12

Evaluation

Evaluation is a kind of policy research. It shares some fundamental features with social impact assessment. First, both are concerned with the impact or effects of different actions on people. Second, both can make use of the same kinds of research methods and techniques. But the two kinds of research are different in certain important ways. SIA is primarily concerned with discovering *before* the fact any costly unintended effects of an activity. For example, governments build dams to impound water so that floods are reduced, agricultural production is increased, or recreational opportunities can be developed. In this situation an SIA might be done to predict whether this would have adverse effects on nearby communities. The purpose of dams is not to displace communities, so it is important to planners to identify all the effects.

Evaluation is most often concerned with determining *after* the fact whether the intended benefits of an activity occurred, or alternatively discovering whether a project with intended benefits is working. For example, an agency might establish a program to increase employment of high school dropouts, and then do research to determine whether the dropouts became employed. In addition, evaluation can be used to examine program operations as well as program effects.

INTEGRATED RESEARCH METHODOLOGY

Our treatment of evaluation will focus primarily on research design. The discussion will start with consideration of classical experimental design. From this base we will consider a number of research design alternatives. Our perspective will be that of general social science as much as anthropology. The specific uses of ethnography in evaluation will be considered. It is important to note that contemporary evaluation research makes use of many different research strategies and that these methods and techniques are used by evaluators regardless of the discipline they were trained in. One is likely to pick up an ethnographically

oriented evaluation research report, based on participant-observation, and find that no anthropologists were involved in the study. And, of course, the reverse is true: it is possible to find anthropologists involved in executing pre-test, post-test, control group experimental designs. Ethnographic practice is one tool, a very useful tool, but only one tool. The implications are clear—evaluation researchers need to know a number of different designs whether they are anthropologists, sociologists, psychologists, political scientists, economists, or other kinds of researchers.

This statement on evaluation takes an integrated research methodology approach, which may combine qualitative and quantitative research (Cook and Reichardt 1979). The integrated research methodology approach requires that we control a variety of research designs and data collection techniques. This implies the possession of the necessary technical skill to process and analyze the data derived through a variety of techniques. The integrated research methodology approach means carefully identifying which research data collection technique is required to solve a research problem. What works? What carries us the furthest in understanding? What is efficient? What research technique is the most credible? These questions represent some of the dimensions that can be considered when we make a judgment about the research approach that will be used.

The criteria that we use to judge which design and technique we will use are quite broad. Of course, basic notions of validity and reliability are among the most important. Another important consideration is cost, both in money and time. The best design in the world is worthless if one can not afford to implement it. Ethnographic practice sometimes involves a great deal of time spent in research. Yet, a good ethnographer can learn a lot about what is going on in a situation by interviewing one person. In any case, there is a large range of legitimate concerns in terms of research design and technique selection. A final, important question is, Does the researcher have the skill to do the task?

Anthropologists can be involved in evaluation in two different ways. First, and most important, is as a broadly trained social scientist who is prepared to do a variety of evaluation research tasks as needed. The second is as a specialized evaluation ethnographer, who contributes to evaluation through nonexperimental, unobtrusive, qualitative, and participatory research techniques as these skills are needed. In this second role, the anthropologist may also be valued because he or she has knowledge of the group within which the evaluation is taking place, in addition to knowledge of technique. While ethnographic skills are very useful in the evaluation process, as we will discuss later, the most promising approach is the first. There are opportunities for professional, ethnographically oriented evaluators, it is just that pathways for career development are more open to people with somewhat broader technical training in research design and data analysis. M. G. Trend speaks of ethnographic evaluators being stuck at the lower levels of large evaluation programs, being relegated to a "go-fer" role with

lower pay and less job security (1976). For more discussion of these issues see Chapter 14, "Making a Living."

One implication of the view expressed above is that there is not an anthropological way of evaluating. It is more useful to think of a multifaceted and loosely structured social science of evaluation in which individual problems in evaluation are addressed using a variety of techniques. Anthropologists can do their job better if they have control of a variety of techniques. The task is not to mimic sociology or psychology, but to participate in a larger contemporary tradition in social or behavioral science. The effect of this on anthropology will be positive.

This chapter will have three parts: (1) a discussion of the evaluation process, with description of alternatives for evaluation design; (2) a discussion of differing perspectives on the role of evaluation; and (3) case studies in evaluation done by anthropologists. The cases used are evaluations of the Administrative Agency Experiment of the Department of Housing and Urban Development, the Tumaco (Colombia) Health Project of Foster Parents Plan International, the Job Shop/Job Club Program of the Community Education Center of West Philadelphia, and a needs assessment done for the United Way of Saskatoon.

EVALUATION PROCESS

Because this single-chapter treatment of evaluation is necessarily brief, our presentation of evaluation may make the field seem much more orderly than it really is. Evaluation encompasses all the disarray that you would expect in a relatively young field in which persons of many disciplines participate. This situation has been exacerbated by the fact that important segments of the field, most notably educational evaluation, have undergone rapid growth forced by huge federal subsidies. For example, the Elementary and Secondary Education Act of 1965 carried a provision that educators receiving grants in support of education programs had to submit evaluation reports that identified the effects of the program. Basically, a lot happened quickly, and there has been little synthesis. There are a number of competing viewpoints and substantial semantic difficulties. Now then, what is evaluation?

At its very core evaluation is what the dictionary says it is, that is, "the determination of the worth of something. While all of us are constantly evaluating things, activities, and ideas, evaluation in a technical sense requires much more than the casually subjective, and largely private, assessments of worth that we produce everyday. Let us consider evaluation as a process.

First, a general technical definition: "Evaluation is the determination of the worth of a thing. It includes obtaining information for use in judging the worth of a program, product, procedure, or objective, or the potential utility of alternative approaches designed to attain specified objectives" (Worthen and Sanders 1973:19). When evaluation is done, it is almost always done in reference to

activity that is intended to affect people in one way or another. Evaluation can be used to determine worth in both negative and positive aspects. While many research designs used in evaluation stress the determination of whether planned objectives were accomplished, evaluation can also be used to discover unintended consequences of programs and projects. The activities evaluated are always motivated by some desired end state.

The evaluation process is a process by which values are rationalized. The idea of treatment borrowed from the literature on experimental design is useful as a label for the actions, projects, programs, and so forth that are carried out to achieve goals. Other dimensions to the evaluation process are the nature and characteristics of both the agency providing the service and the individuals and groups that are the focus of the agency.

At a general level there are three types of evaluations:

I. Effects studies—The basic task here is the determination of whether a program (or other entity) is achieving its goals. This is the classic evaluation task. It has also been referred to as product evaluation (Stufflebeam 1973) or outcome evaluation. Effects studies done during the life of a program that are intended to inform program managers or sponsors about program operations can be thought of as process evaluations (Stufflebeam 1973). Effects studies can be directed at the dissemination of practices to other settings to help guide decisions about continuance, enhancement, curtailment, and modification.

II. Process studies—The basic task here is to determine how a program is operating. This is a managerial task. This type of evaluation is also called operations analysis (Riecken 1972). Both process and effects studies may be designed in the same way. Process evaluations may consist of long-term program monitoring.

III. Needs assessment—The basic task here is to determine the needs of a potentially served population (McKillip 1987; Neuber 1980; Scriven and Roth 1978). One could include needs assessment in a discussion of planning. Needs assessment can also occur during the life of a program so as to allow program redefining. That is, it can be part of program planning and management. Needs assessment can be ongoing.

This general typology implies a number of dimensions. These include: the purpose or role of the evaluation, the timing of the research, and to some indirect extent, the research design. There are a number of very useful discussions in the evaluation literature that address these dimensions. Let us focus on design.

Research design is what is unique about evaluation research when compared to other types of research. Measurement and data analysis techniques are quite comparable in evaluation and basic research. Let us start our discussion of design by considering the classical design pattern and then expanding from that base.

Carol H. Weiss depicts the "traditional formulation" of evaluation research in the following way (1972:6):

1. Finding out the goals of the program;
2. Translating the goals into measurable indicators of goal achievement;

3. Collecting data on the indicators for those who have been exposed to the program;
4. Collecting similar data on an equivalent group that has not been exposed to the program (control group);
5. Comparing the data on program participants and controls in terms of goal criteria.

This is, of course, a generalized version of experimental design used in behavioral science. An understanding of this basic pattern can be supplemented by reading *Experimental and Quasi-experimental Designs for Research* by Donald T. Campbell and J. C. Stanley (1965); *Research Design in Anthropology: Paradigms and Paradigmatics in the Testing of Hypotheses* by John A. Brim and David H. Spain (1974); or *Quasi-experimentation: Design and Analysis Issues for Field Settings*, by Thomas D. Cook and Donald T. Campbell (1979). These volumes very clearly lay out research design alternatives.

Campbell and Stanley (1965) define five different research designs. The list starts with the least rigorous, the one-shot case study. In a one-shot case study of a program one would measure the effects of the program only once, after the research subjects had participated in the program. What is absent from this research is a baseline measurement, called a pre-test, to determine the "before method." One has to assume a great deal about the program participants' prior condition." One has to assume a great deal about the program participants' prior state. Some researchers attempt to strengthen the one-shot case study design by using documentary evidence or reconstructions based on self-reports. At worst, one-shot case studies take the form of program-serving testimonials. Fortunately, there are many circumstances where the one-shot case study can be valuable, because the design is very common. Much ethnographic evaluation takes the form of a one-shot case study. All the examples discussed later in this chapter are of this type.

An enhancement of the one-shot case study is the one-group, pre-test, post-test design. The addition of the pre-test allows one to measure change more objectively. The pre-test, however, does not allow one conclusively to attribute the change to the program. Change can occur because of other events, such as normal change through time, the pre-test's effects, ineffective measurement, and participant fatigue, as well as other factors.

A third type of research design is the static group comparison, which adds a control group to the one-shot case study. In this method, a group that has experienced the program is compared with a group that has not experienced it. The weakness is that the design does not allow certainty about the differences between the groups before the treatment. It is possible to strengthen this design through matching of participants and the use of retrospective measures.

The fourth type is the pre-test, post-test, control group design, which involves setting up two groups before the program. In terms of research design quality, the best way of doing this is by random assignment. There is a before and after measurement to determine effects of the program. This design, although very good, does not control for the effects of the research procedure. This can be controlled for through the use of the final research design, the Solomon four-

231group design, in which the effects of the research are examined along with the effects of the program.

These design alternatives do not address issues relating to needs assessment, but are applicable to many contexts in both effect and process studies. The problems with needs assessment mostly revolve around having the research sample be consistent with the group that actually receives the service after program implementation.

While these are standard designs, and each represents an incremental increase in the capacity to specify cause, there are costs associated with increases in experimental control. This is one reason why much evaluation is carried out using the one-shot case study design. The more complex, error-reducing designs are used quite extensively in education to evaluate curricula in anticipation of wider use. It is clear why; the treatments are usually more readily definable and control groups are easier to find and match. Treatments will consist of a set of test materials administered in an adequately standardized way, and, if you need a control group, other classes of student "subjects" are available. Similar patterns occur in the evaluation of drug treatment programs, in which the participants have diminished control because they are in the program by order of the courts.

Outside of certain specialized areas, it may be very difficult to apply the more complex research designs. While there are many statements in the literature on evaluation that present the more complex designs as ideals, it is very important to view these designs as alternatives to be selected for application as appropriate. Selection should always be based on the most appropriate design, not the most elegant one.

A very large array of factors needs to be considered in evaluation planning, in addition to basic research design. Perhaps most important is the intended purpose of the evaluation. Some researchers may place too high a value on the elegance of their design, and too low a value on assisting the program to serve its clients better. Anthropologists in evaluation seem to be more committed to clients' needs than others. Other factors to consider are cost, available time, and the nature of the service population.

Soft Designs

As we know, in many circumstances soft and fuzzy is good. Much of the research methodology literature is geared toward an idealization of hard and definite. This is changing somewhat as researchers become disenchanted with operationalism. Yet one still finds defensiveness on the part of soft methodologists. Hard versus soft is not the same as good versus bad. Both approaches are subject to their own problems of quality. By soft designs we mean, among other things, research that stresses qualitative methods, naturalistic observation, discovery, induction, and holism. By hard designs we mean, among other things, quantitative methods, structured observation, verification, deduction, and particularism (Cook and Reichardt 1979:10). For example, ethnography, with its

emphasis on key informants and participant-observation, tends toward softness; survey research, with its emphasis on randomly selected subjects and instrumented observation (i.e., questionnaires) tends toward hardness. Again, we are not arguing for anything other than the selection of appropriate methods. When are soft methods appropriate?

Soft methods are useful because they are often less of a burden for the program staff. The more structured the research design is, the greater the chance of the evaluation interfering with program functioning. It is very difficult to burden useful programs with certain kinds of highly structured research. The selection of softer designs is called for where obtrusiveness is an issue. Soft methods are useful where program goals are less well-defined, or are especially complex and diverse. Soft approaches are really useful for discovery. Ethnographers seem to do research more to raise questions than to answer them. Soft methods are often the only way to realistically handle complex situations. The more structured the research design, the fewer variables it can consider. Program goals are often not very well defined. Soft techniques can be fit into ongoing program development better than hard approaches. The before and after measures specified in the experimental designs can be replaced with during-during-during measures, which are more workable with softer techniques.

Softer methods often prepare the way for implementation of results better than hard methods because the researchers often end up with an excellent understanding of the persons managing programs and the constraints under which they must operate. In fact, it may be best to have the evaluation include continual feedback to the program with correction. This kind of arrangement is unworkable with hard designs because it interferes with the outcome of the research. Further, hard research designs assume too much about the stability of programs while they are being evaluated. Mid-study change in program administration disrupts the hard studies, but for the soft designs this kind of activity simply represents more data relevant to the program. We might say that soft techniques are useful in rapidly changing circumstances.

Hard Designs

Hard research designs are especially appropriate if the program has clear-cut, measurable objectives that are identified within the program. Hard designs are appropriate where program staff are familiar with, and value, research along with their commitment to service. This orientation is appropriate to situations where control groups are readily available. The idea of the control group is sometimes antithetical to the service orientation of program administrators. In some circumstances, establishing control groups requires one to deny access. Hard approaches work well where there is relative program stability and a lower expectation or need for mid-course feedback. A useful application of hard approaches is in the production of the final evaluation of demonstration projects, with the goal being to inform potential adopters of the program.

Both hard or soft techniques must be executed in a way that allows their results to be applied to real-world decision making. The inflexibility of the hard techniques can relate to less timeliness—that is, the research has to run its prescribed course. Soft techniques, on the other hand, can be so unfocused that the researcher meanders through a program without attending to the needed research issues. It is important to remember that the primary use of evaluation is to provide information for decision making. Late research is bad research.

PERSPECTIVES ON THE ROLE OF EVALUATION

Evaluation has a number of different roles, both legitimate and illegitimate. A useful understanding of the roles of evaluation can be derived from the ideas of Michael Scriven and Daniel L. Stufflebeam. Both have developed concepts of evaluation that can serve to direct it toward greater utility.

In a very important and frequently reprinted article entitled "The Methodology of Evaluation" (1973), Michael Scriven provides a number of concepts useful for thinking about the role of evaluation. Although his discussion is focused on evaluation in education, with emphasis on the evaluation of new teaching methods and curricula, his ideas are very widely applicable. As he notes, the roles of evaluation can be quite variable. All these roles relate to the primary goal of evaluation: to determine worth. The role of evaluation in some ways structures the evaluation itself.

Scriven conceives of two types of evaluation research: formative evaluation and summative evaluation. Formative evaluation is carried out in the course of a project, with the goal of improving project functions or products. The evaluation may be done by an outside consultant, but the information produced by the evaluation is for the use of the agency. As Scriven notes, "the evaluation feedback loop stays *within* the developmental agency (its consultants), and serves to improve the product" (1973:62). Formative evaluation is conceptualized as a mid-term outcome study of the product or effects of the program, rather than a more general kind of process study, which might answer the question, What is going on here?

Summative evaluation serves to determine worth at the end of the process and is intended to go outside the agency whose work is being evaluated. The evaluation serves to increase utilization and recognition of the project. According to Scriven, program monitoring is a hybrid type of summative evaluation in that it is intended to go outside the agency being evaluated, but at an intermediate time.

Both formative and summative evaluation can make use of the same research designs. However, because of their different roles they require different communication strategies. The essence of the formative-summative contrast rests in the direction and purpose of the communication of evaluation results. Scriven also contrasts what he calls *intrinsic* and *pay-off* evaluation. Intrinsic evaluation evaluates the content of the project's product or treatment, whereas pay-off

evaluation is focused on effects. These four concepts—formative versus summative, intrinsic versus pay-off—are useful because they focus the evaluation on a specific purpose.

Scriven's ideas make us sensitive to the various roles of evaluation; Stufflebeam's work models an entire process of evaluation. His work, also developed in the context of educational evaluation, rests on the assumption that evaluation is done to aid decision making. The information that it provides should be useful to decision makers. Evaluation is a continuing process and is best organized in coordination with implementation. Data collection needs to be consciously targeted on decision-making needs. The total evaluation process ultimately involves collaboration between evaluator and decision maker.

This view of evaluation is integrated by Stufflebeam into a comprehensive process referred to as the CIPP evaluation model. The model specifies different kinds of evaluation, which serve various purposes and inform various types of decisions. These decision types are: (1) "*planning decisions* to determine objectives"; (2) "*structuring decisions* to design procedures"; (3) "*recycling decisions* to judge and react to attainments"; and (4) "*implementing decisions* to utilize, control and refine procedures" (Stufflebeam 1973:133). The four types of decisions are served by four types of evaluation. These are *context evaluation, input evaluation, product evaluation,* and *process evaluation.*

Context evaluation supports planning decisions. This category would include what is called needs assessment by others, but would also identify resources that are not being used, and constraints that affect needs. Products of context evaluation include identification of the client population, and of general goals and objectives. Input evaluation supports structuring decisions. The important task here is the identification of resources that relate to project objectives. Part of the process involves the determination of current agency capability. Also included is the identification of alternate strategies for accomplishing objectives. Input evaluation also involves costing out alternatives. Process evaluation supports implementing decisions. This type of evaluation is used to find defects in procedures and implementation, to inform ongoing decisions, and to document activities of the program. Product evaluation informs recycling decisions. The task here is to evaluate project accomplishments at various points in the life of the project. Product evaluation requires operational definition of objectives, development of a measurement strategy, and standards against which measurements are compared.

CASE STUDIES: FOUR EVALUATIONS

These four projects were selected to illustrate evaluation in different settings, all executed by anthropologists. The fact that all four evaluations are case studies is not an accident. Evaluation anthropologists rarely use control groups. While the evaluations are all case studies, they are quite different. They differ in terms of how they structure analytical comparisons. This is especially apparent in the

first case, on housing and the third case, on employment. The second case, on health care, is primarily concerned with documenting project completion. The fourth case is a needs assessment for a community service agency.

Evaluating Housing Policy Innovations

The Administrative Agency Experiment (AAE) was one of three social experiments initiated by the U.S. Department of Housing and Urban Development (HUD) in the early 1970s to test the use of direct-cash housing allowance payments to assist low-income families improve the quality of their housing through the open market (Trend 1978a). Housing vouchers had been considered as an alternative to subsidized public housing by politicians and policy makers for a considerable amount of time. This program attempted to test the workability of this approach and to develop effective management practices. Most other programs for improving housing for the poor presented beneficiaries with limited choice. The AAE payments were made directly to the families, with the stipulation that they had to spend the money on housing.

The three experiments tested different issues. In addition to the AAE, experiments were designed that tested the effect of direct payments on housing supply and demand. The AAE tested various approaches for management of such a social policy. The AAE was implemented by eight public agencies in different localities throughout the country. The agencies were given substantial control over administrative procedures and implementation at their site. Each local project was allowed to enroll up to nine hundred families for the two-year period of the experiment. The sites were in Springfield, Massachusetts; Jacksonville, Florida; Durham, North Carolina; Tulsa, Oklahoma; San Bernardino, California; Portland, Oregon; Peoria, Illinois; and Bismarck, North Dakota.

The evaluation was carried out by Abt Associates, Incorporated, a large research consulting firm located in Cambridge, Massachusetts. The project made extensive use of anthropologists, especially for on-site observation, but also in subsequent analysis. Most of those employed were fresh out of graduate school.

The AAE was intended to provide information on different methods for administering a housing allowance program. The evaluation task was complex and included documenting both effects and administrative procedures. The AAE was implemented by four different types of organizations: local housing authorities, metropolitan government agencies, state community development agencies, and welfare agencies. HUD provided fewer administrative requirements for implementation than normal. They included certain eligibility rules, payment formulas, a locally defined housing standard requirement, restriction to rental property, a lease requirement, a standard payment plan, and certain reporting procedures. Beyond this, variation was encouraged. Agencies had flexibility in various administrative functions. These included outreach methods, enrollment procedures, certification procedures, payment practices, inspection standards, community

relations activities, and overall program management. Variation in the form and effectiveness of these different functions was an important focus of the evaluation task. The local agencies generated the program features that were to be evaluated.

The AAE was a naturalistic experiment and therefore did not test hypotheses. The evaluation involved after-the-fact determination of success, with emphasis on documenting and evaluating the various solutions to the problems of implementation. Thus the evaluation questions were determined by the design decisions that were made as the programs developed in the eight sites.

The evaluation was based on data from a variety of sources; most important were the on-site observers. These observers spent a year documenting administrative agency activities. While these observations focused on the agency as a whole, the observations and field notes they produced were referenced to fourteen functions, such as outreach, certification, and inspection, as mentioned above. This allowed cross-project comparison. When variations between agencies were apparent, these variations were examined to determine relative effectiveness and impact. These assessments were addressed in reports that compared projects across functions. The reports provided information that facilitated knowledge transfer to other similar projects. This is the summative evaluation function at work.

The evaluation design required that work be done toward two kinds of research objectives. These included data collection that focused on operational measures, such as numbers of enrollees and payment levels, as well as qualitative assessments of hard-to-measure features, such as fairness and dignity of treatment of participants.

Some measures of function were defined early in the experiment and collected through the use of program documents. Background data on participants and the local community was collected throughout the project. In addition to the information on cross-agency differences, information on common features was also collected. In this frame, outcome data was compared. In addition, the evaluation process also produced separate case studies in reference to specific functions and agency programs that were problematic. Many of these case studies were based on observational data.

Six types of data were collected and used: direct observation of agency activities; surveys of participating households; data from administrative agency forms on participants; agency management reports; site environmental data; and samplings of administrative records. The full-time, on-site observers produced field notes that were formatted to allow comparison between programs categorized by the previously mentioned functions. These materials were used for various purposes, but were especially useful for interpretation. The survey component collected information on income, household composition, and attitudes toward the program on the part of participants. In addition, the time of project employees was charged to the various function areas. All in all, the amount of data collected was immense. For example, over twenty-five thousand pages of field notes were accumulated.

The analysis attempted to assess the overall utility of the direct payment approach to meeting housing needs, while making recommendations as to how direct payment might best be done. The variation-focused evaluation helped identify what were thought to be more effective administrative practices. For example, the Bismarck agency found that they obtained reliable certification results with simpler income determination procedures than the other agencies. The Jacksonville agency evaluation addressed the problems of making the approach work in a tight housing market.

Perhaps most important, the experiment demonstrated that poor people could effectively operate in the open housing market in terms of their own choice. The project also showed that the allowance program did not raise housing costs in these markets. On the negative side, the AAE did not reduce housing segregation significantly. The evaluators concluded that AAE was a useful tool in dealing with the housing problems of poor people from a policy perspective. Ultimately, many of the approaches demonstrated in the AAE came to be implemented in federal housing law.

Evaluating a Health Project Sponsored by a Private Voluntary Organization

Foster Parents Plan International (PLAN) is a private organization that sponsors community development projects in twenty-two developing countries. As a private voluntary organization, or PVO, it is not funded as a government agency, nor is it a profit-making firm (Buzzard 1982). Most of the funding for the Foster Parents Plan program comes from individual sponsors (called foster parents) in Australia, Canada, the United Kingdom, the Netherlands, and the United States. These individuals are matched with individual needy foster children. This arrangement is the best-known component of the program. PLAN provides these foster children with various services depending on local resources. A portion of the foster parents' contribution may go directly to the family, often to pay school expenses. In addition, these funds can be aggregated and used for community projects. Community projects may start with a request from the local community. Local PLAN offices are budgeted to fund these projects. In addition, some funds are available from the international headquarters in Rhode Island for special projects. Also PLAN may solicit money from government agencies to fund projects.

The project evaluation considered here was part of the PLAN program funded by the U.S. Agency for International Development (AID) to train community health workers in four field locations—two in Colombia, plus Ecuador and Indonesia. The project discussed here was located in Tumaco, a densely populated seaport town in the Pacific lowlands of Colombia.

Tumaco was faced with a very difficult public health situation. There was limited clean water and no means of sanitary human waste and garbage disposal. Houses were typically built on piles driven into the tidal flats of river estuaries

and beaches. Wastes were disposed of in the water, garbage accumulated near the simple houses, and people defecated on the beaches or in the fields. Although there was a water system, it was old and dilapidated, providing ineffective treatment. In any case, the transport and storage of the water from the tap to its use resulted in more contamination. The effect of this was endemic parasite infestation and high incidence of gastroenteritis, with its debilitating diarrheal disease. Hand washing was infrequent and probably ineffective given the nature of the water. The boiling of water was rare and difficult to get done.

The health problems related to these conditions were exacerbated by poor nutritional status and bad economic conditions. The local diet was based largely on bananas, rice, and cassava. Although the economy was based on littoral fishing and shrimping, many families could not afford these protein-rich foods. As a result, severe protein deficiency could be observed in some children. The generally poor nutritional status of the population tended to increase the effects of the gastrointestinal problems. In this population, diarrheal disease killed children. Other significant health problems were tetanus (especially among newborns), malaria, measles, as well as accidents among the fishermen. Health care facilities were limited.

It was in this context that PLAN initiated the Tumaco Health Project. The project was funded through an agreement with AID. The primary objectives were the improvement of the health and nutritional status of PLAN families, with emphasis on the needs of mothers and children under six years of age. The specific objectives of the five-site project included: (1) decrease the mortality and morbidity from diseases that could be controlled through vaccination; (2) improve sanitation in the family setting to an adequate level; (3) achieve an adequate nutritional level (monitored by age for weight), with special attention to children under six years of age; (4) improve the level of health care and nutrition of mothers during pregnancy and lactation; (5) provide to families health education concerning prevention and treatment of disease; and (6) improve knowledge and access to effective family planning practices.

The Tumaco Health Project sought to achieve these objectives in a number of ways. The core activity was the recruitment, training, and placement of fourteen health promoters from the local community. This required the establishment of an administrative infrastructure and a training program. The functioning of the health promoters and related programs was to be monitored through a record-keeping and evaluation system that would improve quality of management. Program objectives showed a concern for integration of the Tumaco Health Project with the regional system through referral to existing clinics and collaboration with the Ministry of Health. The project was also to result in the construction and staffing of four health posts, installation of public water systems, and the construction of public latrines. All these activities were to be carried out with maximal community participation.

PLAN sought to evaluate the project mid-term, following the training and placement of the health promoters and the establishment of the health posts.

Unlike many other PVOs, PLAN has an ongoing program of evaluation. The evaluation of the Tumaco Health Project reflected the experiences of the PLAN evaluation organization and the reporting requirements associated with the use of AID funds. The evaluation had two primary purposes: the monitoring of the project by PLAN headquarters, and the documentation of the project's functioning for the instruction of others doing similar projects.

The project evaluation was conceived as a case study. Presumably because the evaluation was intended for both internal and external use, there was much reporting of background information on health conditions and the local community. The design of the evaluation was not structured as an experiment to determine program effects. The primary emphasis was the identification of project plan attainments rather than project effects.

The data upon which the evaluation was based was collected using a number of techniques. Participant-observation was used to gain a general familiarity with project functions. The participant-observation data included information gathered by accompanying health promoters on their rounds and the observation of daily activities in the health posts. More data was gathered by interviewing various participants in the program and its planning. This included Ministry of Health administrators as well as health service providers and users. Given the basic task of documenting what went on during the implementation portion of the project, the review of project internal correspondence was quite important. This aspect of the data collection process allowed the evaluator to chronicle the changes in the program. This data was supplemented with reviews of hospital records, health post records, and the data collected through the baseline survey carried out by the health promoters themselves.

The evaluation was more concerned with whether program treatments were implemented than with the actual effects of the treatments, as is often called for in an experimental design. The researcher found that the program was sound. This focus was consistent with the holistic case-study approach used. It shows the importance of the ability to use relatively unstructured inquiry and existing program documents in the evaluation process. The evaluation report shows the utility of an ethnographic approach to evaluation. This particular evaluation was carried out by anthropologist Shirley Buzzard.

Evaluating a Jobs Program in Philadelphia

One of the most significant problems facing inner-city America is unemployment. Programs designed to solve this problem are widespread. The Job Shop Program of the Community Education Center of West Philadelphia is one such program. Awarded a contract from the Pennsylvania Department of Community Affairs, the center was to provide services to unemployed persons in three Philadelphia neighborhoods (Simon and Curtis 1983).

The program was to offer job search assistance, employability training, and

job development services to help unemployed clients. These services were offered as an integrated package of services. The program had four objectives. First, the training of 100 to 120 workers in job readiness, job-seeking skills, and job retention skills, and second, placement of at least 60 percent of these workers in unsubsidized employment or training programs. Third, the program was to assess the employment needs of local employers and to identify potential sources of employees. Fourth, the center was to help program participants retain their jobs by means of counseling and support services.

The center recruited participants through local media, community groups, advertisements, social service organizations, and local churches. The program was intended to serve people in the local neighborhoods. Program participants took classes in getting and keeping jobs. Instructional activities included writing resumes, dealing with applications, reading want ads, the interview experience, and good work habits. Following completion of class work, the participants were to actually search for work using their new skills with the help of project staff. Throughout the project, the staff was to work with local employers to identify possible jobs.

The contract that brought the funds to the center for the project required that the program be evaluated. The evaluation was carried out by Elaine Simon and Karen Curtis, who used a variety of data sources. Program activities were observed and compared with those specified in the proposal. In this way, the "treatment" that the participants received was described.

The proposal presented a rather generalized view of planned program activities. This pattern is very common in evaluation research. It is often the case that programs do not have clearly defined goals. In that goals are frequently the reference point for evaluation, it is not surprising that evaluators often have to spend time identifying goals or using goal-free evaluation techniques.

The evaluation made use of data on individual participants that included referral source, ethnicity, age, education, and previous work history. This was coupled with information on outcomes in employment, referral to other services, and program completion. This data was derived primarily from program records. An interesting component of the evaluation was extensive description of the content of training. The evaluation was further contextualized through the use of a number of participant histories in which their experiences in the program were documented. The individuals selected for profiling in this way were all successful cases. The profiles showed the interrelationship between aspects of the training and preparation and the program's overall success.

The design did not make use of control groups. Moreover, there was no explicit measure of effects of the training on the participants other than the important one, whether they were employed. There was no assessment of levels of new knowledge or changes in attitude. The core of the evaluation consisted of assessing whether the program achieved the numbers of planned participants and whether the predicted placement rates were achieved. These dimensions were

compared in a generalized way with the experiences of other such programs. The quality of the employment that these persons obtained was also considered by examining employment in terms of classification by job type.

The evaluation revealed a program that had largely accomplished its originally stated goals. Coupled with this conclusion were a series of recommendations about the effects of certain participant referral sources. The study showed that participants referred to the program from welfare agencies were less likely to finish the program or to be employed if they did finish.

Assessing Community Needs in Saskatoon

The Saskatoon Needs Assessment Project was carried out by a team from the Department of Anthropology and Archaeology of the University of Saskatchewan, led by Alexander M. Ervin (Ervin et al. 1991). The idea for the project came from the board of the United Way. The executive director of the agency approached Ervin about doing the research. The project was funded in 1990 by a community foundation, the university, and the United Way. Saskatoon's population was about 200,000 at this time. The economy of this prairie city includes agriculture, mining, and forestry, as well a growing manufacturing segment. Unemployment was over 10 percent. There were increases in the use of food banks and soup kitchens.

The project was to provide baseline information for the agency to support their decision making in a number of areas. These included "identifying needs and public perceptions to assist agencies in meeting those needs; allocating funds by working with agencies to target programs in identified needs; evaluating new agencies which have applied to join United Way [and] fundraising by focusing marketing efforts on identified needs which the United Way serve" (Ervin et al. 1991:1).

The research design was developed by the project leader in consultation with an advisory committee, United Way's staff and board, and the research staff. Designs of other Canadian United Way needs assessments were consulted. The assessment process called for six data collection activities. The team reviewed available reports relevant to Saskatoon's needs. These reports, including those from the city government, nongovernment organizations, and academic programs, were abstracted for the final report. The team attempted to review social and economic indicators with the assistance of Statistics Canada. These indicators included census data, household structure, birth rates, labor force, employment, income, disability, and other data. The research team organized three public forums that were highly publicized. They conducted 135 interviews with key informants from community agencies. Five focus groups were held with client groups and one was held with representatives of self-help groups (Greenbaum 1988; Merton, Fiske, and Kendall 1990). Data was also collected using a three-stage Delphi procedure with an expert panel consisting of 28 United Way agency executive directors. Overall the needs assessment had remarkable breadth of

contact with community groups. Over 140 agencies or organizations participated in the interviews and forums, or submitted written briefs.

Delphi procedure was developed as a means of collating the opinions of a panel of experts in a way that allows them to be aware of each others' opinions during the process without them being able to influence each other through their personalities. This technique is also discussed in the chapter on social impact assessment. In this case the process started with a single exercise: "Please list what you consider to be the most important social or human needs that should be addressed in Saskatoon, regardless of what agencies or levels of government are responsible for them." The experts were to type in their answers in ten boxes of equal size. The expert panel never spoke with each other directly, yet they communicated to each other through the research team.

The research team analyzed the responses and put them into standardized phrasing and related clusters. This produced a list of 108 needs. In the second round the experts were asked to choose the top twelve needs in rank order, and then to make comments. This produced another list of 86 needs, ranked in terms of their raw scores. The experts were then asked for any adjustments. The need that was ranked highest was the need "to eliminate hunger and therefore the necessity of food banks." Other highly ranked needs were: "need for more emphasis on preventive services," "need for accessible, affordable, quality accommodation; perhaps based on income (i.e. not low-income ghettos)," and the "need to increase core funding for non-government agencies to enable long term planning and development."

The Delphi panel data was used, along with data from all other sources, to produce an abstract of community needs on "the widespread social problems that are confronting Saskatoon" (Ervin et al. 1991:20). Also reported were the needs of organizations, referred to here as metaneeds. The more than two hundred needs identified were organized into seventeen sectors, including general health, mental health, seniors, native issues, racism and discrimination, immigrant and refugee resettlement, and rehabilitation, among others. The sectors were derived from a directory of community information published by the public library. The research team produced a series of recommendations for the United Way.

SUMMARY

Evaluation research is a rapidly growing area in applied anthropology. Preparation for careers in evaluation should include training in both experimental and case study design, as well as the appropriate data collection techniques. Research methods traditionally associated with anthropology are useful for a number of important evaluation tasks, but these need to be supplemented to meet the entire array of evaluation problems. The utility of ethnographic evaluation methods is directly related to the purpose of the evaluation. Ethnographic evaluation techniques are especially useful when one of the purposes of the evaluation is the documentation of program operations, or the discovery of what went wrong

with a program. Ethnographic techniques serve as a good foundation for providing recommendations for program improvement.

Anthropologists in evaluation do not make extensive use of experimental designs. Usually they rely on various kinds of case study approaches. These approaches are quite variable and represent a significant array of research tools in their own right. The utility of the case study approach can be seen in the interest in these approaches shown by nonanthropologists. The literature on evaluation methodology places an emphasis on the use of experimental designs other than the case study. It is important to recognize that in spite of this, much evaluation is done using the case study approach. The reasons for this are largely practical. In many settings, it is expensive and politically awkward to use the more complex experimental designs. In addition, there are many problems in evaluation where the best and perhaps only approach is the case study method. In spite of the continued importance of the case study method there is relatively little discussion in the literature of refinements to the case study methodology.

FURTHER READING

Cook, Thomas D., and Charles S. Reichardt, eds. 1979. *Qualitative and Quantitative Methods in Evaluation Research.* Sage Research Progress Series in Evaluation, vol. 1. Beverly Hills, Calif.: Sage Publications.

Contains a number of good articles dealing with the qualitative–quantitative contrast. See especially the article by M. G. Trend on the Administrative Agency Experiment.

Fetterman, David M., and M. Pitman, eds. 1986. *Educational Evaluation: Ethnography in Theory, Practice, and Politics.* Beverly Hills, Calif.: Sage Publications.

Fetterman, David M. 1988. *Qualitative Approaches to Evaluation in Education: The Silent Revolution.* New York: Praeger.

David Fetterman has provided leadership in the use of ethnography in evaluation.

13

Technology Development Research

Technology development research is a type of policy or applied research that places the research anthropologist as a communication link between producers and users of new technology. The communication may range from relatively informal interaction to highly organized research and extension projects. The research itself may include elements of needs assessment, evaluation, baseline description, social soundness analysis, as well as extension. This type of policy research requires that the researcher have good understanding of the technology being developed and the occupational culture of the technology developers.

What is called technology development research for the purposes of this chapter includes all those research enterprises that serve the goal of the creation of culturally appropriate technology. These functions are carried out in many contexts, including architecture (Clement 1976), landscape design (Low and Simon 1984), medical treatment (Kendall 1989; Coreil 1989), energy source development (Practical Concepts, Inc. 1980; Roberts 1981), mariculture (Stoffle 1986), reforestation (Murray 1987; Smucker 1981), waste disposal (Elmendorf and Buckles 1978), and housing (Esber 1987; Mason 1979; Wulff 1972). This general type of research is most common in agricultural development (McCorkle 1989; Rhoades 1984).

The achievement of cultural appropriateness is an important goal of most of the intervention and research techniques that are discussed in this book. This value is emphasized in technology development research and cultural brokerage. Cultural appropriateness is, of course, conceptually related to the idea of appropriate technology or intermediate technology, developed by E. F. Schumacher (Schumacher 1973; Dunn 1979; McRobie 1981; Stewart 1977). Developers working within the framework of appropriate technology emphasize the design and manufacture of "small, simple, capital-saving, nonviolent technologies and their supporting institutions" (McRobie 1981: 13).