learning” may not divulge.

Next the evaluator needs to learn about the immediate program at hand. Her approach can vary from an informal and unstructured look-around all the way to systematic, data-based inquiry. At the informal level, observation is much on the order of what writers on organizational management call “management by walking around” (Peters & Waterman, 1982). You walk around, look, and chat. When visits are unscheduled, the observer has a sense that she is getting an inside view of the program. A few meetings with selected groups (staff, administrators, clients) can yield additional information. In a more formal way, a qualitative pilot study can contribute systematic information.

What the observer finds out depends in large measure on where she looks, whom she talks to, and how much prior programmatic knowledge she has. Prior knowledge is important so that the observer knows where to look and whom to talk to. If she is a novice and uninformed about the extent of possible variation from site to site, or unaware of the kinds of misfeasance and nonfeasance that program folks might like to keep under wraps, those who are in charge of the tour can, if they choose, direct her to places that will give the picture they want to paint. But with a little program savvy, the observer can find out a great deal through informal observation and conversation. She just needs to be sure to touch a variety of bases and ask probing questions.

A more systematic procedure is monitoring. Monitoring is similar to process evaluation but usually somewhat less formal and intended for a different purpose. Monitoring is generally undertaken by the funding agency that provides the financial support for the program or the oversight agency responsible for program stan-

dards. These agencies want to know what is happening in the program for accountability purposes (and perhaps as a guide to the kinds of technical assistance the project needs). They have a responsibility to be sure that local projects are doing what they are supposed to do. Legislators want to know whether the intent of the legislation they passed is being carried out.

Monitoring, as Rossi and Freeman (1993) write, is frequently directed at finding out (a) how well the program is reaching the target population it is intended to serve, (b) its fidelity to the original design for activities, and (c) the appropriateness of its expenditure of money and other resources. Learning these things about the program—whether it is serving the right people, delivering the right kind of services, or spending its money in the right ways—gives funders and oversight agencies the information to oversee the appropriateness of local operations.

Note that under the usual definition, monitoring starts with a standard in view. It knows which categories of people are supposed to be served, which kinds of activities should be carried out, and how money should be spent, and it measures what the program does against these yardsticks. The criteria that monitoring applies generally come from the initial charter for the program and the rules of the funding agency. While these existing standards give the monitoring process a clear focus, they are also a set of blinders. They direct attention toward a limited scope of conditions and away from all the other things happening in the program and in the environment.

If program staff take a less blinkered view, agencies can learn more from monitoring than merely whether the rules are being followed. If they find widespread deviations from intended ways of work, it is true that the reason may be the shortcomings of the projects. But another reason may be that the original design was not in tune with reality. The program was designed to serve displaced workers whose companies, or whose entire industry, had gone out of business and offer them training for new occupations. But the project may have found few such displaced workers in the area. Most seem to have found employment in other companies or moved away. When monitoring identifies such deviations from intent, and at least provisional explanations for them, it provides important clues for the modification of existing policy—or even the need for new policies.

For the evaluator, monitoring information can be a significant category of data. She can take advantage of whatever monitoring data other people have collected and she can collect her own. The data will orient her to the project and give her a sense of what is going on. If she has the time, resources, and authorization, she can go further and conduct a process evaluation study, gathering systematic data over a period of time about the implementation of program activities.

What Is the Program Trying to Achieve?

A useful way to begin conceptualizing the program is by looking at what it is trying to accomplish. A sensible place to start is with its official goals. Although official goals are only one source of understanding, they represent an entry point into the matter of program intention.

Most programs have some kind of written statement (applications for funding, recruitment announcements, public relations brochures) that describe their activities,
and often these statements contain a description of program goals. If the goals are framed in clear and specific terms, they are a good source of information. An energy conservation program aims to reduce the amount of electricity that householders use. A road-building program in a developing country seeks to enable small farmers to transport their produce to the city and thereby increase their income. These kinds of statements shed considerable light on what the program is about.

But some programs lack official statements. If the evaluator asks program staff about goals, they may discuss them in terms of the number of people they intend to serve, the kinds of service they will offer, and similar process information. For program implementers, these are program goals in a real and valid sense, but they are not the primary currency in which the evaluator deals. She is interested in the intended consequences of the program.

Other programs have goal statements that are hazy, ambiguous, or hard to pin down. Occasionally, the official goals are merely a list of pious and partly incompatible platitudes. Goals, either in official documents or in program managers' discussion, can be framed in such terms as improve education, enhance the quality of life, improve the life chances of children and families, strengthen democratic processes. Such global goals give little direction for an evaluator who wants to understand the program in detail.

This kind of vague formulation is so commonplace that some writers on evaluation discard the discussion of goals as being too abstract and long term. They prefer to talk about objectives as the near-term down-to-earth effects at which programs set their sights. They say that it is possible for an evaluation to find out the extent to which programs meet their objectives (in this location), where it would be impossible to tell whether they meet their goals of improving education or enhancing the quality of life. I agree that we should concentrate on specific objectives, but I don't want to abandon the useful word goals. Instead, I'll try to bring it down from the stratosphere and use it to mean a program's specific, measurable, near-term aims.

Evaluators wonder why program goals are often stated in fuzzy terms. Part of the explanation probably lies in practitioners' concentration on concrete matters of program functioning and their pragmatic mode of operation. They often have an intuitive rather than an analytic approach to program development and concentrate on formulating activities rather than objectives. But there is also a sense in which ambiguity serves a useful function: It masks underlying divergences in intent. Support from many quarters inside and outside the program agency is required to get a program off the ground, and the glittering generalities that pass for goal statements are meant to satisfy a variety of interests and perspectives. Everyone can agree on improving the neighborhood, where they might come to (verbal) blows over identification of the specific outcomes they have in mind.

Sometimes the divergence in intentions that is papered over through vagueness operates at different levels of the hierarchy. The higher-ups may have one set of purposes—for example, to reduce fraud and overpayment to recipients of public assistance. Direct-service staff may have a different inclination. More sympathetic to the clients with whom they deal, they may want to get them every cent that is legally coming to them and even deliver additional services that improve their lot. If the evaluator listens to official statements, she may learn about the mission of reduc-

ing fraud and abuse, but she may not hear about the tension between reducing welfare outlays and maximizing payments and services.

When goals are unclear or ambiguous, more than the evaluation can be affected. Where there is little consensus on what a program is trying to do, the staff may be working at cross-purposes. One side benefit of evaluation is to focus attention on the formulation of goals in terms of the specific behaviors that program practitioners aim to achieve. The effort may force disagreements into the open and lead to conflict. But if differences can be reconciled (and the program may not be viable if they are not), the clarification can hardly help but rationalize program implementation. It may reveal discrepancies between program goals and program content, in which case either the content or the goals should be changed. When a sense of common purpose is reached, the logic and rationality of practice are likely to be enhanced.

A similar kind of problem for the evaluator arises when conditions change. The official goals were fine when the program began and fully descriptive of what the program aimed to do. However, over time, the environment changed, the type of client changed, the staff changed, the budget changed, and the reigning idea of how the program should operate also changed. The program has become considerably different from what it was when the goals were first enunciated. They no longer fit.

When the evaluator recognizes the obsolescence of the goal statement, she can get contemporary information from program managers and staff. They can be a first-rate source and fill her in on intents, as well as on subtle shadings of intention that the official statements rarely mention. However, direct conversations with the people involved do not necessarily solve all problems. Sometimes staff find it difficult to arrive at a clear-cut formulation of the program's goals. Kallen (1966) tells of working with a committee to discuss evaluation of a program for gang youth. Asked to specify the program's goals, the committee members came up with such things as improving the behavior of the youth, helping them become better citizens, and improving their schoolwork. When they tried to be more specific about goals, they felt that behavior and citizenship were too vague to describe what they really were trying to get youth to learn, and they weren't sure that the program actually did anything directly to improve youths' school grades. Kallen reported the discouraging story:

Finally it turned out that a number of the area residents objected to the young people's use of swear words, and it was decided that one measure of behavioral improvement would be the reduction in swearing. and that this was something the detached worker should aim for in his interaction with the youngsters he was working with. [Was the group identifying program goals or making up new ones?] It was therefore agreed that part of the criteria of success would be a reduction in swearing. I might add that this was the only measure of success upon which the evaluation team and the program advisory committee could agree.

A further complication arises when the program seeks not only to reach official goals but to achieve other things, sometimes in addition, sometimes instead. For example, a social service agency applied for, and received, government funds to run a training program for the unemployed. The goal that it was willing to state up front
was that the program would teach the trainees new skills and help them find new jobs. But the agency had other unstated purposes: to improve community relations with people of color in its neighborhood, meet federal regulations against discrimination, encourage members of minority groups to work on its staff, and gain the additional funding that came with the program grant (Meyers, 1981). These are perfectly legitimate objectives. Agencies have to be concerned with more than reaching instrumental goals; they have to maintain the organization in its environment. An exclusive focus on program goals runs the risk of ignoring the system-maintenance goals that are part of organizational life. Being concerned about these other things doesn’t mean that the agency is lying about its interest in doing a good job training program for the unemployed, but doing good job training is just one of the items on its list.

Sometimes system-maintenance goals get in the way of attaining program objectives. For example, an agency runs an outpatient mental health program to help troubled people cope with daily living. Helping troubled people is one goal. Another is preventing staff members from burnout, being overwhelmed by the demands of patients. In some cases, the goal of making life easier for staff supersedes the therapeutic goal, and the quality of service deteriorates.

Clients of the program are often worth consulting. They may have different ideas about goals. They may have enrolled in the program to accomplish something different from the things that staff are purporting to do. Staff of a methadone maintenance project may discuss their goal of weaning drug abusers from heroin and thereby preventing crimes committed to pay for heroin. The clients may or may not value methadone as a substitute for the highs of illegal drugs, but their expectations from the program may center on its function as a social setting, a hangout, a place to spend time with friends in social activities. To understand what goes on in the program, and the potential conflicts between staff and client expectations, the evaluator may want to hear the goals of other parties besides the staff.

Then, too, many programs have multiple goals. In most cases, the several goals are consistent and represent various facets of a multi-modal program. A program to upgrade a public housing project plans to rehabilitate housing units so that living space is clean, safe, in good working order, and reasonably spacious; it also plans to develop a tenants’ organization so that residents will develop a warmer sense of community and take responsibility for maintaining the apartments in reasonable condition; it wants to encourage the tenants’ organization to take responsibility for policing the public spaces in order to reduce crime; and so on. These manifold purposes should ideally all be met in a successful program.

However, with complex and ambitious goals, managers and tenants have an opportunity to select that part of the charge that they wish to concentrate on. In some locations, effort may be going into something quite different. That is one of the risks in programs with complex goals—that people will choose those aspects of the multiple mission that are congenial to them and ignore the rest. To really understand the program, it pays to know which goals are real.

Goals are real to the extent that people are actively devoting time and effort to working toward them. This is not always easy to see in documents or learn from conversations with funders or program managers. From time to time, budget documents give useful clues, at least if they show on which activities money is being spent. (However, many budgets lump staff salaries into one category without indicating what functions these staff are performing.) Watching the program in operation is probably the best way to understand where energies are being spent, and talking to direct-service staff and clients can give further useful clues.

How the Program Works: Surfacing the Program’s Theories of Change

For evaluation purposes, it is useful to know not only what the program is expected to achieve but also how it expects to achieve it. As Rein (1981) has written, “A program is a theory and an evaluation is its test” (p. 141). In order to organize the evaluation to provide a responsible test, the evaluator needs to understand the theoretical premises on which the program is based.

The notion of defining program theory has not typically been a component of evaluation. Much evaluation is done by investigating outcomes without much attention to the paths by which they were produced. But evaluation is increasingly being called upon not only to answer the question “Did the program work?” but also “What made it work? Why was it successful or unsuccessful?” and even “How can we make it better?” To make a respectable contribution to such a discussion, it helps if the evaluator understands—and investigates—the program’s explicit or implicit theory.

Programs are complicated phenomena, generally born out of experience and professional lore. Teachers, probation officers, social workers, international development specialists, physicians, safety engineers—all develop social programs from a mixture of what they learned in professional school, what they encountered on the job, what stories of other people’s experience suggest, and perhaps some social science and evaluation learnings. Programs are not likely to be laid out in rational terms with clear-cut statements of why certain program activities have been selected and which actions are expected to lead to which desired ends.

Yet when the evaluator undertakes to define program theory, those are just the things she wants to know: What ideas and assumptions link the program’s inputs to attainment of the desired ends? Getting a purchase on the answer is an even harder task than learning about program goals. Goals are at least a familiar and long-standing concept. The program’s theories of change sound strange to many program designers and practitioners.

By theory, I don’t mean anything highbrow or multi-syllabic. I mean the set of beliefs that underlie action. The theory doesn’t have to be uniformly accepted. It doesn’t have to be right. It is a set of hypotheses upon which people build their program plans. It is an explanation of the causal links that tie program inputs to expected program outputs, or as Bickman (1987) has put it, “a plausible and sensible model of how a program is supposed to work” (p. 5). Wholey (1987) says that program theory identifies “program resources, program activities, and intended program outcomes, and specifies a chain of causal assumptions linking program resources, activities, intermediate outcomes, and ultimate goals” (p. 78).
Let's take a simple example. Let's say the program (or policy) is an increase in teachers’ salaries, and the increase is justified on grounds that it will increase students’ achievement. What is the theory that links higher teacher pay to increased student achievement? I can think of a number of possibilities (see Figure 3-1).

One theory might be that higher pay will improve teacher morale; higher morale will lead teachers to work harder at teaching; they will put more effort into preparation and pedagogy; and this will improve students’ understanding of material. The mechanisms here are assumed to be first an increase in morale and then an improvement in pedagogy. A variant of this theory would be that with higher morale comes a more pleasant social climate. Teachers develop more congenial relationships with students, which leads students to seek to maintain good relations with the teachers, so students work harder and therefore learn more. Here the mechanism would be better interpersonal relations.

Another theory would be that higher pay allows teachers to give up their second jobs. Teachers who used to moonlight in order to earn enough money can now put their full energies into teaching. Greater energy and attention lead to more thorough preparation of lessons, greater variety in strategies of pedagogy, more effective teaching, and thus better student learning.

A quite different explanation would be that higher salaries attract abler people into teaching. Because college students learn that they will be better paid in the teaching profession, brighter college students prepare for teaching. Because brighter students become qualified as teachers, school districts hire abler teachers, abler teachers then teach better, and better teaching leads to greater student learning. This program theory has an entirely different time dimension. Evaluation might not be able to locate effects for four to seven years, because it will take time for word of higher salaries to get around, a different caliber of college student to prepare for teaching and be hired, and for the expected “better teaching” to take place and have effects. If the expectation is that the school district offering higher salaries will lure good teachers away from other districts, the time scale would be shorter. But at least one school district relying on such an expectation found that higher salaries induced current teachers to stay on the job longer, thus preventing vacancies from opening up.

One can think of other program theories as well. Higher salaries would, in another formulation, lead to higher social status for teachers, which would lead to greater respect from parents. Greater respect would translate into greater parental support for school demands, such as helping children with homework, encouraging more regular student attendance, support for school disciplinary actions, and more parental participation in school activities. Such increased parent-teacher cooperation would give students greater respect for school and teacher, cause them to behave more properly and study more conscientiously, which would lead to better student achievement.

A final program theory might posit a chain of negative consequences. Higher teacher salaries might provoke parental resentment of the higher taxes they have to pay, which would lead to subtle acts of sabotage against teachers (such as allowing students to get away with truancy, disrespect of teachers, and disruption in the classroom), which would lead to greater classroom discontinuity and disorder, which would reduce opportunities for learning, which would reduce student achievement. To locate such counter assumptions, the evaluator can talk to opponents of the program who have theories about why the program is apt to be a dud—or worse.

Program theory, as I use the term, refers to the mechanisms that mediate between the delivery (and receipt) of the program and the emergence of the outcomes of interest. The operative mechanism of change isn’t the program activities per se but the response that the activities generate. For example, contraceptive counseling is expected to reduce the incidence of pregnancy, but the mechanism is not the counseling itself. It might be the knowledge that participants absorb from the counseling. Or it might be the confidence that women develop to assert themselves in sexual relationships. The undermining of cultural taboos against family planning, or the signal of a shift in power relations between men and women. These or any of several other cognitive/affective/social/cultural responses would be the mechanisms leading to desired outcomes.

Program Theory and Implementation Theory

Program theory puts the emphasis on the responses of people to program activities. But there is another kind of assumption in programming, one that program people are more likely to articulate: that if the program does all the things it intends to do,
good things will happen. Take the example of a job training program whose goal is to enable poor youth to get jobs and become productive earners. The program’s activities are to teach craft skills and job readiness skills, such as regular attendance and appropriate dress. Staff’s assumptions might be that the program has to take the following steps: publicize the program widely, enroll youth who meet eligibility criteria, provide good training for attractive occupations in accessible locations, give stipends, provide child care, give supportive counsel, refer youth to available jobs. If the program does all those things, the expectation is that youth will be employed.

Moving from intentions and plans to sound program activities is the major challenge that agency managers face. Much of their energy is concentrated on ensuring that the program works smoothly and well. Implementation implicitly incorporates a theory about what is required to translate objectives into ongoing service delivery and program operation. Such a theory is what I call implementation theory. The assumption is that if the activities are conducted as planned, with sufficient quality, intensity, and fidelity to plan, the desired results will be forthcoming. A number of evaluations examine implementation theories of this kind (e.g., McGraw et al., 1996; Pentz et al., 1990). Implementation theory does not deal with the processes that mediate between program services and achievement of program goals but focuses on the delivery of program services.

The two kinds of theory, program theory and implementation theory, intertwine in the evolution of the program. As Figure 3-2 shows, each stage of activity assumes an appropriate response from participants, and responses of participants condition the next stages of the program. Unless the agency carries out the scheduled activities well, it is not likely that participants will make progress toward the goal. And if the participants are uninterested, unmotivated, irregular in attendance, unwilling or unable to learn, or resistant to taking jobs, much progress will be made. The two streams of activity have to interact to produce desired outcomes.

I call the combination of program theory and implementation theory the program’s theories of change.

Organizing the Evaluation to Investigate Program Theories

A big advantage of understanding the program’s theories of change is that they can provide the basis for the evaluation. The evaluator can organize the evaluation to trace the unfolding of the assumptions. The evaluation can collect data on the interim markers that are expected to appear. In Figure 3-1, does teachers’ morale improve? Do they put more effort into preparation? If so, the evaluator can follow the chain further and find out whether teachers are in fact improving the caliber of their instruction. If morale rises but teachers are not doing anything different in the classroom, the evaluation can identify the point at which the posited theory breaks down.

In the job training program in Figure 3-2, does information about the program’s availability reach the target audience? Do young people sign up? Does the staff provide state-of-the-art training for occupations for which there is local demand? Do youth attend regularly? Do they learn? Do they develop more positive attitudes toward work? Does the program provide support and assistance? The evaluation traces each step along the route to see whether the stages appear as anticipat-
The evaluator uses program theories to plan points for data collection. (As we will note in later chapters, the evaluator can use other techniques to plan which data to collect as well.) She can collect data to find out if the program carried out each step of the plan of activities and if each step led to the next step and to expected responses from participants. (An excellent example is Holder, 1997.)

It is also wise to build in theories of intended consequences, unplanned and unwanted chains of events that the program may set in motion. Evaluators have learned that unanticipated consequences are all around us. When legislators mandate severe sentences for certain classes of crimes, a frequent consequence is that juries become less likely to convict defendants of those crimes. Evidently a jury can believe that the mandatory sentence is too harsh for the person and/or the circumstances and thus refuse to convict. Instead of increasing the severity of punishment, the new law in actuality reduces the severity of punishment. Or a school system that wants to reduce truancy among high school students adopts a policy that refuses class credit to any student with more than 10 unexcused absences in a class during the semester. Some students thereupon attend more regularly, the goal of the policy, but for those who have exceeded the 10-day maximum, there is no incentive to do the classwork, take tests seriously, or even show up.

Theory, or better still, theories, direct the evaluator’s attention to likely types of near-term and longer-term effects. The two big advantages of developing the evaluation on the basis of program theories are as follows:

First, the evaluation provides early indicators of program effectiveness. It need not wait until final outcomes appear (or fail to appear). It examines the intermediate stages between the initiation of the program and its long-term effects, and in that way gets early clues about what is going well and where anomalies are showing up. If breakdown occurs in the premises about implementation, the implications are immediate and direct: Fix the way the program is being run. If breakdown occurs in markers of early participant progress, the conceptual basis on which the program is predicated may need to be reexamined.

Early data on measures relevant to the program’s long-term success are especially useful when the program has an extended timeline. For example, a program that aims to revitalize an inner-city neighborhood will take a decade or more to achieve results. Theory-based evaluation calls for the collection and reporting of data on early phases that are conceptually linked to the program’s ability to produce long-term success. That is, the data are not just desired interim outcomes; they are the interim outcomes that are considered to be paths to long-term effects. They are part of the assumed causal chain.

The other big advantage of theory-based evaluation is that it helps to explain how and why effects occurred. If events work out as expected, the evaluator can say with a fair degree of confidence how the effects were generated. By following the sequence of stages, it shows the microsteps that led from program inputs through to outcomes. Where several different theoretical assumptions are being tracked, the evaluator can show which of them has the best empirical support.

Such detailed tracking gives a sense of whether the program was responsible for observed changes. That is a big issue in evaluation, whether the program is responsible for whatever outcomes are observed. Many things other than the program might give rise to desired outcomes. The usual way to tell whether the program was responsible is to compare units that received the program with equivalent units that did not. In many programs, such as the neighborhood revitalization program, it is not feasible to use comparison units; similar neighborhoods are too few and too different from each other to provide sensible comparisons. In order to get any purchase on the question of whether the program was the responsible agent, the tracing of program theory can help. A careful tracing of the unfolding of stages of theory provides indications of how well the theories explain what happens and therefore of how implicated the program was in events and outcomes. As Chen and Rossi (1987) state, theory-based evaluation helps to assure that the results identified are firmly connected to what the program has been doing and that they are due to program activities.

A side advantage of setting down the expected paths of change is that it sensitizes the evaluator to shifts in program strategy that make her evaluation design irrelevant. Suppose the job training program has had difficulties in operation, and to overcome them the managers have shifted course. They have found that trainees do not have transportation to the suburbs where good jobs are located. To maintain the viability of the program, managers have shifted to providing subsidized transportation. The original process model is obsolete and must be modified to reflect the new conditions. Some new processes are hypothesized and some old ones discarded; new items and directional lines are added and old ones dropped. The model is adapted to current realities, and measurements and analyses follow the new course. The evaluator, alerted early to the changes in program operations and assumptions, keeps her study flexible enough to be relevant.

In a review of social experiments, Cook and Shadish (1994) discuss the need for generalizing the causal relationships discovered through social experiments. They conclude with a call for “identifying the micro-mediating processes that causally connect a treatment to an outcome, usually through a process of theoretical specification, measurement, and data analysis” (p. 576). That is a good description of basing evaluation on program theory.

Building Program Theory

It is not necessary to come up with a single program theory that seeks to explain how the program will obtain the happy consequences expected from it. As in our teacher salary example, it is often useful to have a set of alternative theories for the evaluation study to examine. If one theory does not work out as hypothesized, there are alternative pathways to explore.

How does the evaluator go about constructing a program theory or theories? Usually she doesn’t construct theories so much as help the program people to surface their own latent theories. They have some ideas in their mind about how program inputs are supposed to lead to program outcomes, and the evaluator can ask and probe and question until they articulate their assumptions. The evaluator may find that different people on the program staff have different assumptions. That’s all right. The evaluator can work with multiple theories. She may find that some of the assumptions sound unpersuasive and overly optimistic. In such a case, she should
ask additional questions to see whether the practitioners have something further in mind, such an intermediate step or a picture of outside conditions that would foster the linkage they are positing. That information should be included in the theory, too. The evaluator should keep probing until she is sure that she has uncovered all the assumptions and expectations that program people hold.

If program people draw a blank, or find the exercise unappealing, the evaluator can send up a few trial balloons. She can offer hypotheses based on her own understanding of what the program is and does. Her knowledge may come from watching the program in operation, from acquaintance with similar programs, from social science, or from logical reasoning. Common sense is not a bad guide. Drawing on her analysis of what the program expects to achieve and how it is going about its work, she can offer speculations about the underlying assumptions. For example, if a job training program brings in local business people to teach the sessions, she may hazard guesses that the strategy is based on the businessmen’s expertise, on their willingness to donate their time free of charge, or on the hope that they will offer the trainees jobs after the program. She tries these hypotheses with program personnel. With this kind of prod, they are likely to be able to elaborate further.

Who shall settle on the final version of program theory is a matter of contention among evaluation authors. Wholey (1987) and Patton (1989) emphasize the role of program personnel and other stakeholders. Chen and Rossi (1980, 1983) put their faith in social science knowledge and theory, and therefore give the evaluator a bigger say. Chen (1990) suggests that it is possible to follow both courses and generate several theories. Then the final synthesis can be done either by the evaluator alone using her professional judgment or by the evaluator working with key stakeholders (Weiss, 1995, 1997). Inasmuch as one of the benefits of constructing program theory is the enhanced communication it generates between practitioners and evaluators (Chen, 1990; McClintock, 1990), it seems sensible to involve a number of people in the final choice.

Use of program theory as a map for evaluation doesn’t necessarily imply that every step of every possible theory has to be studied. Once the theories are in hand, the evaluator can work with funders, program managers, and others to decide which of the several lines of thinking is most central to their concerns. Choices always have to be made in designing an evaluation about which lines of inquiry to pursue. The theory map provides a picture of the whole intellectual landscape so that people can make choices with full awareness of what they are ignoring as well as what they are choosing to study.

Using Theories of Change as a Guide to Evaluation

A program theory usually includes (a) program inputs, such as resources and organizational auspices; (b) program activities, which represent the manner in which the program is implemented; (c) interim outcomes—that is, the chain of responses the activities elicit, which are expected to lead on to (d) desired end results.

Let’s look at how an evaluator can use program theory to guide an evaluation. Figure 3-3 presents a program of teacher visits to students’ homes (Weiss, 1972). The intent of the program is to improve students’ achievement in reading. The home

visits may serve several functions. One assumption may be that teachers provide the parents with knowledge about the school’s expectations for their child’s performance (attendance, homework, and so on). As a consequence of such knowledge, parents may monitor the student’s attendance and help him do assigned work more conscientiously. The student will then do better work and his reading achievement will improve. Another theory could be that the visits give the teacher the opportunity to understand parents and the culture of the home. With better knowledge and understanding, teachers have greater sympathy with student’s worldview and adapt lessons to fit. Students feel appreciated and respected and their reading achievement improves. A third theory would hold that teachers and parents share knowledge about any special problems that impede the student’s achievement, such as dyslexia, emotional problems, or poor eyesight. With this conversation, parents and teachers cooperate to see that the student receives the necessary assistance, and when the problem is ameliorated, the student’s achievement increases.
How did I arrive at these theories? I started with the program of teacher visits and the expected outcome of improved student reading performance. Then I filled in the space between with guesses about how one might get from (a) to (d). I used a single outcome, but one can easily use two or more outcomes. (It might take a separate analysis for each one.) As more evaluations develop and test program theories, the hope is that less guesswork will be involved and more knowledge, at least provisional knowledge.

One can also include contingencies in the theory—that is, conditions under which one thing will happen and conditions under which something else will happen. For example, contingencies might come into play in the job training program in Figure 3-2. In communities where the economy is good and jobs are available, the theory might state that trainees go out looking for jobs and find them. However, in communities in recession or with high unemployment, trainees seek jobs but do not necessarily get them.

Similarly, one can envision a theory that makes distinctions among subgroups of people. For example, delinquents assigned to a community service program may respond to the program differently depending on whether they are first-time offenders or repeaters. First-time offenders may absorb the lessons of the program and view their work in the community as atonement for violation of the law. Delinquents with a longer record may regard their hours of community service as punishment and scoff at efforts to redirect their beliefs and values. This illustration suggests the possibilities of elaborating program theories as the evaluator and program staff gain experience, knowledge, and confidence.

Again, it is useful to posit some theories that produce unintended effects, or in the literal words of a cultural anthropologist, "seek to distinguish actors’ purposes from the unsought entailments of their acts" (Barth, 1990: 650). Discussion about what might go astray can be useful on several counts. It can help program staff try to foresee unwanted side effects and guard against them, and it can call such possible effects to the attention of the evaluator, so that she can study them. If there is a likelihood that negative effects can ensue, the evaluator can trace them with the same care devoted to the chains of desirable effects.

The evaluator uses the diagram in Figure 3-3 to decide what data to collect. If she pursues all the chains, she will need data on the subject of the conversations between teachers and parents. She will also need to collect data before and after the home visits about parents’ knowledge of school expectations, parental support and encouragement for the student, the conscientiousness of student work, pupil attendance, teachers’ knowledge of students’ home culture, teachers’ style of teaching, and pupils’ feelings of being heard and respected. After the visits, the evaluator will have to find out about the identification of any physical or emotional problem, referrals for care, help received, and improvement of the problem.

Suppose that after home visits, student reading achievement improves significantly more than that of a control group of students who did not have the visits. The usual conclusion would be that the visits (and all the assumptions and expectations surrounding them) were justified. But suppose also that records of teachers’ visits showed that all the measures of interaction, communication, and understanding between parent and teacher were at very low levels. Teachers and parents were really not communicating. The source of the student’s improved reading ability must be sought elsewhere. (One explanation might be that the students misperceived the intent of the teachers’ visits; they may have thought teachers were complaining about their work and trying to persuade parents to punish poor achievers. Improved schoolwork could have been a response to the perceived threat.)

If the predicted sequence of events does not work out, further investigation is needed. But when the theory proves to be a reasonable picture of how things happen, it gives some notion of the reasons why. However, even with the best and most supportive data, theories are never proved. At best, they are not disconfirmed by the data. There may be alternative theories that would provide equally plausible or better interpretations of the available facts. Scientific generalizations are built up by developing hypotheses and then submitting them to successive tests in an effort to disprove them or find the limits of their applicability.

In undertaking theory-based evaluation, it is probably best to start with a relatively simple theory. In the evaluation of a community education program to reduce cardiovascular disease (Farquhar et al., 1990), the focus was on reducing cholesterol intake and smoking. One theory that underlay the program was that intensive communication would increase residents’ knowledge of risk factors, and increased knowledge about risk factors would lead to changes in behavior (such as reduced intake of cholesterol). With more healthful behaviors, residents would improve in blood pressure, weight, and pulse rate.

So one part of the theory was that communication leads to knowledge, which leads to behavior change. That is the theory of many programs. Sometimes it is assumed that communication leads to knowledge, which leads to attitude change, which leads to behavior change. The evaluator can elaborate the theory with more detail and design the evaluation to investigate each of the several links. The important point is that the theory gives guidance to the evaluation about where to look and what to find out.

When a program is very complex and multi-faceted, like the introduction of case managers to integrate the delivery of social services to multi-problem families, the evaluator may find it impossible to follow every branch of every subtheory underlying the program. Sheer limitations on time and resources will prevent an effort to test each link. Choices will have to be made about which causal pathways to pursue. How should an evaluator decide?

Possible criteria are the following:

1. Uncertainty. Some linkages between program service and client progress are more problematic than others. Sometimes no knowledge exists on the viability of the assumptions. Sometimes available knowledge from social science research or previous evaluations suggests that the assumption is unwarranted. This may be a good set of linkages for the evaluator to study, preferably early in the evaluation. If it does not hold up, the program can find out about failings quickly and modify what it is doing.
Comparing Program Theory to Actual Developments

Once the data are in hand, the intent of the analysis is to see how well the theory describes what actually happened. Which parts of the theory are supported, and which parts did not turn out as expected? This is important information for the directors and staff of the program under study, so that they can rethink the understandings and replan the activities that did not work out as anticipated. It is important information for people engaged in other programs so that they can take stock of what they are doing in light of the theoretical understandings generated. On a broader scale, it is important for the growth of social knowledge. If, for example, evaluations of programs in many different fields show that the supposed link between change in knowledge and change in behavior is weak or nonexistent, then policymakers and program designers have to develop other theories (and the activities to implement them) when they seek to bring about behavior change. For social scientists, too, this kind of knowledge is grist to the mill of better understanding of human behavior.

The mere construction of such a theory can expose naïve and simplistic expectations. Do program developers, for example, believe that those parents who do not have the values, the background experience, or the skills to help their children with academic studies can be persuaded to do so by one or two visits from a teacher? Certainly, more must be built into a program with such great expectations—training for the teachers, for one thing, and reinforcements, rewards, and possibly skill training for the parents. The evaluation theory can be a learning tool long before the evaluation begins.

During the conduct of the home visit program, it may turn out that few teachers actually make the home visits. They may try and fail to find parents at home, or they may shy away from the effort entirely. In that case, it makes little sense to continue to study the effects of what is now known to be a phantom program. Or evidence may show that teachers make the visits, and the evaluator follows along the hypothesized chain of events. She finds out teachers do indeed show enhanced understanding of the culture of students' home, they have more sympathy with the students, but they continue teaching in the old traditional ways without attempting to suit instruction to the pupils' subculture. If the students' reading performance does not improve, we have some clue about the why of it. We can tell where the project's chain has foundered. Similarly, if some parents do learn about the school's expectations for their children's behavior and performance and do try to encourage them to do better homework and schoolwork, yet their children's work is no better done than that of other students, we have a place to look for further insight into the breakdown of the expected chain of events.

How does one analyze the viability of program theory? Writers on evaluation have offered several methods of statistical analysis (Chen & Rossi, 1987; Lipsey, 1989). But even without sophisticated statistical techniques, it is possible to learn a great deal through what Campbell (1966) has called "pattern matching." As the words imply, the evaluator compares the expectations generated by theory with empirical data to see how well they fit. Riggan's (1990) evaluation of a mandatory job search program for welfare recipients shows a simple manner in which evaluators can use program theory as a basis for recommendations to program managers. The study found that the program wasn't doing what its own theory demanded; its implementation theory was not being realized. The evaluators' recommendations centered on bringing the program into compliance with its theory—for example, applying sanctions where clients were not following the rules and hiring more counselors to perform necessary activities going unfulfilled.

Advantages of Explicating Theories of Change

Mapping the theories of change inherent in a program has advantages not just for evaluators but also for the other players in the program world.

Program Designers

During the period when the program is first being developed, program designers can profit from the disciplined thinking that this approach stimulates. The theories-of-change model forces them to make their assumptions explicit, and in doing so, it gives them an opportunity to examine the logic of their ideas. Are they making too many leaps of faith? In Figure 3-1, is it reasonable, based on past experience and research findings, to expect that a few thousand dollars of additional pay will lead teachers to transform their styles of teaching? Thinking through the logic of their expectations may help program designers to consider more powerful or comprehensive interventions to achieve the goals they seek, or it may lead them to scale back their expectations for what they can accomplish with the means at hand. It might lead them to call on different categories of personnel or different organizational structures. In short, by asking them to make implicit assumptions explicit, it encourages them to think harder and deeper about the programs they design.

Practitioners

During the implementation of programs, practitioners, too, are asked to confront their theories about how the program will lead to change. It may turn out in conversation that members of the same staff have different ideas about how the program
will achieve its effects. If that is the case, they may well be working at cross-purposes. If practitioners can work through their differences and agree on a common set of assumptions about what they are doing and why, they can increase the force of the intervention.

Also, practitioners can profit in much the same way as program designers. They are in a sense redesigning the program each day by what they do and how they do it. When they surface their assumptions about how the program will work, they have to face their weak or questionable premises and the leaps of faith embedded in their expectations. This confrontation can help them to improve not only their theories and plans but also their regular practice.

Program Managers
A major advantage for program managers is that the evaluation provides feedback about which chain of reasoning breaks down and where it breaks down. Do trainees who are exposed to the training program fail to attend sessions regularly, or do they attend but fail to learn the work skills, or do they learn the skills but fail to look for a job? This kind of information helps managers determine what specifically needs to be fixed. If the evaluation shows that one whole line of theory receives little support in the data, the manager can reconsider the utility of efforts devoted to this line of work.

Managers and Funders of Similar Programs Elsewhere
Information about the theoretical underpinnings of the program has special importance to those who are running or considering the start of similar programs. Not only do they want to know whether such programs are likely to work, but they also want to know how they work. If they understand the what, how, and why of program success (and failure), they can undertake new ventures better prepared to replicate those elements of the program that are associated with successful progression and to rethink and rework those program elements that do not lead on to the next stage.

Policymakers and the Public
Another side benefit of theory-based evaluation is that it provides explanations, stories of means and ends, that communicate readily to policymakers and the public. Its findings can be encapsulated in morals, like Aesop’s fables: Teachers who earn higher salaries feel better appreciated, and they work harder at teaching. These kinds of explanatory stories are often more convincing and memorable than are statistical findings alone. They may stand a better chance of influencing the course of future policy.

Of course, there are limitations in the use of programmatic theories of change as a basis for evaluation. One is that it is data greedy. Because theory-based evaluation seeks to follow many steps of the assumed sequence of events, it demands large quantities of data. Heavy demand for data and, later on, for analysis of the data makes claims not only on the evaluator but also on evaluation resources and on the people in the program who must supply the data. Another limitation is that while some of the data called for may well be quantitative in nature, some will probably be narrative and qualitative. The combination calls for ingenuity in analytic strategy and places additional burdens on the evaluator.

Still, the definition of a program’s theories of change is worth attention. Looking at program theories is not the only way to go about deciding on the evaluation focus, but it is a good way to clarify and systematize the factors that are worth examining.

Critiques of Program Theory
Some observers question whether an emphasis on program theory will advance the state of the evaluation art. They doubt that people are good at identifying the causal processes at work (Shadish, Cook, & Leviton, 1991). They are skeptical that program practitioners have any special insight into the processes through which change occurs, and they have little more faith that social scientific knowledge is up to the task. Nor do critics believe that theory-based evaluation will necessarily contribute to a growth in program knowledge. They say that while evaluations of one project at a time can yield information specific to each project, it is not clear that findings will generalize beyond the project studied. They doubt that evaluators will be able to distinguish between program features that are common across a family of programs and those that are unique to the local setting. Programs are too complex to be decomposed into a finite set of variables. Therefore, there is little chance for cumulation of findings.

Many qualitative researchers insist that all truths are local and contingent. Each program is its own world and has to be understood in its own terms. The evaluator, in this argument, has to work within the uniqueness of setting, participants, program operations, and time, and these multiple idiosyncrasies preclude drawing generalizations. They say that efforts to generalize about complex undertakings like social programs are not likely to succeed at any high level of abstraction.

I agree that hopes for accumulation of knowledge through repeated tests of program theory are optimistic. For such a desirable end to occur in a reasonable time span, we should start with strong theory at the outset. Unfortunately, program theory in many fields is at a low stage of development. In those fields, the theories that evaluations are likely to be testing are low-level approximations, riddled with inaccuracies and false paths. To discover their flaws and painstakingly revise and improve them will take multiple iterations. This is an arduous way to generate generalizable knowledge.

But individual studies can be pooled with hundreds of other studies of programs of like kind through meta-analysis. Through analysis of the whole body of studies, it may be possible to discover which intervening processes are most strongly associated with achievement of desired outcomes. Another hope for understanding causal theory is that some effects emerge from particular types of program operations consistently enough to be dependable, even if a larger explanatory system is not fully known (Cook & Shadish, 1994).

Program-theory approaches have the advantage of breaking program experience into a series of small chunks— that is, the links between one step and the next. Because the evaluation focuses on these links (rather than just the global assumption that the program leads to desired outcomes), it can examine linkages across a range of program and implementations. For example, many programs rely on case man-
agers to coordinate the services that a family receives from numerous care providers. Common assumptions are that the case manager increases (a) the number and (b) the appropriateness of services the family receives. Theory-based evaluations can test those assumptions in a variety of settings and thus contribute to generalizations about what case managers do and do not accomplish.

Furthermore, the local reasons for basing evaluation on program theory remain compelling. For the evaluator studying a particular program or project, unearthing (or constructing) program theories is a helpful way to figure out what to study, how to allocate evaluation resources, where to concentrate data collection, and how to make sense of findings. And it just may happen that evaluators will learn things of use to others.

With all their complexities and quirks, programs display discernible commonalities. Just as the basic social sciences find patterns of behavior that hold across individuals, organizations, polities, and societies, it should be possible to find features and relationships that characterize programs of various kinds. Human behavior is complex, but it is not random. We ought to be able to make headway in understanding what happens under particular sets of circumstances when interventions work to redress social problems.

The search for better program theory is well worth the pursuit. Given the sorry state of social intervention in some fields and the lack of explicit theorizing to guide program development, even a modest advance would mark significant improvement.

**Summary**

This chapter has highlighted the importance of understanding the program that is being evaluated. It has pointed to several sources of knowledge about the program, including written information, people, and direct observation. The chapter has stressed the value of reviewing past evaluations of the same or similar programs as a way for the evaluator to orient herself to activities, expectations, past outcomes, and explanations for achievements and shortfalls. As a side benefit, past evaluations also offer useful insights into the research methods that other evaluators have used, the instruments they relied on for collecting data, the modes of analysis they employed, and their occasional suggestions for how to do the study better the next time around.

A good method for formalizing knowledge about the program is to construct an outline of the program's underlying theory (or theories). The evaluator should draw on the implicit and explicit knowledge of program personnel in explicating program theory. A statement of theory should lay out the sequence of assumptions that show how program inputs (staff, resources, activities) translate through a series of intermediate steps to desired program outcomes (improvements in people, organizations, or communities). The microsteps of the theory can then become the framework for the evaluation study. The evaluation tracks developments to find out whether the assumed linkages in fact occur.

By using program theory as one basis for structuring the evaluation, the study allows the evaluator to reach conclusions not only about the processes by which services were delivered and the outcomes that they achieved, but also about the valid-