CHAPTER ONE

Process Evaluation for Public Health Interventions and Research

An Overview

Laura Linnan and Allan Steckler

Improving and sustaining successful public health interventions relies increasingly on the ability to identify the key components of an intervention that are effective, to identify for whom the intervention is effective, and to identify under what conditions the intervention is effective. The published literature includes a plethora of reports about interventions that have successful outcomes. A limited number of studies, however, disentangle the factors that ensure successful outcomes, characterize the failure to achieve success, or attempt to document the steps involved in achieving successful implementation of an intervention. To truly advance science and our understanding of applied interventions, we must learn a great deal more about public health intervention successes and failures. Process evaluation efforts can assist in making these discoveries.

In the last decade, the literature on process evaluation related to public health interventions has grown considerably. In the late 1990s and in early 2000, there has been an explosion in the number of published studies that include extensive process evaluation components. There are several plausible explanations for this noticeable increase in the use of process evaluation. Social and behavioral interventions have become increasingly complex, making it important for researchers to know the extent to which all intervention components are actually implemented. This complexity stems from the fact that projects are often implemented at multiple locations, so that process evaluation becomes essential for ensuring that planned interventions are carried out equally at all sites. Complexity

also results when interventions are implemented at multiple levels and with multiple audiences. Many contemporary studies use ecological approaches that intervene among individual, dyadic, group, organization, community, and population levels (McLeroy, Bibeau, Steckler, and Glanz, 1988). Accounting for the extent to which the intervention occurs at each level of influence (and among different subgroups) becomes increasingly complicated. As we attempt to eliminate inequalities in income and education level, we can, with thoughtful process evaluation efforts, obtain important clues about the influence of interventions among selected subgroups.

Another plausible explanation for why process evaluation efforts have proliferated is that we are looking for ways to explain why certain results were achieved. Specifically, when interventions lead to significant outcomes, it is important to understand which components of the intervention contributed to the success. Similarly, when large sums of taxpayer money are allocated to conduct multilevel community trials and the primary outcomes are insignificant (or the magnitude of the effect is small), there is an increased demand among researchers, funders, and members of the general public to understand why these interventions did not produce their intended effect (Fisher, 1995; Susser, 1995). Thus, process evaluation can help explain positive, modest, and insignificant results.

Process evaluation also provides important links to understanding and improving theory-informed interventions. Because more programs are developing theory-informed interventions, there is a greater need to understand which theoretical constructs make a difference (Glanz, Lewis, and Rimer, 1997). Process evaluation efforts can provide a link between theoretical constructs thought to be essential for intervention success and the final study outcomes. Understanding the mechanisms for how and why these constructs produce successful change (or fail to produce change) is key to refining theory and improving intervention effectiveness. Understanding under what conditions certain mediators are more (or less) influential in the achievement of certain study outcomes will inform the next and future generations of theory-informed interventions (Baranowski, Anderson, and Carmack, 1998; Baranowski and others, 1997).

Process evaluation efforts also help us understand the relationships among selected intervention or program components. In a comprehensive multimethod intervention, one might offer group classes, self-help programs, individual counseling sessions, and print materials. Process evaluation can help disentangle the effects of each method singly, and it can clarify the possible interactions that can occur to produce a synergistic effect. Few studies have attempted to reach into the "black box" of intervention effectiveness to explore which program components have been more or less effective, yet doing so is a powerful option within a comprehensive process evaluation effort (Harachi and others, 1999).

Assessing the quality and accuracy of the intervention delivered to program participants can also be achieved with process evaluation, which is another reason why these studies have proliferated. Increasingly, funders and program participants at all levels want assurances that the intervention being delivered is of a high quality and is highly accurate. In addition, attention to the cost-effectiveness of interventions is an increasingly important component of intervention planning and evaluation. Process evaluation efforts can assist with each of these requirements.

Finally, we contend that another reason for the rise in the use of process evaluation is the increasing recognition of the value of qualitative research methods (National Institutes of Health, 2001). Process evaluation frequently uses both quantitative and qualitative methods. Qualitative methods such as field site visits, structured observations of interventions, and open-ended interviews are often employed in conducting process evaluations. As the value attached to qualitative methods has increased, opportunities for publishing these findings and learning from their results have increased as well. Integrating different methods—such as qualitative and quantitative methods—yields rich detail about study outcomes that neither method could achieve alone (Tashakkori and Teddlie, 1998; Steckler and others, 1992).

Taken together, the recent increase in published literature on process evaluation results reflects the complexity of public health interventions today, and it reflects the many ways in which thoughtful, comprehensive process evaluation efforts can shed light on questions that will inform improvements in theory, intervention design, and methods in the future.

The remainder of this chapter presents a brief overview of the history of process evaluation as it relates to public health. After identifying current gaps in knowledge and clarifying barriers to effective process evaluation, the chapter presents definitions of key process evaluation terms and describes an approach that program planners, researchers, and evaluators may find helpful when planning and conducting process evaluation efforts.

Brief History and Review

Process evaluation is not a new concept. As early as during the 1960s, an explanation of process evaluation appeared in a widely used textbook on program evaluation (Suchman, 1967), although Suchman does not label it "process evaluation" per se. Suchman writes:

> In the course of evaluating the success or failure of a program, a great deal can be learned about how and why a program works or does not work. Strictly speaking, this analysis of the process whereby a program produces the results it

> does is not an inherent part of evaluative research. An evaluation study may limit its data collection and analysis simply to determining whether or not a program is successful. . . . However, an analysis of process can have both administrative and scientific significance, particularly where the evaluation indicates that a program is not working as expected. Locating the cause of the failure may result in modifying the program so that it will work, instead of its being discarded as a complete failure [p. 66].

This early definition of process evaluation includes the basic framework that is still used today; however, as is discussed later in this chapter, the definitions of the components of process evaluation have been further developed and refined.

Few references to process evaluation were made in the literature during the 1970s. In evaluation research, the 1970s were devoted to the issues of improving evaluation designs and measuring program effects. For instance, Struening and Guttentag's *Handbook of Evaluation Research* (1975) does not contain any reference to process evaluation (see also Glass, 1976; Guttentag and Saar, 1977; Green, 1977).

In their influential book, Green, Kreuter, Deeds, and Partridge (1980) define process evaluation in a somewhat unusual way: "In a process evaluation, the object of interest is professional practice, and the standard of acceptability is appropriate practice. Quality is monitored by various means, including audit, peer review, accreditation, certification, and government or administrative surveillance of contracts and grants" (p. 134).

The emphasis on professional practice as the focus of process evaluation as suggested by Green, Kreuter, Deeds, and Partridge (1980) faded as attention returned to the idea of assessment of program implementation. By the mid-1980s, the definition of process evaluation had expanded. Windsor, Baranowski, Clark, and Cutter (1984) explain the purpose of process evaluation in the following way:

> Process produces documentation on what is going on in a program and confirms the existence and availability of physical and structural elements of the program. It is part of a formative evaluation and assesses whether specific elements such as facilities, staff, space, or services are being provided or being established according to the given program plan. . . . Process evaluation involves documentation and description of specific program activities—how much of what, for whom, when, and by whom. It includes monitoring the frequency of participation by the target population and is used to confirm the frequency and extent of implementation of selected programs or program elements. Process evaluation derives evidence from staff, consumers, or outside evaluators on the quality of the implementation plan and on the appropriateness of content, methods, materials, media, and instruments [p. 3].

McGraw and others (1989) write that "'process' or 'implementation' evaluation is not a new concept. . . . Yet in the broad field of social experimentation, it is only recently—since the late 1970s, that this aspect of program evaluation is accorded more formal recognition" (p. 460). This historical assessment appears to be accurate. Starting in the mid- to late 1980s, we can see the beginnings of contemporary process evaluation theory and methods and their application to applied public health interventions.

One of the key publications in the mid-1980s that laid the groundwork for modern process evaluation was an article titled "Avoiding Type III Errors in Health Education Program Evaluations: A Case Study" (Basch and others, 1985). Researchers are familiar with a Type I error (for example, rejecting a "true" null hypothesis) or a Type II error (for example, failing to reject a "false" null hypothesis), but a Type III error ensues from "evaluating a program that has not been adequately implemented" (p. 316). Although the idea of the Type III error was not new in 1985 (Dobson and Cook, 1980; Rezmovic, 1982), our research suggests that this was the first time this idea appeared in the public health literature. The article argues that measuring program implementation is critical to avoiding a Type III error and thus drawing incorrect conclusions about the effectiveness of a given intervention. That is, in addition to answering the evaluation question Did the program work? evaluators must first answer the question Was the program actually carried out as planned?—and if it was not carried out as planned, they must answer the question How did the program vary from the original plan?

The community-based cardiovascular disease prevention (CVD) demonstration studies funded by the National Heart, Lung, and Blood Institute (NHLBI) in the early 1980s through the next decade represented an important step forward in planning, implementing, and evaluating community-based public health interventions. Investigators from the Stanford Five-City Program, the Pawtucket Heart Health Program, and the Minnesota Heart Health Program had the foresight to get together periodically and look for ways to collaborate on the development of research questions, data collection, interventions, and program evaluation. The program evaluation units in each of the studies were independent (Pirie and others, 1994), but as studies evolved, investigators began to realize the importance of developing a consistent approach to assessing the *dose* of intervention delivered by the program staff in each of the communities. As a result, the three demonstration studies combined key components of their process tracking systems so that they could compare the extent of their activities by targeted risk factors (smoking, nutrition, physical activity, blood pressure, and blood cholesterol levels) over time. It was a big effort to join components of the three tracking systems that were already in the field, but the benefits of this effort were

significant. The intervention staff of each project obtained feedback on whether they were meeting expectations for program delivery. Corrective feedback was possible, based on the results.

In addition, these studies were among the first intervention studies to conceptualize and measure the importance of the dose received by program participants. Previously, program evaluators were content to measure the extent to which the intervention was delivered as planned. However, investigators were aware that many interventions were delivered that participants never received. For example, programs were scheduled and offered, but no one attended them. It was important to recognize that if programs were not received, important corrective action needed to be taken to improve the intervention or marketing of these programs. Moreover, when interpreting results, one must take into account the fact that participants may not have received what was delivered. As a result, the CVD demonstration studies advanced the field by collecting data from community members (through surveys) to assess participation in recently sponsored programs and exposure to various media messages. Pirie and others (1994) have published a detailed descriptive paper on program evaluation strategies of the demonstration studies which underscores the numerous contributions made by these investigators to conceptual thinking and the operationalization of key program evaluation components. The individual projects also published numerous papers that offer great insights about developing and implementing process evaluation efforts that will benefit process evaluators today. Of the many articles published, we recommend the one by Finnegan, Murray, Kurth, and McCarthy (1989), which describes the tracking of program implementation for the Minnesota Heart Health Program, the one by McGraw and others (1989), which describes the process evaluation system of the Pawtucket Heart Health Program, and the one by Flora and others (1993), which describes the community education monitoring system of the combined demonstration studies.

Because these demonstration projects were the largest federally funded trials, many beginning investigators received their training in public health interventions and evaluations by getting involved with these projects. By the early 1990s, investigators who had worked on these studies were now publishing papers on components of the process evaluation data and were conducting community-based studies of their own. A number of investigators from the CVD demonstration studies were collaborators on the next large, federally funded multi-institute trial. Corbett, Thompson, White, and Taylor (1991) describe the process evaluation used in the National Cancer Institute (NCI)-funded Community Intervention Trial for Smoking Cessation (COMMIT) (1988–1991). For this large community-based intervention study, they applied what they learned from the limitations of the CVD demonstration study tracking system and state in their 1991 article that process evaluation

> addresses what the intervention program consists of, how activities serve short-term objectives, how activities are carried out, and what other factors contribute to outcomes. It may entail "implementation evaluation," "quality control," "quality assurance review," "program utility assessment," "process analysis," and other assessments. Process evaluation can have a "formative" role during the development and unfolding of a program as well as a "summative" function. Quantitative process objectives along a timeline are often employed to facilitate evaluation of program delivery. Process evaluation may be designed to provide information to feed back suggestions to program designers, for mid-course corrections. For some researchers, process evaluation may also refer to qualitative assessment of the dynamics of program operation. Ultimately, what is needed is to know what specific results were and how they came about [p. 293].

The early 1990s represented an important period of growth in the number and complexity of community- and school-based intervention studies. At approximately the same time as the COMMIT study, the NCI funded the national Working Well Trial (1989–1994)—the largest federally funded worksite-based intervention trial (Abrams and others, 1994; Heimendinger and others, 1995; Sorensen and others, 1996) (see Chapter Six for a description of the process evaluation efforts of Working Well)—and an NHLBI-funded school-based study titled CATCH (Child and Adolescent Trial for Cardiovascular Health) (1986–1994) was conducted. The study design had been published earlier (Stone, McGraw, Osganian, and Elder, 1994), but the main process evaluation results were published in 1997 (Perry and others, 1997). The CATCH intervention had been implemented in fifty-six schools in four states. The process evaluation had four main objectives: (1) *participation*—Did teachers, food service personnel, and PE specialists attend the training sessions? (2) *dose*—Were prescribed components of the CATCH program implemented? (3) *fidelity*—Were the prescribed intervention components implemented according to protocol? and (4) *compatibility*—Did the CATCH programs fit the context of the schools as well as the needs, expectations, and values of the staff members and teachers? Because of its extensive process evaluation, the CATCH study made important contributions to the development of process evaluation theory and methods (Perry and others, 1997). McGraw and others (1994) describe the overall model for the CATCH process evaluation, which includes the measurement of student characteristics, intervention activities, student outcomes, the exogenous and competing effects, and school characteristics/outcomes. In addition, the interventions were based on principles of organizational change and social cognitive theory. The CATCH process evaluation data were used to describe program implementation, quality control and monitoring, and program effects. The CATCH process evaluation data were also used to assess the

environmental context in which the interventions took place (Elder and others, 1994).

By 2000, the design and implementation of process evaluation efforts became quite complex, reflecting, in part, the complexity of the interventions they sought to monitor (Bartholomew, Parcel, Kok, and Gottlieb, 2001). Important conceptual work on process evaluation has been the hallmark of recent published reports on process evaluation. Perhaps the most thorough explanation of the components of process evaluation is offered by Baranowski and Stables (2000), who list eleven components of process evaluation:

1. *Recruitment*—attracting agencies, implementers, or potential participants for corresponding parts of the program
2. *Maintenance*—keeping participants involved in the programmatic and data collection
3. *Context*—aspects of the environment of an intervention
4. *Resources*—the materials or characteristics of agencies, implementers, or participants necessary to attain project goals
5. *Implementation*—the extent to which the program is implemented as designed
6. *Reach*—the extent to which the program contacts or is received by the targeted group
7. *Barriers*—problems encountered in reaching participants
8. *Exposure*—the extent to which participants view or read the materials that reach them
9. *Initial use*—the extent to which a participant conducts activities specified in the materials
10. *Continued use*—the extent to which a participant continues to do any of the activities
11. *Contamination*—the extent to which participants receive interventions from outside the program and the extent to which the control group receives the treatment

This list provides a useful beginning framework for organizing conceptual thinking about process evaluation and for developing consistent definitions to be used in the measurement of key process evaluation components.

Gaps in Current Knowledge

A number of gaps in current knowledge about process evaluation must be addressed if the field is to move forward. We review a number of these gaps, particularly focusing on (1) the lack of clear, consistent definitions for key process

evaluation components and (2) the lack of a systematic process for planning and developing a process evaluation effort.

A selected review of recent process evaluations revealed an impressive array of process measures. For example, process evaluation efforts have measured the fidelity of the intervention implementation (that is, the quality of the program implementation), its reach, the time spent on program activities, the use of intervention materials, the level of participation, the dose delivered, the external factors, the program's penetration, its quality measures—such as the accuracy of the information and services provided, its impact, the relationships among program components, the training results, and the costs. At this point, there appears to be considerable overlap in how key terms like fidelity, dose, exposure, and reach are defined. Because of the diversity of process evaluation measures, methods used to collect the data, and ways of reporting results, it is difficult to compare findings across studies, and all but the most experienced evaluators may be overwhelmed by it. Creating clear, consistent definitions of existing process evaluation outcome measures would fill an existing gap in knowledge about process evaluation.

The lack of a systematic approach to guiding process evaluation efforts causes another serious gap in current knowledge about process evaluation. Most published literature describes the process evaluation for a single project or research effort. Few resources that take a systematic approach to designing a process evaluation effort are available in the current literature. Without guides for the planning and development of a process evaluation effort, project staff members are often left to reinvent an approach as well as the forms and systems needed to carry out the evaluation effort. Thus, a gap in current knowledge about process evaluation results from the lack of a stepwise approach to creating and implementing a process evaluation effort.

Independent of the approach used are key components that evaluators must consider when planning and carrying out a comprehensive process evaluation effort. First, the role of theory is unclear in many of the most recent process evaluation efforts (Weiss, 1998). When theory guides intervention development, process evaluation efforts that measure the implementation of each intervention component will also be theory-linked. If evaluators fail to specify the underlying theory upon which interventions are developed, they will miss the opportunity to advance our understanding of likely mechanisms of change and how theory—and interventions—can be improved to reflect that new understanding.

Second, in view of the resource intensity of comprehensive process evaluation efforts, researchers should attempt to strike a practical balance between the data that is clearly needed and the data that is merely "nice to have." A "less is more" approach to data collection may be preferable to collecting all possible data. The burden on the project staff (for collecting and managing the data) and on participants (for providing data) can often be a big barrier to collecting all the required

data and for maintaining the quality of the collected data. Embedded in the larger stepwise approach to process evaluation must be a process for prioritizing the research questions that are addressed, thus giving priority to the type and amount of data collected and the methods used to collect data.

Process evaluation typically yields a wealth of information—particularly because a wide range of methods are used to collect data. Typical quantitative data collection methods include surveys, reports, checklists, attendance logs, self-administered forms, project archives, and community profiles. Qualitative data collection methods include observations, structured interviews, focus groups, and content analysis of audiotapes and videotapes. These lists are by no means exhaustive, but very little is known about which methods are more (or less) appropriate in certain situations. For example, when is it best to use qualitative versus quantitative methods? Which methods are more successfully used in middle schools than in elementary schools? Which methods are quite appropriate in faith-based settings but are not appropriate in worksites? Although using a mixture of the qualitative and quantitative methods brings the strengths of both approaches to an evaluation effort, limited resources often require deciding which method is the most cost-effective for answering a particular question. Little information is presently available to guide the decision-making process.

Wide variation in analytic strategies for presenting process evaluation results exists, yet it is fair to say that most process evaluation analyses employ descriptive statistics. As powerful analytic strategies for assessing study outcomes are developed, it is desirable to employ these same techniques, when appropriate, to the process evaluation data. For instance, McGraw and others (1994) and Baranowski, Anderson, and Carmack (1998) clarify the role and benefits of mediator analyses by using process evaluation data. Process evaluation is often focused on the mechanisms of change, and because new analytic strategies are being developed on a regular basis to deal with these mediator questions, evaluators are encouraged to explore the full range of analytic strategies available to answer prioritized questions.

Finally, related to a lack of guidance about how to plan, develop, and implement process evaluation efforts is the fact that there are few training opportunities available for students or professionals who want this expertise. In typical public health training programs, one course on program planning and evaluation may be required. With the many essential topics that must be covered in that type of course, process evaluation may be covered in one or two lectures. Even with advanced training, unless a student works on a project or study in which a process evaluation effort is being designed, there are simply few opportunities to learn about how to build effective process evaluation efforts. On many projects, an outside consultant may be hired to assist with the process evaluation effort. Yet when time and resources are spent getting the evaluator "up to speed" on a particular

set of intervention goals, less time and money are available to focus on the process evaluation effort itself. In general, most public health professionals would benefit from additional training on program evaluation—including process evaluation—to build a larger pool of professional resources with expertise in this area.

Advancing Future Public Health–Linked Process Evaluation Efforts

Like Patton's utilization-focused evaluation (1997), process evaluation will be most effective when it takes into account the needs, wants, and concerns of the potential users of the system. Key stakeholders in designing an effective process evaluation effort include all the project staff members who deliver services or handle data, project managers and investigators who have been involved in designing an intervention or project evaluation effort, and participants who may be asked to collect data and review reports. We view the development process as highly iterative—trying out various techniques, revising them, and reaching consensus on key forms and tools to be used to collect the data, the questions to be answered with these process evaluation data, and reports generated to inform stakeholders about progress and problems. In addition to using a collaborative process, including extensive pretesting, reaching consensus on a set of terms that can be universally recognized and applied to process evaluation efforts is desirable.

Defining the Components of Process Evaluation

As shown in Table 1.1, *context* refers to the larger physical, social, and political environment that either directly or indirectly affects an intervention program. Since process evaluation is concerned with answering how and why an intervention was successful or not, an understanding of the context is often necessary. To assess the context, process evaluators determine which environmental factors might influence program implementation and then determine how the appropriate data might be collected. For instance, in a program designed to increase physical activity among young people, access to recreational facilities would be an important contextual factor to take into account. Even if the intervention to increase physical activity was not successful overall, perhaps certain subgroups of young people—for example, those with greater access to recreational facilities—were more likely to experience increases in physical activity, whereas young people with little access to such facilities might be less likely to experience such positive outcomes. Brainstorming a list of potentially important contextual factors prior to

TABLE 1.1. KEY PROCESS EVALUATION COMPONENTS.

Component	Definition
Context	Aspects of the larger social, political, and economic environment that may influence intervention implementation.
Reach	The proportion of intended target audience that participates in an intervention. If there are multiple interventions, then it is the proportion that participates in each intervention or component. It is often measured by attendance. Reach is a characteristic of the target audience.
Dose delivered	The number or amount of intended units of each intervention or each component delivered or provided. Dose delivered is a function of efforts of the intervention providers.
Dose received	The extent to which participants actively engage with, interact with, are receptive to, and/or use materials or recommended resources. Dose received is a characteristic of the target audience and it assesses the extent of engagement of participants with the intervention.
Fidelity	The extent to which the intervention was delivered as planned. It represents the quality and integrity of the intervention as conceived by the developers. Fidelity is a function of the intervention providers.
Implementation	A composite score that indicates the extent to which the intervention has been implemented and received by the intended audience.
Recruitment	Procedures used to approach and attract participants. Recruitment often occurs at the individual and organizational/community levels.

intervention delivery is the ideal situation. Sometimes, existing archival data can be used retrospectively to understand contextual influences.

Reach concerns the degree to which the intended audience participates in an intervention (Glasgow, Vogt, and Boles, 1999; Glasgow, McCaul, and Fisher, 1993). Reach is often measured as the percentage or proportion of the target audience that attends a given intervention or part of an intervention. Effective intervention programs aim to reach as many participants as possible; therefore, measurement of reach is critical for estimating total program implementation.

Knowing which subgroups of the intended target population actually participate is also critically important. For example, if overall reach is moderately high but only the healthiest subgroup of individuals participates, evaluators would be interested in taking corrective action to extend the reach among the entire population. At a minimum, evaluators typically assess certain characteristics of the population, such as health status or health risk level, age, race/ethnicity, gender,

and income and education level. In calculating the reach, it is sometimes difficult to know precisely what the correct denominator is for determining the percentage of a target audience that has participated. For example, in faith-based settings, the church membership may change on a weekly basis. How one determines the actual membership may be determined by averaging attendance over an eight-week period or through some other reasonable estimating procedure. Or if the goal of a program is to reach senior citizens in a community, census data from government agencies may be available to estimate how many community members fifty-five years of age and older reside in a given community.

Dose delivered is a term that is commonly used in the process evaluation literature and it refers to the amount or proportion of the intended intervention that is actually delivered to program participants. Dose delivered is directly related to program implementation. By this we mean that dose is usually determined by the actions or behaviors of those who deliver the intervention. The process evaluation question that dose delivered answers is, What proportion of the intended intervention was actually delivered to the intended audience? For instance, how many lessons from the curriculum were actually delivered by the trained teachers to the fifth-grade students? Related to dose delivered is the concept of *dose received*. Baranowski and Stables (2000) term dose received as "exposure," or a measure of the extent to which participants receive and use educational materials or other recommended resources. Like Pirie and others (1994), we prefer to think about dose delivered and dose received as two important, conceptually similar, yet different process evaluation components. To measure dose received, evaluators would ask the question What proportion of the educational materials did participants actually receive? and What proportion was read, viewed, or otherwise used by participants? Pirie and others (1994) suggest that surveys of participant program awareness, message awareness, or other surveillance/monitoring strategies are common methods for ascertaining dose received.

Of the process evaluation components shown in Table 1.1, perhaps the most difficult to measure is fidelity. *Fidelity* refers to the quality of the implementation of an intervention. Although it appears fairly straightforward, devising appropriate measures is often difficult because quality may appear to be a subjective notion (Dusenbury, Brannigan, Falco, and Hansen, 2001). Measures of fidelity include addressing whether the intervention is carried out according to a prespecified plan and whether it is carried out in both the manner and the spirit in which it was intended. It is the manner and the spirit that often prove difficult to assess. Some projects develop checklists of core intervention components or minimum requirements that an intervention must include in order to receive a high fidelity rating (Baranowski and Stables, 2000). Fidelity may be assessed by observations of intervention implementation by a trained observer using a structured

observation guide (Resnicow and others, 1998). A less expensive method is to have program implementation staff members fill out some type of survey or questionnaire to assess how an intervention was implemented. For example, they indicate whether all the components of the intervention were carried out and rate how well they were carried out. Of course, the problem with this approach is the possibility of biased response or recall. Identifying multiple indicators of fidelity may strengthen convictions about results. Balancing cost considerations with creative thinking about efficient, unobtrusive ways to collect data on fidelity is an area of ongoing research.

Program *implementation* includes a combination of reach (who participated), dose (what the program delivered), dose received (what participants received), and fidelity (the quality of the intervention delivered). Whereas fidelity is often a difficult process evaluation component to measure, program implementation is difficult to operationalize (or calculate). Program implementation relies on accurate measurement of the four components and must then add a weighting factor to determine the final implementation score. Some authors have recommended that implementation be the result of the product of reach, dose, and fidelity (Baranowski and Stables, 2000; Glasgow, Vogt, and Boles, 1999). An alternative approach would be to average the four to assess implementation. Using the multiplicative approach, if 75 percent of the audience is reached, with 75 percent dose delivered, 75 percent dose received, and 75 percent fidelity, then a program implementation "score" would be 0.32 (for example, $0.75 \times 0.75 \times 0.75 \times 0.75$). Using the averaging approach, such as $0.75 + 0.75 + 0.75 + 0.75/4$, the program implementation score would be 0.75—for example, the average of the four process measures. Program evaluators should decide a priori what method of calculating implementation will be used in a given project and what the acceptable levels of implementation will be. For instance, is an implementation rate of 75 percent desirable? If not, then what is the acceptable rate? If intervention staff members have a realistic implementation score to aim for, they are in a much better position to achieve the objectives. In establishing realistic implementation score objectives, researchers should take into account evidence on participation rates linked to certain intervention dose levels available in the literature, or they should talk with colleagues who are attempting similar interventions.

Baranowski and Stables (2000) argue that *recruitment* is a key process evaluation component. Recruitment refers to the procedures used to approach and attract prospective program participants. Examining the resources that were employed and the reasons for nonparticipation among individuals and/or organizations that were approached can be used as measures of recruitment effectiveness. Typical process evaluation questions related to recruitment include: Which subgroups of individuals or organizations were more (or less) likely to be

successfully recruited? Why were certain groups of individuals or organizations more (or less) likely to be recruited? Was the recruitment process consistently applied across all individuals or organizations? If recruitment efforts yield a biased sample, it is important to understand the implications of the bias on the final outcomes. With appropriate process-related recruitment results, evaluators/investigators will avoid overgeneralizing findings to all subgroups or attributing widespread success to a project that was not truly tested in all populations.

Thus, we contend that at the very least, process evaluators should collect data to determine the context (including documentation of recruitment efforts), the reach, the dose (delivered and received), and the fidelity of the intervention. In addition, accurate information should be gathered to describe the context in which the intervention occurred, decisions should be made early in the study about how to determine an implementation score for the intervention, and recruitment procedures should be documented.

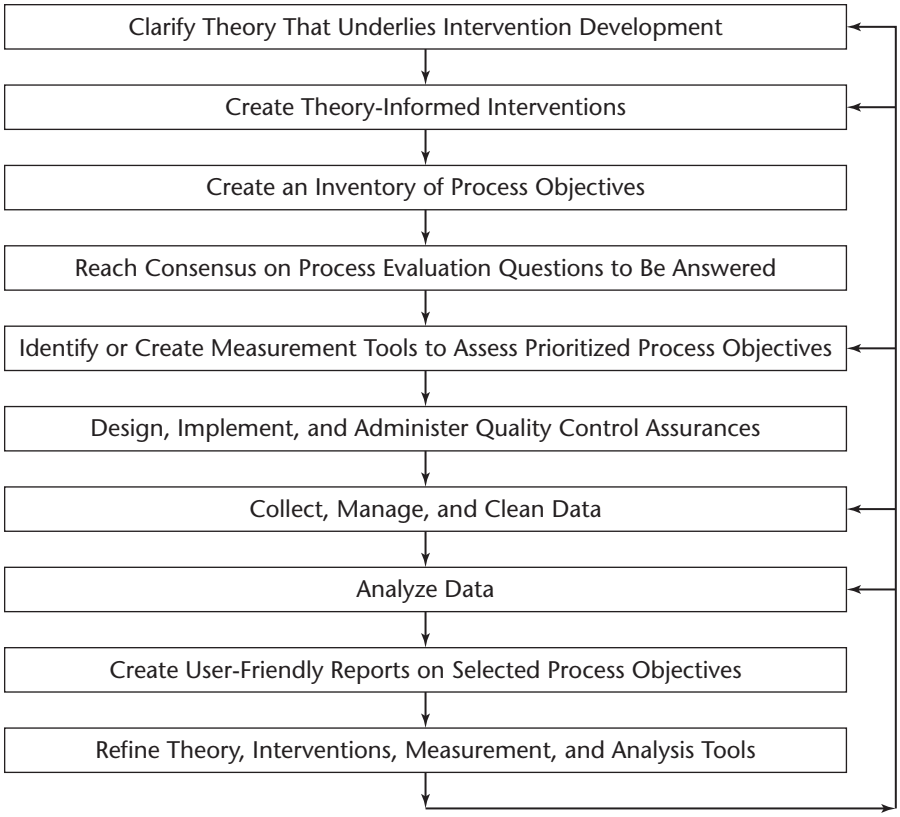
In addition to these minimum requirements, a host of additional process evaluation variables may be collected, including changes in intermediate outcomes that lead to hypothesized changes in final outcomes, training results, program-specific results (embedded within the larger study outcomes), formative/pretesting procedures, and various quality assurance measures. Because there are often limits in terms of resources, time, and personnel to collect these data, difficult decisions must be made about which process evaluation data are to be collected and analyzed. The remainder of this chapter focuses on a stepwise approach that evaluators may use to plan for, prioritize, and implement a successful process evaluation effort.

Designing and Implementing Process Evaluation Efforts

A clear definition of terms, a collaborative process that includes the evaluator and all key stakeholders, assurance that pretesting will occur, and an effort to use the following systematic approach will assist program evaluators in designing and implementing effective process evaluation efforts.

Clarify theory. Figure 1.1 represents a recommended process for designing and implementing process evaluation efforts. Building on the guidance of Weiss (1998), McGraw and others (1989), and Helitzer and Yoon (Chapter Four), we begin with theory. Specifically, evaluators are asked to specify the theory that underlies the intervention being delivered. Often, it is useful to draw a conceptual model of the intervention (Earp and Ennett, 1991), or logic model, clarifying the specific theoretical constructs of interest, those expected to change, and possible mediators of the change process. From those conceptual plans, *interventions that influence specified theoretical constructs of interest can be developed* and pretested with

FIGURE 1.1. A PROCESS FOR DESIGNING AND IMPLEMENTING EFFECTIVE PROCESS EVALUATION EFFORTS.



Note: It is recommended that key stakeholders be involved in all aspects of this process.

intended audiences. The role of theory does not end at this first step. Process evaluation results can be used to test theory (or parts of theory) as well as to create new theory. This process encourages the use of theory to guide the planning and implementation of a process evaluation effort.

Create an inventory of process objectives. For any public health intervention, there are likely to be several important intervention components. For example, in a comprehensive tobacco control worksite-based intervention, one might offer smoking cessation classes, self-help quitting programs, print materials with information about the dangers associated with environmental tobacco smoke, and a twenty-four-hour quit challenge. The program developers are in the best position to write

realistic, measurable process objectives for each intervention component. Using theory and evidence from previous (and similar) interventions, program developers have the required expertise to know what to expect from an intervention of a certain intensity, frequency, and duration.

Following the flow diagram (Figure 1.1), the program developers would write one or more process objectives for each key intervention component. Like all objectives, these should be measurable and time-sensitive, and they should include information about what will specifically be delivered, who will receive the intervention, by when will they receive it, and what method will be used. If the first three steps of the process are followed, the process objectives will provide a measurable and theoretical link to the intervention. The process objectives will also help guide the evaluation effort.

Reach consensus on process evaluation questions to be answered. Once an inventory of process objectives is compiled, the team has provided a great deal of clarity about the intervention plan and its theoretical and conceptual links. Next, the team should reach consensus on the process evaluation questions to be addressed. As mentioned previously, the tendency will be to try to answer all possible questions, but this is not recommended in most cases.

Instead, consider brainstorming a list of possible process evaluation questions. To create the list of possible questions, start with the key components of a process evaluation that are defined in this chapter. For example, at a minimum, the team should consider placing a priority on getting answers to questions about reach, dose, and fidelity. In addition, if possible, get answers to questions about the context in which the intervention is taking place, as well as some information about the recruitment process.

As the team reviews the full inventory of process objectives, they might be tempted to create questions that address all of them. However, prioritizing the questions and performing a good measure of a few process objectives is likely to yield far more meaningful results than trying to measure all possible objectives. The burden on staff members and participants to collect and manage enormous amounts of data can become prohibitive. Thus, evaluators should prioritize by establishing criteria (a priori) for selecting which questions will yield the most valuable information for a particular study. The priority list of evaluation questions to be answered should be determined by using a consensus-building process with the project team. For example, if the intervention project is focused on a required training, there may be less focus on recruitment and much more focus on the costs and quality of the training approach. Here, the project team will make decisions based on consideration of the intervention, organizational demands or barriers, funding agency requirements, research study aims, available resources, and potential burden to the participants or study staff. In other words, decisions about

what evaluation questions should be prioritized represent an effort to match the art and the politics with the science of developing effective process evaluation efforts. A collaborative process using a team approach to prioritize all possible questions will help focus the process evaluation effort. Moreover, this process will help ensure that all team members have the same expectations about what the process evaluation effort will achieve.

Creating a number of decision rules—or criteria for selecting priority questions up front—may make the process easier for all concerned. For example, theory may be a guide for prioritizing among particular questions. Alternatively, if the evaluation budget is very tight, the team may agree to assess the cost of answering each question as part of the decision rules for prioritizing evaluation questions. Another method for prioritizing among all possible research questions could follow the approach offered by Green and Kreuter (1999) in the third phase of the PRECEDE planning process. They inventory all possible behavioral and environmental determinants influencing a specific public health problem and then prioritize both within and between those categories of determinants. Applying that notion to the evaluation effort, a grid could be used to rate each question by the changeability and importance of the question to the overall process evaluation effort. Thus, the project staff would consider all possible process evaluation questions and would follow a systematic approach for prioritizing the questions to be answered. The bottom line is that the team decides on the decision rules up front and has all the stakeholders who are involved in the process prioritize the key questions.

Identify measurement tools. All process objectives should be measurable so that well-written objectives will indicate how the objective is to be measured. Once process evaluation questions are prioritized, the evaluator reviews the process objectives that line up with a particular research question and either *identifies* or *creates* measurement tools to assess them. It is possible that no appropriate tools are available—or easily constructed. In this case, that particular evaluation question may be reprioritized and another from the inventory may be moved forward. Clarifying the ways in which the data will be collected to assess a particular process objective must take into account (1) the type of data collection required (paper and pencil forms, electronic forms, or some other technology), (2) the frequency of data collection, (3) who is responsible for data collection, (4) the reliability and validity of the data collection measures, (5) the cost, and (6) the potential burden to participants and staff members. These are not simple issues. The evaluator is encouraged to review sample forms (like those found in this book) to reduce development time. It is a good idea to talk with other process evaluators about potential problems encountered with a particular data collection method, as well as talk with potential end users for the measurement tool—and get their input

early in the process. When the measurement tools for each process objective are identified, it is imperative to pretest the use of these tools under "real-world" conditions. An iterative process of testing, revising, retesting, and improving these forms is ideal. Allowing enough time to accomplish a thorough pretest of the data collection tools cannot be stressed enough. The time taken up front should be viewed as an investment that will reap great returns when it comes time to analyze and make sense of the data.

Quality assurance. Quality assurances about the data collection, coding, and management steps should be built into the process evaluation effort early and should be monitored throughout the study implementation period. In the Working Well Trial (see Chapter Six), several quality assurance steps were instituted by the data coordinating center at the Fred Hutchinson Cancer Research Center. First, field staff at all four study centers were required to complete a standardized training and certification program for using the study-wide process tracking system before they engaged in data collection. If staff members did not "pass" the certification test, they were asked to repeat the test until they met proficiency requirements. Second, process tracking coordinators at each study center reviewed the printed data collection forms and screened them for missing information and potential coding errors before the data were entered into the database locally. A majority of errors were identified through this local checking process. Third, the local database was then uploaded to the data coordinating center and the staff there did a second quality assurance check to identify missing data, potential coding errors, or any other systematic errors that might occur. Fourth, a monthly conference call between the data coordinating center staff and the local process tracking coordinator identified problems, provided updates on new coding information, and offered ongoing training and collaboration on the use of the tracking system. Fifth, reports were generated on a regular basis for the process tracking coordinator and the project directors/principal investigators at each study center to keep them apprised of the status of the process evaluation effort. These reports also served as a quality assurance check when problems were identified and solved.

It may not be realistic to assume that in all projects it will be possible to establish the extensive quality assurance steps undertaken in a large national trial like Working Well, but a series of quality assurances must be in place to ensure that the process data collected is meeting established quality standards and that project staff members are aware of and committed to the quality assurance process. Having a project staff person dedicated to the data management task is an ideal situation. This individual is responsible for troubleshooting problems, upholding quality standards, and monitoring/reporting on the status of the process evaluation effort over time. Technology has advanced at a fast pace, and the tools

available to manage large datasets typically found in a process evaluation effort are becoming more readily available.

Although cost and complexity are always important considerations, using a relational database that allows evaluators to link different sources of data by unique identifiers is recommended. Flexibility is also essential when selecting either the technology to collect the data or the software packages to manage the data. Often, a process evaluation effort will be modified over the course of a project or study. If the system one is using to manage the data is inflexible, or if the software is so rigid that modifications are not possible (without great expense to the project), then data management efforts can become difficult. Evaluators are encouraged to talk with colleagues who have managed process evaluation data on similar projects to get an idea of what has worked well—or not so well. Blaine, Guire, and Forster (Chapter Eleven) review the steps required to develop and revise a statewide data management tool for tracking local smoking control efforts in Minnesota. This chapter offers useful insights and practical tips about how to create these systems and what to avoid in the process.

Data analysis and reporting. The analysis and reporting of process evaluation data does not typically receive much attention in the published literature. A few important points are worth consideration as evaluators face these steps in a process evaluation effort. First, evaluators should consider moving beyond descriptive statistics when analyzing process evaluation data. In some cases, this will be impossible—or unnecessary, but in the cases where sample size allows for multivariate statistics, it is critical to apply the most powerful analytic strategies available when reporting results. Moreover, to fully understand the mediating mechanisms of change, structural equation modeling techniques might prove particularly useful. Most evaluators are beginning to recognize the importance of using mixed methods and mixed analytic strategies for making sense of available process evaluation data.

Report generation is an essential component of a comprehensive process evaluation effort. If key stakeholders are involved in prioritizing the process evaluation questions to be addressed, it makes sense to discuss what type of reports they would like to see generated at that same point in the process. Getting end users involved in the process of creating practical reports will increase buy-in for all aspects of the data collection process. Evaluators typically have flexibility in the format and "look" of the report. End users can help clarify the best ways to create practical, easy-to-use reports.

Reports can be generated to assist with project management tasks, to motivate the staff to collect and enter data on a timely basis, and to identify possible problems in collecting or coding data. In addition, reports may be shared with study participants to help them stay involved in the study or to meet data collection

timelines. Moreover, timely feedback on participation, or the extent to which the intervention has been delivered as planned, can generate great enthusiasm for the project among participants and other key stakeholders. Corrective feedback when problems are identified is also possible when reports are generated on a timely basis. Reports are also great tools for sharing success stories; everyone involved will benefit if reports are used to share good news, progress, or special achievements that would have gone unrecognized otherwise.

Summary

Process evaluation is a growing and important component of a comprehensive evaluation effort. This overview has briefly reviewed the history of process evaluation as it has been applied to public health interventions, has identified gaps in knowledge concerning process evaluation, has recommended definitions and the consistent use of key process evaluation components, and has presented a strategy for designing and implementing effective process evaluation efforts. Improvements in the conceptual thinking about process evaluation will drive changes in the design, measurement, data collection, analysis, and timely reporting of the process evaluation data. Process evaluation data can be used to answer important questions that advance understanding of how and why public health interventions work, which has positive implications for both research and practice.

References

Abrams, D. B., and others. "Cancer Control in the Workplace: The Working Well Trial." *Preventive Medicine,* 1994, *23,* 15–27.

Baranowski, T., and Stables, G. "Process Evaluation of the 5-a-Day Projects." *Health Education and Behavior,* 2000, *27*(2), 157–166.

Baranowski, T., Anderson, C., and Carmack, C. "Mediating Variable Framework in Physical Activity Interventions: How Are We Doing? How Might We Do Better?" *American Journal of Preventive Medicine,* 1998, *15*(4), 266–297.

Baranowski, T., and others. "Theory as Mediating Variables: Why Aren't Community Interventions Working as Desired?" *Annals of Epidemiology,* 1997, *7*(S7), S89–S95.

Bartholomew, L. K., Parcel, G. S., Kok, G., and Gottlieb, N. H. *Intervention Mapping: Designing Theory and Evidence-Based Health Promotion Programs.* Mountain View, Calif.: Mayfield, 2001.

Basch, C. E., and others. "Avoiding Type III Errors in Health Education Program Evaluations: A Case Study." *Health Education Quarterly,* 1985, *12*(4), 315–331.

Corbett, K., Thompson, B., White, N., and Taylor, R. "Process Evaluation in Community Intervention Trial for Smoking Cessation (COMMIT)." *International Quarterly of Community Health Education,* 1991, *11*(3), 291–309.

Dobson, D., and Cook, T. J. "Avoiding Type III Errors in Program Evaluation: Results from a Field Experiment." *Evaluation and Program Planning,* 1980, *3,* 269–276.

Dusenbury, L., Brannigan, R., Falco, M., and Hansen, W. B. Unpublished manuscript. Tanglewood Research, Greensboro, N. C., November 2001.

Earp, J. A., and Ennett, S. T. "Conceptual Models for Health Education Research and Practice." *Health Education Research,* 1991, *6*(2), 163–171.

Elder, J. P., and others. "Process Evaluation of Environmental Factors and Programs." *Health Education Quarterly,* 1994 (supp. 2), S107–S128.

Finnegan, J. R., Murray, D. M., Kurth, C., and McCarthy, P. "Measuring and Tracking Education Program Implementation: The Minnesota Heart Health Program Experience." *Health Education Quarterly,* 1989, *16,* 77–90.

Fisher, E. B. "The Results of the COMMIT Trial." *American Journal of Public Health,* 1995, *85,* 159–161.

Flora, J. A., and others. "A Community Education Monitoring System: Methods from the Stanford Five-City Project, the Minnesota Heart Health Program, and the Pawtucket Heart Health Program." *Health Education Research,* 1993, *8,* 81–95.

Glanz, K., Lewis, F. M., and Rimer, B. K. "Linking Theory, Research, and Practice." In K. Glanz, F. M. Lewis, and B. K. Rimer (eds.), *Health Behavior and Health Education: Theory, Research, and Practice.* San Francisco: Jossey-Bass, 1997.

Glasgow, R. E., McCaul, K., and Fisher, K. J. "Participation in Worksite Health Promotion: A Critique of the Literature and Recommendations for Future Practice." *Health Education Quarterly,* 1993, *20*(3), 391–408.

Glasgow, R. E., Vogt, T. M., and Boles, S. M. "Evaluating the Public Health Impact of Health Promotion Interventions: The RE-AIM Framework." *American Journal of Public Health,* 1999, *89*(9), 1322–1327.

Glass, G. V. (ed.). *Evaluation Studies Review Annual.* Vol. 1. Thousand Oaks, Calif.: Sage, 1976.

Green, L. W. "Evaluation and Measurement: Some Dilemmas for Health Education." *American Journal of Public Health,* 1977, *67,* 151–156.

Green, L. W., and Kreuter, M. W. *Health Promotion Planning: An Ecological Approach.* Mountain View, Calif.: Mayfield, 1999.

Green, L. W., Kreuter, M. W., Deeds, S. G., and Partridge, K. B. *Health Education Planning: A Diagnostic Approach,* Mountain View, Calif.: Mayfield, 1980.

Guttentag, M., and Saar, S. (eds.). *Evaluation Studies Review Annual.* Vol. 2. Thousand Oaks, Calif.: Sage, 1977.

Harachi, T. W., and others. "Opening the Back Box: Using Process Evaluation Measures to Assess Implementation and Theory Building." *American Journal of Community Psychology,* 1999, *27,* 715–735.

Heimendinger, J., and others. "The Working Well Trial: Baseline Dietary and Smoking Behaviors of Employees and Related Worksite Characteristics." *Preventive Medicine,* 1995, *24,* 180–193.

McGraw, S. A., and others. "Methods in Program Evaluation: The Process Evaluation System of the Pawtucket Heart Health Program." *Evaluation Review,* 1989, *13,* 459–483.

McGraw, S. A., and others. "Design of Process Evaluation Within the Child and Adolescent Trial for Cardiovascular Health (CATCH)." *Health Education Quarterly,* 1994, *2* (supp.), S5–S26.

McLeroy, K. R., Bibeau, D., Steckler, A., and Glanz, K. "An Ecological Perspective on Health Promotion Programs." *Health Education Quarterly,* 1988, *15,* 351–377.

National Institutes of Health. *Qualitative Methods in Health Research: Opportunities and Considerations in Application and Review.* Washington, D.C.: National Institutes of Health. Office of Behavioral and Social Sciences Research, 2001.

Patton, M. Q. *Utilization-Focused Evaluation.* (3rd ed.) Thousand Oaks, Calif.: Sage, 1997.

Perry, C. L., and others. "The Child and Adolescent Trial for Cardiovascular Health (CATCH): Intervention, Implementation, and Feasibility for Elementary Schools in the United States." *Health Education and Behavior,* 1997, *24*(6), 716–735.

Pirie, P. L., and others. "Program Evaluation Strategies for Community-Based Health Promotion Programs: Perspectives from the Cardiovascular Disease Community Research and Demonstration Studies." *Health Education Research,* 1994, *9*(1), 23–36.

Resnicow, K., and others. "How Best to Measure Implementation of Health Curricula: A Comparison of Three Measures." *Health Education Research,* 1998, *13,* 239–250.

Rezmovic, E. L. "Program Implementation and Evaluation Results: A Reexamination of Type III Error in a Field Experiment." *Evaluation and Program Planning,* 1982, *5,* 111–118.

Sorensen, G., and others. "Worksite-Based Cancer Prevention: Primary Results from the Working Well Trial." *American Journal of Public Health,* 1996, *86,* 939–947.

Steckler, A., and others. "Toward Integrating Qualitative and Quantitative Methods: An Introduction." *Health Education Quarterly,* 1992, *19*(1), 1–8.

Stone, E. J., McGraw, S. A., Osganian, S. K., and Elder, J. P. "Process Evaluation in the Multicenter Child and Adolescent Trial for Cardiovascular Health (CATCH)." *Health Education Quarterly,* 1994 (supp. 2), 1–148.

Struening, E. I., and Guttentag, M. (eds.). *Handbook of Evaluation Research.* Thousand Oaks, Calif.: Sage, 1975.

Suchman, E. A. *Evaluative Research: Principles and Practice in Public Service and Social Action Programs.* New York: Russell Sage Foundation, 1967.

Susser, M. "Editorial: The Tribulations of Trials—Interventions in Communities." *American Journal of Public Health,* 1995, *85,* 156–158.

Tashakkori, A., and Teddlie, C. *Mixed Methodology: Combining Qualitative and Quantitative Approaches.* Thousand Oaks, Calif.: Sage, 1998.

Weiss, C. *Evaluation.* (2nd ed.) Englewood Cliffs, N.J.: Prentice Hall, 1998.

Windsor, R. A., Baranowski, T., Clark, N., and Cutter, G. *Evaluation of Health Promotion and Education Programs.* Mountain View, Calif.: Mayfield, 1984.