

1

Program Evaluation: A Prelude

A Reader's Guide to Chapter 1

What Is Program Evaluation?

The Program or Intervention
Program Objectives and Outcomes
Program Characteristics
Program Impact
Program Costs
Program Quality
Program Value

Evaluation Methods

Posing evaluation questions and hypotheses; deciding on evidence of merit: effectiveness, quality, value; designing the evaluation; selecting participants for the evaluation; collecting data on program merit; managing data so that it can be analyzed; analyzing data to decide on program merit; reporting on effectiveness, quality, and value.

Who Uses Evaluations?

Baseline Data, Formative Evaluation, and Process Evaluations

Summative Evaluation

Qualitative Evaluation

Mixed-Methods Evaluation

Participatory and Community-Based Evaluation

Evaluation Frameworks and Models

PRECEDE-PROCEED, RE-AIM, Centers for Disease Control Framework, Practical Program Evaluation, and Logic Models

Evaluation Reports Online

Exercises

Suggested References and Readings

Suggested Websites

What Is Program Evaluation?

Program evaluation is an *unbiased* exploration of a program's *merits*, including its *effectiveness*, *quality*, and *value*. An *effective* program provides substantial benefits to individuals, communities, and societies and these benefits are greater than their human and financial costs. A *high-quality* program meets its users' needs and is based on sound theory and the best available research evidence. A program's *value* is measured by its worth to individuals, the community, and society.

The Program or Intervention

At the core of an evaluation is a *program*, *intervention*, *treatment*, or *policy*. Programs, interventions, and treatments are systematic efforts to achieve explicit objectives for improving health, education, and well-being. They occur in all fields, including medicine, education, and business and law and involve individuals, communities, and society. A program may be relatively small (e.g., a course in web design for seniors in two high schools; a new community health center for persons over 75 years of age), or relatively large (e.g., a nation's health plan or a global initiative to eliminate poverty). Programs can take place in differing geographic and political settings, and they vary in their purposes, structures, organization, and constituents. A policy is a system of laws, regulatory measures, courses of action, and funding priorities associated with private and public governing agencies, trusts, and boards. A policy can be used to support or discontinue programs, interventions, treatments, and evaluations.

Are the following objectives likely to come from program evaluations?

Objective: To determine the effectiveness of an abuse-prevention curriculum designed to empower women with mental retardation to become effective decision makers.

The answer is yes. The evaluation is for an abuse-prevention program.

What about this objective?

Objective: To investigate the effectiveness of acupuncture compared with sham acupuncture and with no acupuncture in patients with migraine.

This objective is also likely to come from an evaluation. The investigators compare three interventions: acupuncture, sham acupuncture, and no acupuncture. (No acupuncture is considered an intervention because the absence of acupuncture does not mean the absence of anything at all. The no acupuncture group may be on medication or other forms of therapy.)

Finally, is this objective likely to come from a program evaluation?

Objective: To assess whether the Acquittal Project has effectively exonerated wrongfully convicted people through DNA testing.

Similarly, this objective is likely to come from a program evaluation. The program, the Acquittal Project, is designed to achieve a specific objective: to exonerate wrongfully convicted people.

Now, consider whether this objective is typical of program evaluations.

Objective: To clarify the concepts of coping with pain and quality of life, and to present a literature review of the strategies that children with recurrent headaches use to cope with their pain; the impact of recurrent headaches on children's quality of life; and the influence of personal characteristics (i.e., age, family support) on headaches, coping, and quality of life in children.

No. This objective is not typical of program evaluations. The researchers are not planning to investigate the effectiveness, quality, or value of a specific program.

Program Objectives and Outcomes

A program's *objectives* are its anticipated *outcomes*—for example, to improve skills in primary school math, prevent gambling problems in adolescents, or provide increased access to social services for young families. The aim of a major program evaluation is to provide data on a program's progress toward achieving its objectives.

The ultimate, desired outcomes of most social programs are usually lofty goals, such as providing efficient, high-quality health care and education to all people. These outcomes are often difficult to measure (or achieve) because of a lack of consensus on definitions, and because evaluators rarely have sufficient time to observe and assess the programs accurately. As a result, many evaluations focus on the extent to which programs achieve more easily measured goals and objectives, such as improving 4th grade reading skills, helping adolescents stop smoking, or teaching older adults to become better consumers of online health information. The idea is that if programs can foster the achievement of these interim objectives, accomplishment of the loftier outcomes may eventually become possible.

Program Characteristics

Evaluations answer questions about a program's characteristics and social and cultural contexts. Typical questions of this type include: Is the individual, communal, or societal need for the program explained? Is it justified? Who is responsible for program development and program funding? Which principles of learning, social justice, or health-behavior change guide the program development? Was the program implemented as planned? What were the barriers to

implementation? Were changes made to the original objectives? If so, why were the changes needed, and who made them? What is the duration of the program? What is its content?

Program Impact

Evaluators often examine a program's impact—that is, the scope of its effects, the duration of its outcomes, and the extent of its influence in varying settings and among different groups of people. For example, consider the evaluations of two programs to improve mental health status. Evaluation A reports that Program A improved mental health status for its participants, and that the gains were sustained for at least 3 years; moreover, when Program A was tried out in another country, participants in that country also improved. Evaluation B reports that Program B also improved mental health status and sustained the improvement for 3 years, however, for fewer participants. When program B was tested in another country, the evaluators found there were few gains. The evaluators of Programs A and B agreed that Program A had greater impact because its benefits reached more people over a longer period of time.

Program Costs

Evaluations are also concerned with how much a program costs and the relationship of cost to effectiveness and benefit. Program costs include any risks or problems that adversely affect program participants. For instance, program participants in a group therapy program may feel embarrassed about revealing personal information, or they may become unexpectedly ill from the treatment being evaluated. Program costs also include the financial costs of facilities, staff, and equipment. Typical questions about costs include: If two programs achieve similar outcomes, which one is least costly? For each dollar spent, how much is saved on future use of services?

Program Quality

High-quality programs meet their users' needs and are based on accepted theories of human behavior and the best available research evidence. They have sufficient funding to ensure that their objectives are achieved and have strong leadership, trained staff, and a supportive environment.

Commonly asked questions about program quality include:

- Has the program been studied systematically before implementation so that its risks and benefits are predictable?
- Is the program grounded in theory or supported by the best available research?

- Does the program provide a safe, healthy, and nurturing environment for all participants?
- Is the infrastructure well-developed, and is the fiscal management sound?
- How well does the program develop and nurture positive relationships among staff, participants, parents, and communities?
- Does the program recruit, hire, and train a diverse staff who value each participant and can deliver services as planned at the highest level?
- Has the program established a partnership with communities in order to achieve program goals?
- Does the program have a coherent mission and a plan for increasing capacity so that the program is sustained or continues to grow?
- Is a system in place for measuring outcomes and using that information for program planning, improvement, and evaluation?
- Are resources appropriately allocated so that each component of the program and its evaluation are likely to produce unbiased and relevant information?

Program Value

Value is defined as the importance, worth, or usefulness of something. The word “evaluation” in program evaluation implies that the discipline’s purpose is to analyze and judge the value of a program. The term value is subjective, and the whole enterprise of program evaluation is based on identifying strategies to minimize the subjectivity—bias—that can consume the process of analyzing or judging a program’s merit.

Despite the term’s subjectivity, most evaluators agree that “value” should be defined to suit the recipients of services (students, patients) rather than the suppliers (teachers, nurses, physicians, social workers, psychologists, funders, policy makers). Typical evaluation questions about program value include:

- Do the program’s risks or benefits outweigh its costs?
- Does the program meet a need that no other service can or does provide?
- Does the program provide the most improvement and benefits possible with its available resources?

Evaluation Methods

Program evaluators use many of the methods social and health scientists, educators, and psychologists rely on to gather *reliable* and *valid* evidence. These methods typically include the following activities:

1. Selecting questions for formulating hypotheses about program and participant characteristics, outcomes, impact, and costs
2. Deciding on evidence of program merit: effectiveness, quality, value

3. Designing the evaluation
4. Selecting participants for the evaluation
5. Collecting data on program merit
6. Managing data so that it can be analyzed
7. Analyzing data to decide on merit
8. Reporting the results

Evaluation Questions and Hypotheses

Evaluations directly or indirectly answer questions about a program's implementation, outcomes, impact, and costs. Some evaluations design their studies to test hypotheses rather than ask questions, although the two are related.

Typical evaluation questions and hypotheses include:

- **Question:** Did the program achieve its goals and objectives?
Hypothesis: When compared to a similar program, program A will achieve significantly more goals and objectives than Program B.
- **Question:** Which program characteristics (e.g., theoretical foundation, use of technology, funding) are most likely responsible for the best and worst outcomes?
Hypothesis: The online course will achieve significantly better results than the traditional course.
- **Question:** For which individuals or groups was the program most effective?
Hypothesis: Boys will learn as quickly as girls.
- **Question:** How applicable are the program's objectives and activities to other participants in other settings?
Hypothesis: Participants in other schools will do as well as participants in the local school.
- **Question:** How enduring were the program's outcomes?
Hypothesis: Participants will maintain their gains over a five-year period after the program's conclusion.
- **Question:** What are the relationships among the costs of the program and its outcomes?
Hypothesis: For every dollar spent, there will be at least one reading level improvement.
- **Question:** To what extent did social, political, and financial support influence the program's outcomes and acceptability?
Hypothesis: Local support is associated with greater program satisfaction.

- **Question:** Is the program cost-effective?
Hypothesis: New Program A and old Program B achieve similar outcomes, but Program A costs less to implement.
- **Question:** Were there any unanticipated outcomes (beneficial as well as harmful)? This is a research question. No hypothesis is associated with it because the evaluators have no basis for stating one. They do not have a theory or any research evidence to support assumptions about outcomes.

Some evaluations answer just a few questions or test just a few hypotheses, while others answer many questions and test numerous hypotheses.

Evidence of Merit: Effectiveness, Quality, Value

Evidence of merit consists of the facts and information that demonstrate a program's effectiveness, quality, and value. Consider each of the following six possible indications of merit for a program whose objective is "to improve children's dietary and other health habits":

1. Testimony from children in the program (and from their parents and teachers) that their habits have improved.
2. The evaluator's observations of improved health habits (e.g., through studies of children's choices of snacks during and between meals).
3. Proof of children's improved health status found in physical examinations by a nurse practitioner or a physician.
4. The evaluator's finding of statistically significant differences in habits and in health status between children who are in the program compared with children who are not. Children in the program do significantly better.
5. The evaluator's finding of statistically significant and *sustained* differences in habits and in health status between children who are in the program compared with children who are not. Children in the program continue to do significantly better over time.
6. Statistical and qualitative evidence that Program A achieves the same aims as Program B, and demonstrates that it is less costly.

Which of these indications of merit is best? How much and what types of evidence are needed? Merit is a subjective term: It varies across individuals, communities, institutions, and policy makers.

The evaluator's challenge is to identify evidence that is unbiased, convincing to the evaluation's users and funders, and possible to collect with the available resources. For instance, evaluators are unlikely to be able to provide data on sustained program benefits in evaluations that are scheduled to last a year or less even if they have the resources, and even if that is what the users indicate they want. Bias in evaluations often comes from faulty research methods or failure to properly implement the program or the evaluation.

Many evaluators consult and form partnerships with users and funders to ensure that the evidence they plan to collect is appropriate and likely to meet expectations. Evaluators find that working with clients typically creates mutual respect, promotes client cooperation with data collection during the evaluation's implementation, and improves the usefulness of the results.

Designing the Evaluation

An *evaluation's design* is its structure. Evaluators do their best to design a project so that any benefits that appear to result from the program are real and not influenced by expectation or preference. A standard evaluation design includes comparing the participants in a new program with participants in an alternative program. The comparison can occur once or several times. For example, suppose five universities plan to participate in an evaluation of a new program to teach the basic principles of program evaluation to Education Corps trainees. In designing the study, the evaluator has to answer questions like these:

- Which program is a fair comparison to the “new” one? Evaluators sometimes compare the new program to an already existing one with similar characteristics, or they compare the new program to “usual practice.” If the resources are available, they may compare the new program to an already existing program and also to usual practice. Another option is to compare two versions of the new program and usual practice. For instance, a crime prevention program for teens may compare a smartphone app with peer counseling [version 1], the same app without the counseling [version 2], and the school's usual monthly webinar [usual practice].
- Which criteria are appropriate for including institutions? (size, resources, location)
- Which criteria are appropriate for excluding institutions? (unwillingness to implement program as planned; lack of staff commitment to the evaluation)
- What should I measure? (understanding of principles of program evaluation, application of the principles when designing an evaluation)
- When should I measure learning? (before and after program participation? How long after program participation?)

Selecting Participants for the Evaluation

Suppose you are asked to evaluate a program to provide school-based mental health services to children who have witnessed or have been victims of violence in their communities. Here are some questions you need to ask:

- Who should be included in the evaluation? (Which grades should be included? How much exposure to violence should eligible children have?)
- Who should be excluded? (Should children be excluded if, in the opinion of the mental health clinician, they are probably too disruptive to participate in the program's required group therapy sessions?)
- How many children should be included? (What is a sufficient number of participants to allow the evaluation to detect change in children's behavior if the program is effective?)

Collecting Data on Program Merit

Conclusions about a program's merit, quality, and value come from the data an evaluator collects to answer questions and test hypotheses. Data collection includes

- identifying the variables (individual knowledge, attitudes, or behaviors; community practices; social policies) that are the program's target outcomes;
- identifying the characteristics of the participants who will be affected by the program (men between the ages of 45 and 54; rural and urban communities);
- selecting, adapting, or creating measures of the variables (knowledge tests, direct observations of behavior; analysis of legal documents);
- demonstrating the reliability (consistency) and validity (accuracy) of the measures;
- administering the measures; and
- analyzing and interpreting the results.

Some common measures or sources of evaluation data are

- literature reviews;
- archival reviews (school and medical records);
- existing databases, such as those maintained by governments and schools;
- self-administered questionnaires (including in-person, mailed, and online surveys);
- interviews (in-person and on the telephone);
- achievement tests;
- observations;
- physical examinations; and
- hypothetical vignettes or case studies.

Managing Data So That It Can Be Analyzed

Data management includes the following activities

- Drafting an analysis plan that defines the variables to be analyzed
- Creating a codebook
- Establishing the reliability of the coders or coding
- Entering data into a database and validating the accuracy of the entry
- Reviewing the evaluation's database for incomplete or missing data
- Cleaning the data
- Creating the final data set for analysis
- Storing and archiving the data set and its operations manual

Analyzing Data to Decide on Program Merit

Data analysis consists of the descriptive (qualitative) and statistical (quantitative) methods used to summarize information about a program's effectiveness. The choice of which method of analysis to use is dependent on several considerations.

- The characteristics of the evaluation question and evidence of effectiveness (Do the questions ask about differences over time among groups or about associations between program characteristics and benefits? If the questions ask about differences, then a statistical method that tests for differences is needed. If the questions ask about associations, then different statistical methods are probably warranted.)
- How the variables are expressed statistically: categorically ("did or did not pass the test"); with ordinals (Stage I, II, or III of a disease; ratings on a scale ranging from 1 to 5); or numerically (average scores on a mathematics test)
- How the variables are expressed qualitatively (e.g., themes from an analysis of a focus group)
- The reliability and validity of the data

Reporting on Effectiveness, Quality, and Value

An evaluation report describes, justifies, and explains the purposes of the evaluation, the program, the setting, and the methods that are used to arrive at unbiased conclusions about effectiveness, quality, and value. The methods include descriptions and explanations of the evaluation question and evidence-selection processes, the research design, sampling strategy, data collection, and data analysis. The report also states the

results and arrives at conclusions about program merit based on the evidence. Many scholarly journals also require proof that the evaluation respected and protected participants from risk. This is done by asking for the evaluator to state that the evaluation received a formal review by an ethics board.

Evaluation reports may be oral or may be presented in written form, as books, monographs, or articles. Consider the summaries in Example 1.1.

Example 1.1 Summaries of Program Evaluations

1. Evaluation of a Healthy Eating Program for Professionals Who Care for Preschoolers (Hardy, King, Kelly, Farrell, & Howlett, 2010)

Background. Early childhood services are a convenient setting for promoting healthy eating and physical activity as a means of preventing overweight and obesity. This evaluation examined the effectiveness of a program to support early childhood professionals in promoting healthy eating and physical activity among children in their care.

Setting and Participants. The evaluation included 15 intervention and 14 control preschools with 430 children whose average age was 4.4 years.

Methods. Preschools were randomly allocated to the intervention or a control program. The evaluators did not know which schools were in each program. They collected data before and after program implementation on children's lunchbox contents; fundamental movement skills (FMS); preschool policies, practices, and staff attitudes; knowledge and confidence related to physical activity; healthy eating; and recreational screen time.

Results. Using statistical methods, the evaluators found that over time, FMS scores for locomotion and object control, and total FMS scores significantly improved in the intervention group compared with the control group by 3.4, 2.1, and 5.5 points (respectively). The number of FMS sessions per week increased in the intervention group compared with the control group by 1.5. The lunchbox audit showed that children in the intervention group significantly reduced sweetened drinks by 0.13 servings.

Conclusion. The findings suggest that the program effectively improved its targeted weight-related behaviors.

2. A Psychological Intervention for Children With Symptoms of Posttraumatic Stress Disorder (Stein et al., 2003)

Context. Are psychological interventions effective for children with symptoms of post-traumatic stress disorder (PTSD) resulting from personally witnessing or being personally exposed to violence?

Objective. To evaluate the effectiveness of a collaboratively designed school-based intervention for reducing children's symptoms of PTSD and depression resulting from exposure to violence.

Design. A randomized controlled trial conducted during one academic school year.

Setting and Participants. Sixth-grade students who reported exposure to violence and had clinical levels of symptoms of PTSD at two large middle schools in a large U.S. city.

Intervention. Students were randomly assigned to a 10-session standardized cognitive-behavioral therapy early intervention group (61 students), or to a wait-list delayed intervention comparison group (65 students) conducted by trained school mental health clinicians.

Main Outcome Measures. Students were assessed before the intervention and 3 months after the intervention on measures assessing child reported symptoms of PTSD and depression.

Results. The evaluation found that compared with the wait-list delayed intervention group (no intervention), after 3 months of intervention, students who were randomly assigned to the early intervention group had significantly lower scores on symptoms of PTSD and depression.

Conclusion. A standardized 10-session cognitive-behavioral group intervention can significantly decrease symptoms of PTSD and depression in students who are exposed to violence and can be effectively delivered on school campuses by trained school-based mental health clinicians.

By examining the summaries in Example 1.1, it is apparent that *doing* an evaluation involves the use of multiple skills in research design, statistics, data collection, and interpretation. Since very few individuals have perfected all of these skills, evaluators almost always work in teams, as is illustrated in Example 1.2.

Example 1.2 Program Evaluations as an Interdisciplinary Discipline

- A 4-year evaluation of a new workplace literacy program was conducted by a team composed of two professional evaluators, a survey researcher, a statistician, and two instructors. The evaluation team also consulted an economist and an expert in information science.
- A 3-year evaluation of a 35-project program to improve access to and use of social services for low-income women relied on two professional evaluators, a social worker, an epidemiologist, a nurse practitioner, and an economist.

- An evaluation of a program using nurses to screen community dwelling elderly individuals for hypertension, vision, and hearing disorders relied on a nurse, a nurse practitioner, a statistician, and a professional evaluator.

Who Uses Evaluations?

At least seven different groups use the information that results from program evaluations:

1. Government agencies
2. Program developers (a director of a community health clinic, a curriculum committee, or a nursing school's curriculum committee)
3. Communities (geographically intact areas, such as a city's "skid row"; people with a shared health-related problem, such as HIV/AIDS; or individuals with a common culture, such as Armenians or Latinos)
4. Policy makers (i.e., elected officials; the school board)
5. Program funders (philanthropic foundations or trusts and the various agencies of the National Institutes of Health)
6. Students, researchers, and other evaluators (specific to schools and universities, government agencies, businesses, and public agencies)
7. Individuals interested in new programs

Baseline Data, Formative Evaluation, and Process Evaluation

The need for a program is demonstrated when there is a gap between what individuals or communities need and their current services. *Baseline data* are collected to document program participants' status before they begin the program. *Interim data*, which are collected during the course of the program, show the program's progress in meeting the participants' needs. These interim data are used to evaluate the program while in its formative stage.

Baseline Data

Baseline data are collected before the start of the program to describe the characteristics of participants (e.g., their social, educational, health status, and demographic features,

such as age), information that is important later on when the evaluator is interpreting the effects of the program. Example 1.3 illustrates some of the reasons program evaluators collect baseline data.

Example 1.3 Baseline Data and Program Evaluation

The Agency for Drug and Alcohol Misuse has published extensive guidelines for identifying and counselling adolescents whose alcohol use is interfering with their everyday activities, such as attendance at school. An evaluation of the guideline's effectiveness is being conducted nationwide. Before the evaluators begin the formal evaluation process, they collect baseline data on the extent to which health care professionals in different settings (e.g., schools and community clinics) already follow the practices recommended by the guidelines, the prevalence of alcohol misuse among adolescents in the communities of interest, and the number of adolescents that are likely to use services in the evaluation's proposed duration of 3 years.

Interim Data and Formative Evaluation

In *formative evaluation*, data are collected after the start of a program, but before its conclusion—for example, 12 months after the beginning of a 3-year intervention. An evaluator collects interim data to describe the progress of the program while it is still developing or “forming.” Formative evaluation data are mainly useful to program developers and funders. Program developers and funders may want to know if a new program is *feasible* as is, or whether it needs to be improved. A feasible program is one that can be implemented according to plan and is likely to be beneficial.

Data from formative evaluations are always preliminary, and require cautious interpretation. Example 1.4 illustrates why evaluators need to take care in interpreting formative findings.

Example 1.4 Formative Evaluation and Interim Data: Proceed With Caution

In a 3-year study of access to prenatal care, the results of a 14-month formative evaluation found that three of six community clinics had opened on schedule and were providing services to needy women exactly as planned. Preliminary data also revealed that 200

women had been served in the clinics and that the proportion of babies born weighing less than 2,500 grams (5.5 pounds) was 4%, well below the state's average of 6%. The evaluators concluded that progress was definitely being made toward improving access to prenatal care. After 3 years, however, the evaluation results were quite different. The remaining three scheduled clinics had never opened, and one of the original three clinics had closed. Many fewer women were served than had been anticipated, and the proportion of low birth weight babies was 6.6%.

As this example shows, data from a formative evaluation can be misleading. Good interim results may be exhilarating, but poor ones can adversely affect staff morale. With programs of relatively short duration—say, 2 years or less—the collection of interim data is expensive and probably not very useful. Consider the evaluation described in Example 1.5.

Example 1.5 Questions Asked in a Formative Evaluation of a Program for Critically Ill Children

Many experts agree that the emergency medical services needed by critically ill and injured children differ in important ways from those needed by adults. As a result, a number of health regions have attempted to reorganize their emergency services to provide better care to children. One region commissioned a 3-year evaluation of its program. It was specifically concerned with the characteristics and effectiveness of a soon to be implemented intervention to prevent transfers from adult inpatient or intensive care units and to maximize quality of care for children with cardiopulmonary arrest in hospital emergency departments and intensive care units.

In planning the evaluation, the evaluators decided to check a sample of medical records in 15 of the state's 56 counties to see whether sufficient information was available for them to use the records as a main source of data. Also, the evaluators planned to release preliminary findings after 12 months, which involved reviews of records as well as interviews with physicians, hospital administrators, paramedics, and patients' families. An expert's review of the evaluation's design raised these questions for the evaluators:

1. Does the description of this evaluation as a "3-year evaluation" mean that there will be 3 years of data collection, or do the 3 years include evaluation planning, implementation, and reporting as well as data collection? Assume interim data are promised in a year. Can you develop and validate medical record review forms in time to collect enough information to present meaningful findings?

2. Can you develop, validate, and administer the survey forms in the time available?
 3. To what extent will the interim and preliminary analyses answer the same or similar questions? If they are very different, will you have sufficient time and money to effectively conduct both?
 4. Will a written or oral interim report be required? How long will that take to prepare?
-

Some program evaluations are divided into two phases. In Phase 1, the evaluation is designed to focus on feasibility and improvement, and in Phase 2, it focuses on effectiveness, cost, and value. Some funders prefer to have Phase 1 done internally (i.e., by the participating schools or clinics), and Part 2 done externally (by professional evaluation consultants). External evaluations are presumed to be more objective and less inclined to bias than internal evaluations. Increasingly, however, many agencies and institutions involve program evaluators in both study phases for continuity and efficiency.

Process or Implementation Evaluation

A *process evaluation* is concerned with the extent to which planned activities are implemented, and its findings may be reported at any time. Process evaluations are almost always useful. For example, in an evaluation of three interventions to increase the rates at which women returned to follow up on Pap smears, a process evaluation concluded that implementation of the intervention protocols was less than perfect and thus introduced a bias into the results of the outcome evaluation. This study is the subject of Example 1.6.

Example 1.6 Process or Implementation Evaluation: Follow-Up of Abnormal Pap Smears

During the course of a 2-year evaluation, all women were to be surveyed at least once regarding whether they received the program and the extent to which they understood its purposes and adhered to its requirements. Telephone interviews after 18 months revealed that 74 of 100 women (74%) in the slide-tape intervention had seen the entire 25-minute presentation, 70 of 111 (63%) had received mailed reminders from their physicians' offices to come back for another Pap smear, and 32 of 101 (about 32%) had received phone calls from their physicians' offices. These findings helped explain the apparent failure of the third intervention to achieve positive results when compared with the other two.

Summative Evaluation

Summative evaluations are historical studies that are compiled after the program has been in existence for a while (say, two years), or all program activities have officially ceased. These evaluations sum up and qualitatively assess the program's development and achievements.

Summative evaluations are descriptive rather than experimental studies. Funding sources sometimes request these evaluations because summative reports usually contain details on how many people the program served, how the staff was trained, how barriers to implementation were overcome, and if participants were satisfied and likely to benefit. Summative evaluations often provide a thorough explanation of how the program was developed and the social and political context in which the program and its evaluation were conducted.

Qualitative Evaluation

Qualitative evaluations collect data through interviews, direct observations, and review of written documents (for example, private diaries). The aim of these evaluations is to provide personalized information on the dynamics of a program and on participants' perceptions of the program's outcomes and impact.

Qualitative evaluation is useful for examining programs where the goals are in the process of being defined, and for testing out the workability of particular evaluation methods. Because they are personalized, qualitative methods may add emotion to otherwise purely statistical findings and provide a means of gauging outcomes when reliable and valid measures of those outcomes are unlikely to be available in time for inclusion in the evaluation report.

Qualitative methods are employed in program evaluations to complement the usual sources of data (such as standardized surveys and medical record reviews, physical examinations, and achievement tests). Example 1.7 illustrates four uses of qualitative methods in program evaluation.

Example 1.7 Uses of Qualitative Methods in Program Evaluation

1. To evaluate the effectiveness of a campaign to get heroin addicts to clean their needles with bleach, the evaluators spend time in a heroin "shooting gallery." They do not have formal observation measures, although they do take notes. The evaluators discuss what they have seen, and although needles are being cleaned, agree that the addicts use a common dish to rinse needles and dilute the drug before shooting. The evaluators recommend that the community's program should be altered to take into account the dangers of this practice.

2. To evaluate the quality and effectiveness of an education counseling program for mentally ill adults, the evaluation team lives for 3 months in each of five different residential communities. After taping more than 250 counseling sessions, the evaluators examine the tape to determine if certain counseling approaches were used consistently. They conclude that the quality of the counseling varies greatly both within and among the communities, which helps to explain the overall program's inconsistent results.
3. To evaluate the impact of a school-based health program for homeless children, the evaluators teach a cohort of children to keep diaries over a 3-year period. The evaluation finds that children in the program are much more willing to attend to the dangers of smoking and other drug use than are children in schools without the program. The evaluators do an analysis of the content of the childrens' diaries. They find that children in the program are especially pleased to participate. The evaluators conclude that the children's enjoyment may be related to the program's positive outcomes.
4. An evaluation of the impact on the county of a program to improve access to and use of prenatal care services asks "opinion leaders" to give their views. These people are known in the county to have expertise in providing, financing, and evaluating prenatal care. The interviewers encourage the leaders to raise any issues of concern. The leaders share their belief that any improvements in prenatal care are probably due to medical advances rather than to enhanced access to services. After the interviews are completed, the evaluators conclude that major barriers to access and use continue to exist even though statistical registries reveal a decline in infant mortality rates for some groups of women.

In the first evaluation in Example 1.7, the evaluators are observers at the heroin shooting gallery. They rely on their observations and notes to come to agreement on their recommendations. In the second illustration, the evaluators tape the sessions, and then interpret the results. The interpretations come after the data are collected; the evaluators make no effort to state evaluation questions in advance of data collection. In the third illustration, diaries are used as a qualitative tool, allowing participants to say how they feel in their own words. In the fourth illustration in Example 1.7, experts are invited to give their own views; the evaluators make little attempt to require the opinion leaders to adhere to certain topics.

Mixed-Methods Evaluation

Mixed methods is most commonly interpreted as a type of research in which qualitative and quantitative or statistical data are combined within a single study. Example 1.8 outlines at least three reasons for mixing methods: to better understand experimental study results, to incorporate user perspectives into program development, and to answer differing research questions within the same study. Consider these examples.

Example 1.8 Reasons for Mixed-Methods Evaluations

1. Mixed Methods to Incorporate User Perspectives into Program Development

The study's main purpose was to develop online education to improve people's use of web-based health information. The investigators convened five focus groups and conducted in-depth interviews with 15 people to identify preferences for learning [*user perspectives*]. They asked participants questions about the value of audio and video presentations. Using the information from the groups and interviews, the investigators developed an online education tutorial and observed its usability in a small sample. Once they had evidence that the education was ready for use in the general population, they evaluated its effectiveness by using statistical methods to compare the knowledge, self-efficacy, and Internet use among two groups. Group 1 was assigned to use the newly created online tutorial, and Group 2 was given a printed checklist containing tips for wise online health information searches.

2. Mixed Methods to Answer Different Research Questions (Marczinski & Stamates, 2012; Yu, 2012)

- A. The investigators in this study want to find out if alcohol consumed with an artificially sweetened mixer (e.g., diet soft drink) results in higher breath alcohol concentrations (BrACs) compared with the same amount of alcohol consumed with a similar beverage containing sugar [*Research Question 1*]. They were also interested in determining if individuals were aware of the differences [*Research Question 2*]. BrACs were recorded, as were self-reported ratings of subjective intoxication, fatigue, impairment, and willingness to drive. Performance was assessed using a signaled go/no-go reaction time task. Based on the results, the investigators found that mixing alcohol with a diet soft drink resulted in elevated BrACs, as compared with the same amount of alcohol mixed with a sugar-sweetened beverage. Individuals were unaware of these differences, a factor that may increase the safety risks associated with drinking alcohol.
- B. A mixed-methods project was devoted to understanding college students' justification for digital piracy. The project consisted of two studies, a qualitative one and a quantitative one. Qualitative interviews were conducted to identify main themes in students' justification for digital piracy; the findings were subsequently tested in a quantitative manner using a different sample of students.

3. Mixed Methods to Better Understand Experimental Results

The investigators found that experimental program participants reported significantly more discomfort with study participation than did control program participants. This finding surprised the evaluation team. To help them understand the findings, the team conducted interviews with each of the experimental program participants and asked them about the causes of their discomfort.

Participatory and Community-Based Evaluation

A *participatory evaluation* invites representatives of the organizations and communities that will be affected by the evaluation's findings to join the evaluation team as partners in some or all of the evaluation activities. Proponents of community-based evaluations assert that when community participation is encouraged, there are at least four reasons why an evaluation's findings can be particularly useful in helping to reduce disparities in health, education, and well-being based on characteristics, such as race, ethnicity, age, sexual orientation, socioeconomic status, and geography.

1. Participation helps to improve the quality and validity of research by giving it a basis in local knowledge, culture, and history. In participatory evaluations, public concerns are viewed ecologically—that is, in their political and social context as well as in their clinical setting.
2. Including the expertise of community members enhances the relevance of the evaluation questions, the quality and quantity of data gathered, and the use of the data. Community members as well as researchers “own” the data and therefore want to see the data used.
3. Participatory evaluation projects can assist in providing community members with resources and possible employment opportunities. For example, community members can help evaluators in translating surveys and in conducting interviews.
4. Participatory evaluations can lead to improvements in the health and well-being of communities by studying and addressing important community needs and increasing community members' power and control over the research process. The community can keep the evaluators on track, preventing them from taking an approach that is too academic or theoretical.

Participatory evaluators must be skilled in working with diverse groups of individuals. They must learn how to lead meetings, encourage consensus, and inform participants about the objectives and purposes of evaluation studies in general and their own evaluations in particular. At the same time, participatory evaluators must lead the process of collecting unbiased data and interpreting those data objectively. Not all programs—no matter how well-intentioned—are effective, and even those that have positive effects may not be cost-effective, and so the participatory evaluator must be prepared to be the bearer of bad news. Participatory evaluations themselves also tend to be extremely costly because they are labor-intensive: They require individuals from the evaluation team and the community to spend time agreeing on evidence of effectiveness and assisting with technical activities, including research design, data collection, and report writing.

Example 1.9 provides illustrations of participatory evaluation in action.

Example 1.9 Participatory Evaluations in Action

1. An evaluation of a cancer control program involves the community in all phases of the project, from the development of the grant proposal through to interpretation of the data. The purpose of the project is to evaluate the effectiveness of a culturally appropriate intervention as a means of increasing breast and cervical cancer screening practice among the communities' women. The results show a community-wide impact on cancer-related knowledge, attitudes, and behaviors; increased research capabilities; and improvements to the health systems and services available to the community.
2. A mental health intervention is designed to diminish symptoms of depression in urban schoolchildren who have witnessed or participated in community violence. A group of parents assist the evaluators in developing evaluation questions, translating some of the surveys into Spanish and Russian, and collecting data from other parents. They also review the evaluation's findings and comment on them. The comments are incorporated into the final report of the intervention's effectiveness.
3. The directors of a health care clinic, interested in improving patient education, intend to organize a series of staff seminars and then evaluate whether patient education improves after all staff have attended the seminars. As part of the evaluation, the evaluation team convenes a series of four noon meetings with clinic staff to identify the nature and extent of current problems in the clinic's education for patients and to examine alternative solutions. The clinic staff agrees to form a committee to work with the evaluators and decide on evidence of effectiveness for the seminars and the patient education. The staff also agrees to advise the evaluators on questions to ask patients about their experiences at the clinic and to review and comment on the report of the evaluation's findings.

In the first illustration in Example 1.9, members of the community are actively included in all phases of the evaluation study, including the writing of the proposal for funding and the interpretation of the data. In the second instance, parents work with the evaluators on many activities, including the formulation of evaluation questions, data collection, and reporting. They are not necessarily involved in designing the evaluation (e.g., determining which children are eligible for participation and the characteristics of the control or comparative intervention) or in the data analysis. The third illustration is a participatory

evaluation because the staff and evaluators work together to decide on evidence of effectiveness, identify appropriate questions to ask patients about their experiences, and review the evaluation report.

Evaluation Frameworks and Models

The PRECEDE-PROCEED Framework

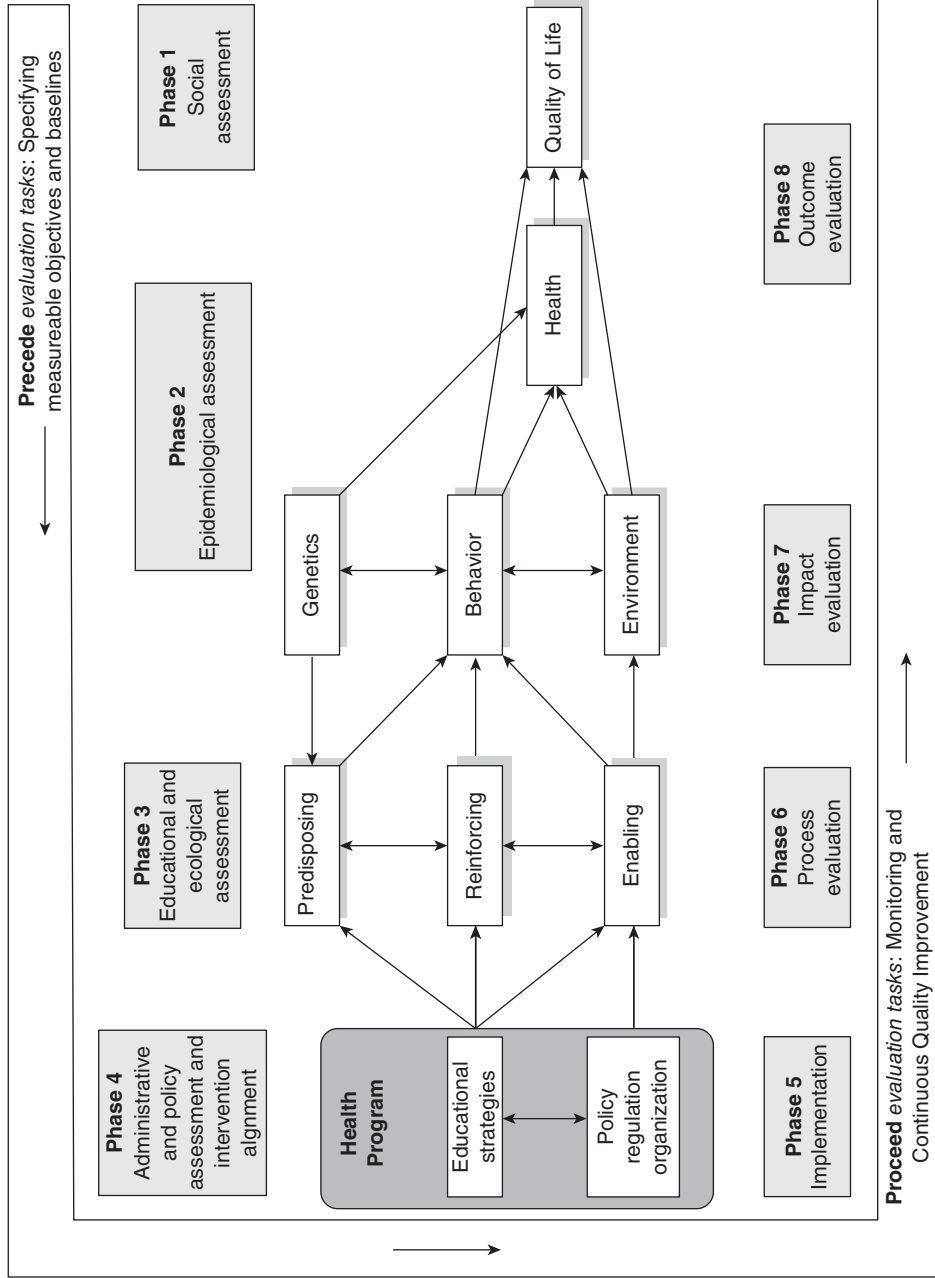
Evaluation frameworks provide guidance for program planners and evaluators, helping ensure that the evaluation's overall design considers the origins and contexts of the programs examined. One commonly used framework is the PRECEDE-PROCEED Framework (Figure 1.1).

The acronym PRECEDE stands for predisposing, reinforcing, and enabling constructs in education/environmental diagnosis and evaluation. The acronym PROCEED stands for policy, regulatory, and organizational constructs in educational and environmental development. Although developed to study how effectively programs affect changes in health behavior, PRECEDE-PROCEED is being used increasingly in the fields or disciplines of education and psychology.

The PRECEDE-PROCEED MODEL begins on the far right of the figure and moving *counter* clockwise has 8 phases.

1. *Social assessment* to determine perceptions of people's needs and quality of life. For instance, evaluators use focus groups with parents, students, and teachers to find out how to improve attendance at after school programs.
2. *Epidemiological assessment* to identify the problems that are most important in the community. For instance, evaluators conduct interviews with providers at local clinics to find out why neighborhood children visit the clinics; evaluators review county records to study inoculation rates; program planners conduct interviews with children and families to learn more about their culture, family history, and lifestyle.
3. *Educational and ecological assessment* to identify the factors that might be needed to foster changes in behaviors. These may include assessments of knowledge, beliefs, and self-efficacy (referred to as *predisposing factors*); social support (*reinforcing factors*); and programs and services necessary for good outcomes to be realized (*enabling factors*).
4. *Administrative and policy assessment and intervention alignment* to review policies and resources that facilitate or hinder program implementation.
- 5–8. *Implementation and evaluation of process, impact, and outcomes*. Using the assessments as a guide, program developers implement programs and evaluators study the programs' activities and immediate and long-term outcomes.

Figure 1.1 PRECEDE-PROCEED Model of Program Planning and Evaluation



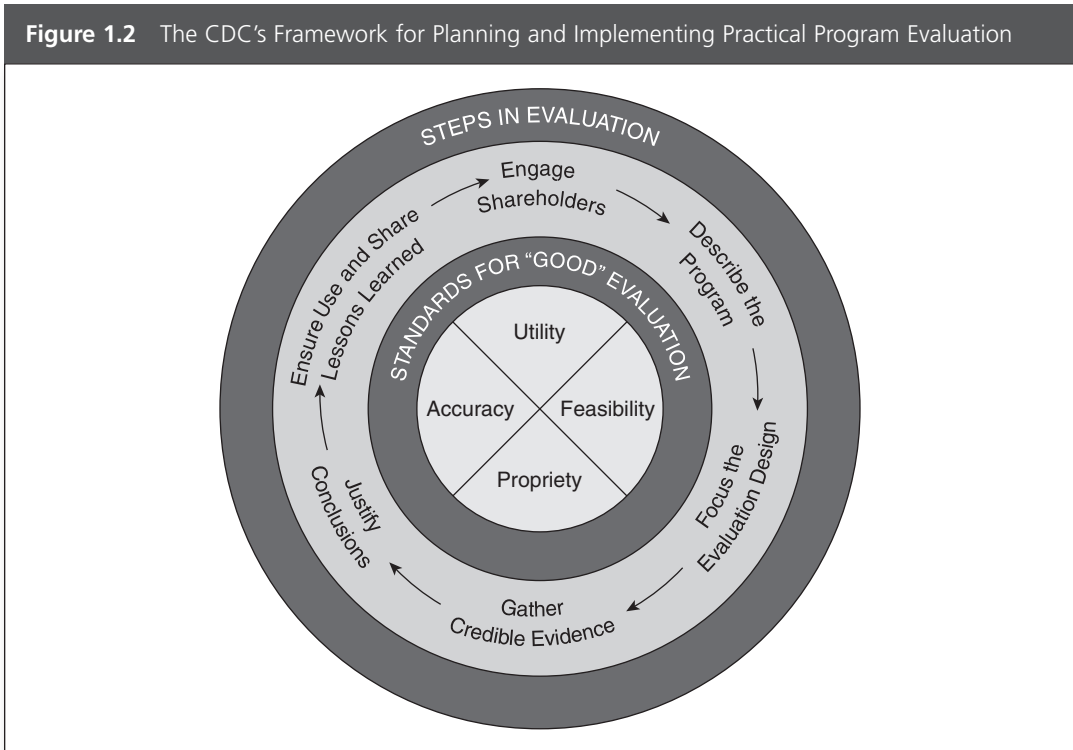
Source: Adapted by the University of Kansas from L. Green and M. Kreuter. (2005). *Health Promotion Planning: An Educational and Ecological Approach* (4th ed.). Mountain View, CA: Mayfield.

RE-AIM

The acronym RE-AIM stands for *reach, efficacy (or effectiveness), adoption, implementation, and maintenance*. *Reach* refers to the percentage of potential participants who are exposed to an intervention and how representative they are of others who might benefit from the program. *Efficacy*, or *effectiveness*, concerns both the intended effects of a program and the possible unintended outcomes. *Adoption* refers to the participation rate of eligible subjects and how well the setting and the people who deliver the intervention reflect future participants. *Implementation* denotes the extent to which various components of the program are delivered as intended. *Maintenance* is related to two questions: What are the program’s long-term effects? To what extent is the program continued after the completion of the evaluation? All five of these dimensions are considered equally important in the RE-AIM framework.

The Centers for Disease Control’s Framework for Planning and Implementing Practical Program Evaluation

Figure 1.2 illustrates the framework for planning and implementing “practical” program evaluation recommended by the U.S. Centers for Disease Control and Prevention (CDC).



This framework consists of six steps for accomplishing the evaluation (e.g., beginning with engaging stakeholders) and includes four standards for assessing the evaluation: accuracy, utility, feasibility, and propriety.

The three frameworks described above share several important features. First, they are all more precisely described as frameworks rather than models. That is, their purpose is to provide guidance in program planning and evaluation. Strictly speaking, models “predict” behavior or outcomes and are based on theoretical expectations or empirical evidence gained through experience and experiment. Frameworks leave the theories and methods of implementation to the evaluator.

These three frameworks are all-inclusive. PRECEDE-PROCEED, for example, contains a comprehensive set of factors that should be considered in program planning and evaluation. It is unlikely, however, that evaluators will ever find themselves involved in all aspects of program development and evaluation as characterized in this framework. In many cases, evaluators are called in to appraise the merits and effectiveness of existing programs. In other cases, evaluators are asked to be part of research teams that are developing interventions. No one really expects an individual evaluator to be *the* expert in the planning process or in the development of a program. The evaluator’s primary domain is collecting and interpreting valid data on the implementation of the program and its effectiveness.

Frameworks, such as PRECEDE-PROCEED, RE-AIM, and the CDC’s approach to practical evaluation may be useful in encouraging evaluators to pay attention to the origins and development of the programs they examine (even if the evaluator had little to do with establishing the need for the programs or with their implementation). Any knowledge an evaluator gains may help to design more realistic and relevant studies.

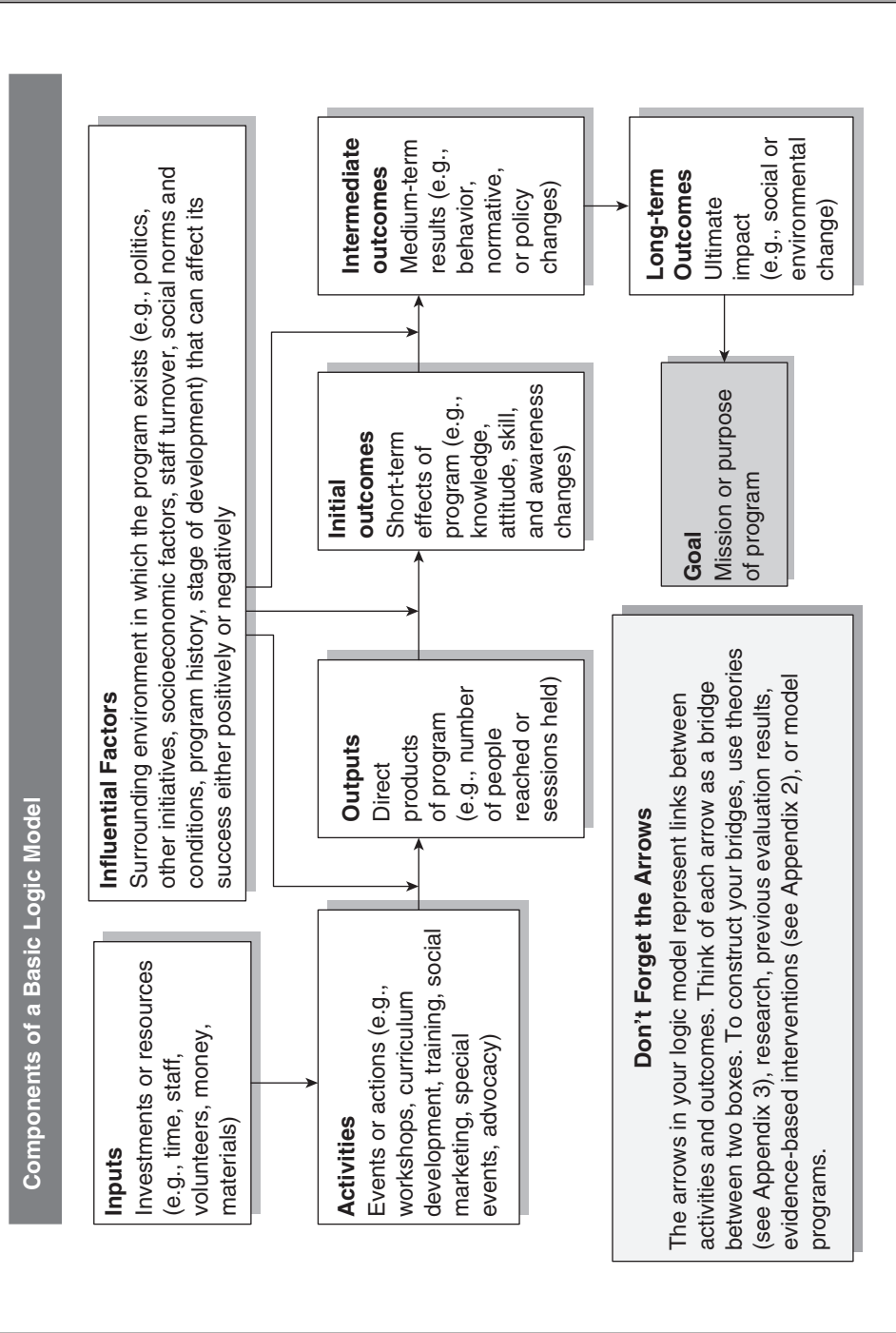
The CDC’s framework is different from PRECEDE-PROCEED and RE-AIM in that it incorporates standards for a good evaluation that specifically include propriety. *Propriety* refers to the legal and ethical considerations involved in evaluation research. With the exception of very small, local studies, and some larger studies conducted under certain circumstances, most evaluation studies are now required by law and institutional practice to demonstrate their ethical nature in writing, specifying how they will show respect for their participants and protect participants’ privacy.

Logic Models

A logic model is a planning tool to clarify and graphically display what your evaluation intends to do and what it hopes to accomplish. The most basic model consists of a depiction (often in graphic form) and an explanation of the resources that go into a program, the activities it undertakes, and the changes or benefits that result. The relationships are logical. In most cases, the relationships have not been tested empirically. Figure 1.3 shows the components of a basic logic model developed by the Centers for Disease Control and Prevention.

According to its supporters, a logic model describes the sequence of events presumed to produce or generate benefits or change over time. It portrays the chain

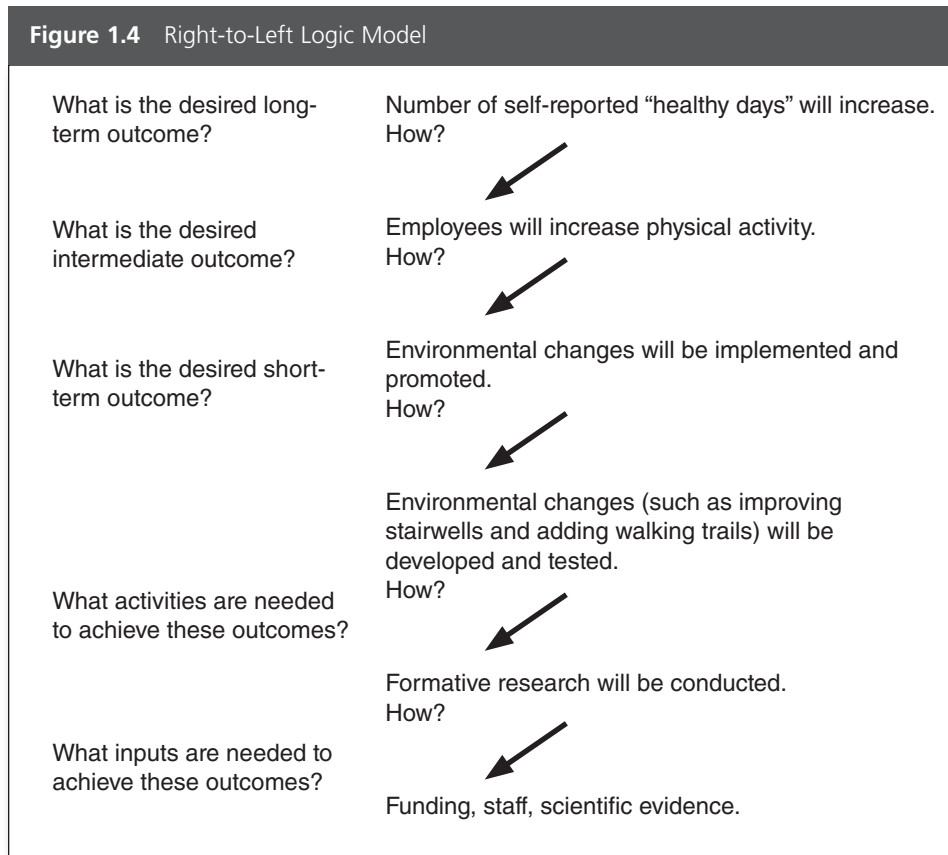
Figure 1.3 A Basic Logic Model



of reasoning that links investments to results. Additionally a logic model is termed a systems model because it shows the connection of interdependent parts that together make up the whole.

There is no single correct way to create a logic model. The stage of development of the program (i.e., planning, implementation, or maintenance) leads to one of two approaches used to create the model: right-to-left or left-to-right (Figure 1.4).

Right-to-Left Logic Model. This approach, also called *reverse logic*, starts with desired outcomes and requires working backwards to develop activities and inputs. Usually applied in the planning stage, this approach ensures that program activities logically lead to the specified outcomes if the arrow bridges are well-founded. As you progress from left to right in the logic model, ask the question: “How?” This approach is also helpful for a program in the implementation stage that still has some flexibility in its program activities.



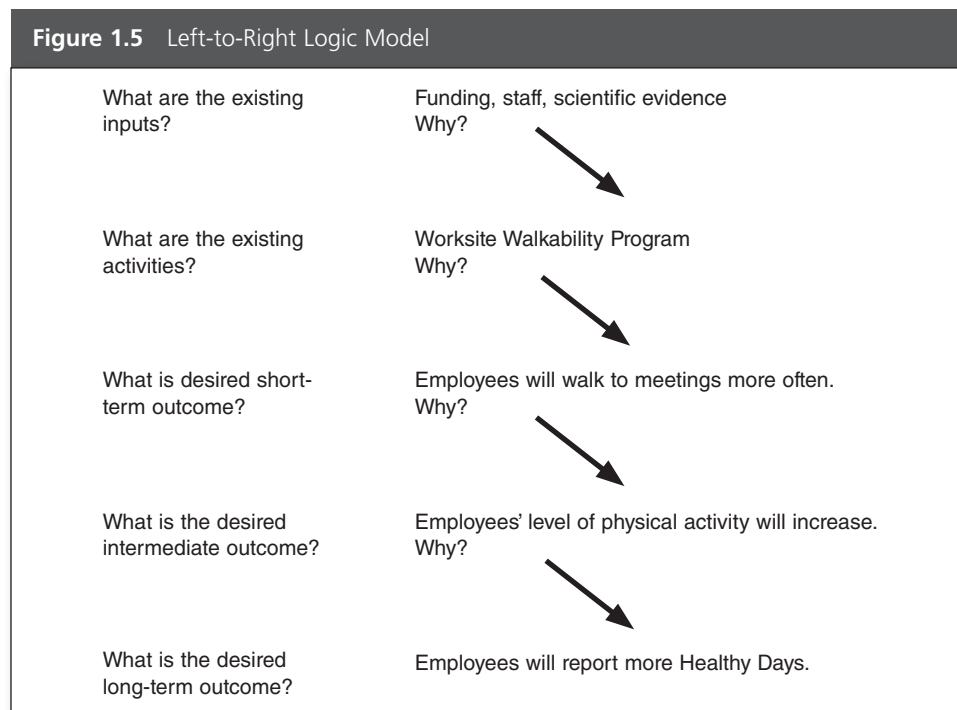
Left-to-Right Logic Model. This approach (Figure 1.5) also called *forward logic*, may be used to evaluate a program in the implementation or maintenance stage that does not already have a logic model. You start by articulating the program inputs and activities. To move to the right in your model, you continually ask the question, “Why?” You can also think of this approach as an “If-then” progression.

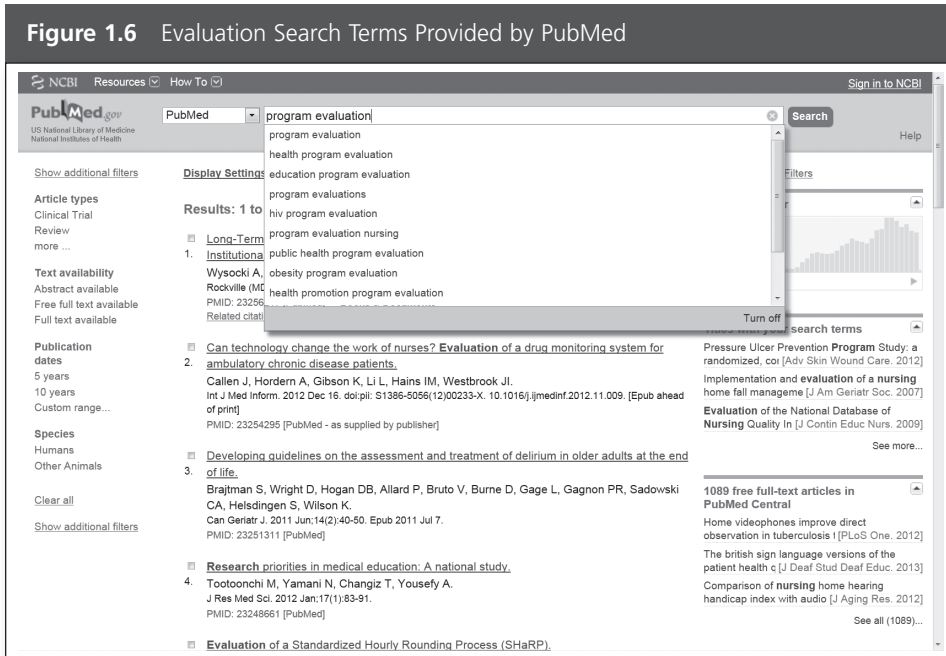
Evaluation Reports Online

Published evaluation reports provide information on promising programs and effective research methods. You can find evaluation reports in their entirety online. One good place to start is PubMed, the free access bibliographic database of the National Library of Medicine (<http://www.ncbi.nlm.nih.gov/pubmed>). If you search for program evaluation, you will be given detailed options as shown in Figure 1.6.

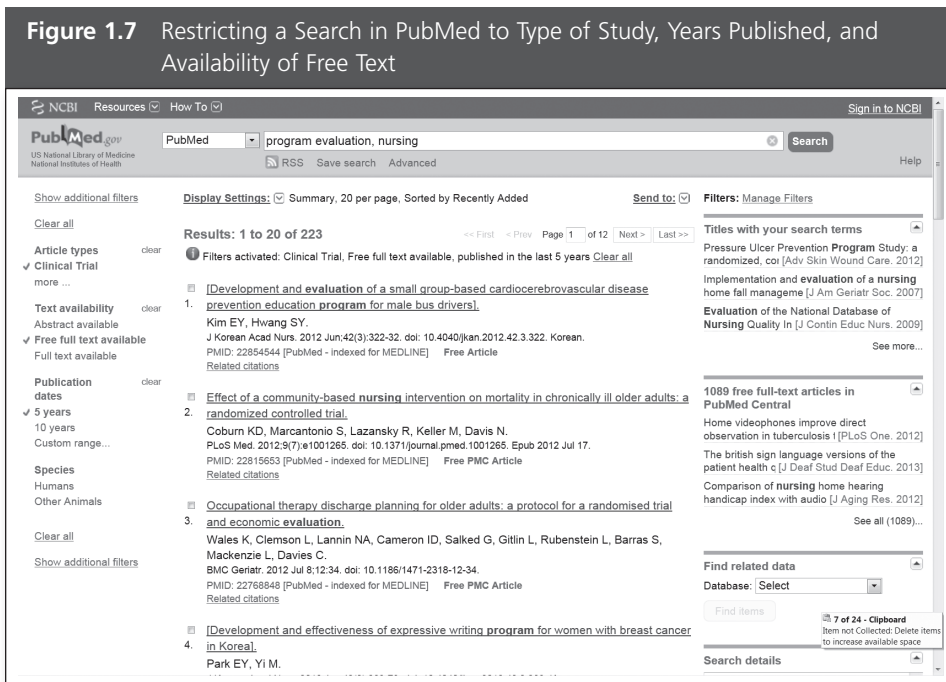
Suppose you are interested in reviewing program evaluations in nursing. You enter the words: “program evaluation, nursing.” This yields over 27,000 articles pertaining to program evaluation in nursing. Most reviewers would find this number overwhelming.

PubMed allows you to filter the search or narrow it down so that it produces fewer studies, but studies which are more likely to be on target. For example, you can ask PubMed to provide only evaluations that are clinical trials, published in the past five years and for which a full text is available (Figure 1.7). This produces 223 evaluations. You can





Source: Pubmed at <http://www.ncbi.nlm.nih.gov/pubmed>



Source: Pubmed at <http://www.ncbi.nlm.nih.gov/pubmed>

reduce the number of citations even more by adding other filters (publication date within one year rather than five) or other terms like “community-based.” If you simply add “community-based,” you are left with 22 evaluations to consider (Figure 1.8). The challenge is to weigh the need for a comprehensive search with the available resources (time, skill).

Another useful database for evaluation reports is ERIC (Education Resources Information Center): Go to ERIC and enter the words, “program evaluation.” If you limit your search to the last three years (as of December 2012), you will get 9,127 potentially usable evaluations. Similar to PubMed, ERIC (an online library of education research and information, sponsored by the Institute of Education Sciences (IES) of the U.S. Department of Education) gives you the opportunity to filter your search so that you do not have to review over 9,000 articles, many of which will not be relevant. For instance, if you restrict the search to evaluations that are primarily of interest to teachers, you will find 147 articles; for researchers, there are 38.

Figure 1.9 shows an ERIC search for journal articles published in the last six months (December 2012) on the effectiveness of elementary education programs. The search uncovers 10 articles.

Other databases containing program evaluations in the social, psychological, health, and educational fields include the Web of Science and PsycINFO. Almost all databases have tutorials to help you use a search strategy that will retrieve the articles and evaluations you need. Learning to be efficient in conducting online literature reviews is becoming an increasingly important evaluation skill (Fink, 2013).

Figure 1.8 Filtering Evaluation Reports

The screenshot shows the PubMed search interface. The search query is "program evaluation nursing community based". The results are filtered to show 15 articles. The first article is "Effect of a community-based nursing intervention on mortality in chronically ill older adults: a randomized controlled trial" by Coburn KD, Marcantonio S, Lazansky R, Keller M, Davis N. The second article is "Healthy eating and active living for diabetes in primary care networks (HEALD-PCN): rationale, design, and evaluation of a pragmatic controlled trial for adults with type 2 diabetes" by Johnson ST, Mundt C, Soprovich A, Wozniak L, Plotnikoff RC, Johnson JA. The third article is "Preventing mental health problems in children: the Families in Mind population-based cluster randomised controlled trial" by Hiscock H, Bayer JK, Lycett K, Ukoumunne OC, Shaw D, Gold L, Gerner B, Loughman A, Wake M. The search filters include "Clinical Trial", "Free full text available", and "published in the last 5 years".

Source: Pubmed at <http://www.ncbi.nlm.nih.gov/pubmed>

Figure 1.9 An ERIC Search for Recent Articles Evaluating Elementary Education Programs

The screenshot shows the ERIC search interface. On the left, there are filters for 'Narrow Your Search' including Author, Thesaurus Descriptor, Dates, and Source. The main search criteria are: ((Keywords: "program evaluation")), and Thesaurus Descriptor: "Program Effectiveness" [x], and Publication Type: "Journal Articles" [x], and Dates: In the last 6 months [x], and Education Level: "Elementary Education" [x]. The search results show 10 results per page, with the first result being 'Necessary, but Not Sufficient: The McKinney-Vento Act and Academic Achievement in North Carolina (EJ984614)'. The article details include authors Hendricks, George and Barkley, William, published in 2012-07-00, in the journal Children & Schools, v34 n3 p179-185 Jul 2012. The abstract discusses the McKinney-Vento Homeless Education Assistance Improvements Act of 2001 (MCKV) and its impact on local educational agencies (LEAs).

Source: Eric.ed.gov sponsored by the Institute of Education Sciences (IES) of the U.S. Department of Education.

Summary and Transition to the Next Chapter on Evaluation Questions and Evidence of Merit

Program evaluation is an *unbiased* exploration of a program's *merits*, including its *effectiveness*, *quality*, and *value*. An *effective* program provides substantial benefits to individuals, communities, and societies, and these benefits are greater than their human and financial costs. A *high-quality* program meets its users' needs and is based on sound theory and the best available research evidence. A program's *value* is measured by its worth to individuals, the community, and society.

To conduct evaluations, researchers pose questions and decide on evidence of effectiveness, quality, and value; choose study designs and sampling methods; and collect, analyze, interpret, and report information. The information produced by program evaluations is used by the financial supporters of the programs as well as by consumers (patients, students, teachers, and health care professionals), program developers, policy makers, and other evaluators and health researchers.

Collecting and analyzing interim or formative evaluation data is expensive and may produce misleading results; evaluators who choose to collect interim data should proceed

with caution. However, formative evaluation data are helpful in determining feasibility. Process evaluations are useful because they provide data on what and when something was done, which is helpful in understanding the program's dynamics.

Evaluations may use qualitative or statistical data or both in the same study. Participatory evaluations involve the community in all phases of the evaluation. Comparative effectiveness evaluations compare two existing programs in naturalistic settings and provide information designed to help consumers make informed choices.

Before evaluators can decide on the evaluation's design and data collection methods, they must choose evaluation questions and decide on evidence of program effectiveness. The next chapter discusses how to select and state evaluation questions and choose appropriate, justifiable evidence.

Exercises

Exercise 1

Directions

Read the two evaluations below and using only the information offered, answer these questions:

1. What are the evaluation questions?
2. What is the evidence of merit?
3. What data collection measures are being used?

1. Effectiveness of Home Visitation by public-health nurses in prevention of the recurrence of child physical abuse and neglect (MacMillan et al., 2005)

Objective: Recurrence of child maltreatment is a major problem, yet little is known about approaches to reduce this risk in families referred to child protection agencies. Since home visitation by nurses for disadvantaged first-time mothers has proved effective in prevention of child abuse and neglect, the aim is to investigate whether this approach might reduce recurrence.

Programs: 163 families with a history of one child being exposed to physical abuse or neglect are randomly assigned so as to compare standard treatment with a program of home visitation by nurses in addition to standard treatment. The main outcome was recurrence of child physical abuse and neglect based on a standardized review of child protection records.

Results: At 3-year follow-up, recurrence of child physical abuse and neglect in the intervention group did not differ between groups. However, hospital records showed significantly higher recurrence of either physical abuse and/or neglect.

2. Evaluating a Mental Health Intervention for Schoolchildren Exposed to Violence: A Randomized Controlled Trial (Stein et al., 2003)

Objective: To evaluate the effectiveness of a collaboratively designed school-based intervention for reducing children's symptoms of posttraumatic stress disorder (PTSD) and depression that resulted from exposure to violence.

Program: Students were randomly assigned to a 10-session standardized cognitive-behavioral therapy (the Cognitive-Behavioral Intervention for Trauma in Schools) early intervention group or to a wait-list delayed intervention comparison group conducted by trained school mental health clinicians.

Results. Compared with the wait-list delayed intervention group (no intervention), after three months of intervention, students who were randomly assigned to the early intervention group had significantly lower scores on the Child PTSD Symptom Scale, the Child Depression Inventory, Pediatric Symptom Checklist, and the Teacher-Child Rating Scale. At six months, after both groups had received the intervention, the differences between the two groups were not significantly different for symptoms of PTSD and depression.

EXERCISE 2

Directions

Define program evaluation.

EXERCISE 3

Directions

Explain whether each of these is an evaluation study or not.

1. The purpose of the study was to evaluate a randomized culturally tailored intervention to prevent high HIV risk sexual behaviors for Latina women residing in urban areas.
2. The researchers aimed to determine the effectiveness of an intervention regarding the use of spit tobacco (ST) designed to promote ST cessation and discourage ST initiation among male high school baseball athletes.
3. To study drivers' exposure to distractions, unobtrusive video camera units were installed in the vehicles of 70 volunteer drivers over 1-week time periods.

References and Suggested Readings

- Fink, A. (2013). *Conducting research literature reviews: From the Internet to paper*. Thousand Oaks, CA: Sage.
- Galvagno, S. M., Jr, Haut, E. R., Zafar, S. N., Millin, M. G., Efron, D. T., Koenig, G. J., Jr., . . . Haider, A. H. (2012, April 18). Association between helicopter vs ground emergency medical services and survival for adults with major trauma. *JAMA: The Journal of the American Medical Association*, 307(15), 1602–1610.
- Hammond, G. C., Croudace, T. J., Radhakrishnan, M., Lafortune, L., Watson, A., McMillan-Shields, F., & Jones, P. B. (2012, September). Comparative effectiveness of cognitive therapies delivered face-to-face or over the telephone: An observational study using propensity methods. *PLoS One*, 7(9).
- Hardy, L., King, L., Kelly, B., Farrell, L., & Howlett, S. (2010). Munch and move: Evaluation of a preschool healthy eating and movement skill program. *International Journal of Behavioral Nutrition and Physical Activity*, 7(1), 80.
- MacMillan, H. L., Thomas, B. H., Jamieson, E., Walsh, C. A., Boyle, M. H., Shannon, H. S., & Gafni, A. (2005). Effectiveness of home visitation by public-health nurses in prevention of the recurrence of child physical abuse and neglect: A randomised controlled trial. *The Lancet*, 365(9473), 1786–1793.
- Marczinski, C. A., & Stamates, A. L. (2012). Artificial sweeteners versus regular mixers increase breath alcohol concentrations in male and female social drinkers. *Alcoholism: Clinical and Experimental Research*, 37(4), 696–702.
- Porter, M. E. (2010). What is value in health care? *New England Journal of Medicine*, 363(26), 2477–2481.
- Stein, B. D., Jaycox, L. H., Kataoka, S. H., Wong, M., Tu, W., Elliott, M. N., & Fink, A. (2003, August 6). A mental health intervention for schoolchildren exposed to violence: A randomized controlled trial. *Jama: The Journal of the American Medical Association*, 290(5), 603–611.
- Volpp, K. G., Loewenstein, G., & Asch, D. A. (2012). Assessing value in health care programs. *JAMA: The Journal of the American Medical Association*, 307(20), 2153–2154.
- Yu, S. (2012, October 1). College students' justification for digital piracy: A mixed methods study. *Journal of Mixed Methods Research*, 6(4), 364–378.

Suggested Websites

PubMed: <http://www.ncbi.nlm.nih.gov/pubmed>

ERIC: <http://www.eric.ed.gov>

PsychINFO

Web of Science

The PRECEDE-PROCEED model: Community Tool Box: http://ctb.ku.edu/en/table_contents/sub_section_main_1008.aspx

The RE-AIM framework, go to <http://www.re-aim.org> and <http://www.cdc.gov/aging/caregiving/assuring.htm>

CDC's Practical Framework: <http://www.cdc.gov/eval/materials/frameworkoverview.PDF>

Logic Models: (http://www.cdc.gov/nccdphp/dnpao/hwi/programdesign/logic_model.htm)

http://www.childtrends.org/files/child_trends-2008_02_19_eva18programquality.pdf

Purpose of This Chapter

A program evaluation is an unbiased exploration of a program's effectiveness, quality, and value. Evaluations answer questions like: Did the program benefit all participants? Did the benefits endure? Is the program sustainable? Did the program meet the needs of the community, and was it done more efficiently than current practice? The answers to these questions require not just a "yes" or a "no," but evidence for the answers. This chapter begins with a discussion of commonly asked evaluation questions and hypotheses. It continues with an explanation on how to select and justify evidence of program merit.

Evaluation questions focus on programs, participants, outcomes, impact, and costs and provide data on whether a program achieves its objectives, in what context, with whom, and at what cost. Evidence of merit may be based on statistical significance, but the evidence should also have practical or clinical meaning. Sources of practical or clinical significance include experts and consumers, large databases ("big data") and the research literature. Experts can provide information on what to expect from a program, and consumers can tell the evaluator what is acceptable. Large databases provide information on populations and programs that can assist in program development ("What have others done?") and guide the evidence-selection process ("What did others achieve?").

The relationships among evaluation questions and hypotheses, evidence of merit, and independent and dependent variables are also examined in this chapter. Their connection is illustrated through the use of a special reporting form: The QEV or Questions, Evidence, Variables Report.
