



EVALUATION MATTERS

One of the great mistakes is to judge policies and programs by their intentions rather than their results.

—Milton Friedman

Upon completion of this chapter, you should be able to

- Define evaluation.
- Identify programs and policies that might be evaluated.
- Describe the purpose of evaluation and its relationship to research.
- Distinguish between formative and summative evaluation.
- Compare and contrast internal and external evaluation.
- Discuss the embedded evaluation model.
- Explain the first step in embedded evaluation.

1.1 WHAT IS EVALUATION?

Welcome to the field of evaluation! Whether you are new to the field and this is your first course in evaluation or you are a seasoned evaluator looking to explore new approaches, we are thrilled to walk with you on this journey and hope that you find the material in this text helpful to you. So, a good place to start . . . what is evaluation? Merriam-Webster online defines **evaluation** as the “determination of the value, nature, character, or quality of something.” We all do that on a daily basis:

- We estimate if a product is worth buying (are the upgraded features on this new iPhone version worth the additional \$100?);
- We judge whether spending extra time on a homework assignment is worth a higher grade and if a lower grade will impact our overall course average;
- We rate our professors (yes, professors really do look at these ratings from time to time);
- We appraise the work ethic and quality of one of our coworkers or fellow students;
- We assess the extent to which we will use a textbook in the future and determine whether it is more advantageous for us to rent the book for the semester or purchase a copy; and
- We make decisions about whether we can afford to rent an apartment on our own or if we have to get a roommate.

If you have contemplated any of the above, you are already an evaluator of sorts. While we will not spend class time debating the merits of the new iPhone, we will provide you with strategies to systematically make evaluative decisions.

I am sure you have noticed the “valu” embedded in the word evaluation. Like many words, “value” has multiple dimensions. The Etymology Dictionary online asserts that the term “value” is derived from the Latin word *valere*, meaning to be well, strong, and of worth. Well and strong relate to merit and have to do with inherent value, while worth is typically interpreted within a certain context. Thus, what is being evaluated, the evaluand (Scriven, 1979), may have an inherent value that is free of any context. On the other hand, the evaluand may only be of value to a particular group or in a specific context. In their analysis of the two aspects of value—merit and worth—Lincoln and Guba (1980) use the example of gold. They explain that gold can be judged on its merit or its worth. Judged on its merit, gold has inherent beauty. Judged on its worth, gold has a variable value according to the gold trading markets. Likewise, an SAT or GRE prep course may be judged on its merits based on its coverage of material and clarity of instruction. However, judgments based on worth to you will likely relate to how well you performed on the SAT or GRE. While Lincoln and Guba recognize that both the merit and worth of an evaluand can change over time, they emphasize the importance of deliberately considering context, and perhaps even multiple contexts, when making evaluative judgments. So, whether you are evaluating a new purchase or a course you are taking, consider both its intrinsic value and its value to you at this time in your life.

As described above, evaluation is a method of determining value and the evaluand is the subject of the evaluation. While we all evaluate as part of our human nature, this

Evaluation: a method used to determine the value of something.

textbook focuses on **program evaluation**. That is, the evaluand is the program and the focus is on determining the merit or worth of that program. For the purposes of this textbook, a **program** will be defined broadly to include

- A group of activities;
- Small, focused interventions;
- Organization-wide projects;
- Statewide initiatives;
- National reforms and policies; and
- International programs.

The tools explored in this text will be applicable to evaluating a set of activities or small interventions, as well as larger initiatives and multifaceted policies. Examples of programs include focused interventions such as the Olweus Bullying Prevention Program (Olweus & Limber, 2007) as well as national reforms such as Head Start (2019). See the “In the Real World” box for several additional examples of programs.

Program evaluation: evaluation used to determine the merit or worth of a program.

Program: defined broadly in this text to include a group of activities ranging from a small intervention to a national or international policy.

IN THE REAL WORLD . . .

First Year Experience (FYE) Courses are required courses for freshman at over half of four-year institutions. FYE courses intend to aid students both academically and socially. The What Works Clearinghouse (2006b) examined several quasi-experimental evaluations of FYE programs and found potentially positive effects for credit accumulation, degree attainment, and general academic achievement.

The Too Good for Drugs (TGF) program is designed to help students develop the skills to resist negative peer influences and the use of drugs, alcohol, and tobacco. The What Works Clearinghouse (2006a) examined two evaluations of TGF programs using randomized controlled

designs and found a potentially positive impact on behavior. No discernable effects were found for values, or knowledge of and attitudes towards drugs and alcohol.

Financial Incentives for Teen Parents to Stay in School is a strategy used by state welfare programs to promote school attendance and graduation. The belief is that by supporting graduation from high school, the teens will be less likely to depend on welfare in the future. The What Works Clearinghouse (2016) examined two evaluations of such programs. While they found potentially positive effects for teens staying in school, they found no discernable effects for students progressing to the next grade level or graduating from high school.

Source: What Works Clearinghouse (<http://ies.ed.gov/ncee/wwc>).

1.1.1 What Is the Purpose of Evaluation?

Evaluation intends to determine merit and worth. Yet, as discussed in the previous section, evaluation is also contextual. Determining the merit and worth of a program and focusing on its value for a group of people, under a certain set of circumstances, and in a specific context, is no easy task. This is especially true when the consequences of your evaluation might have human and financial implications. Thus, the way we go about evaluating a program is critical to making evaluative determinations. We trust that medical schools effectively evaluate their students to ensure that the cardiothoracic surgeon operating on a loved one has the skill to do so. We often have no choice but to trust that the auto mechanic evaluating why our car broke down is competent at repairing engines and ethical in quoting prices. As a medical school has procedures to evaluate its students and mechanics have protocols to assess car problems, program evaluators have methods, processes, standards, and tools to guide their evaluation.

The evaluative methods I have adopted over the course of my career have been heavily influenced by two factors. The first is my undergraduate training as an engineer. As an engineer, I learned how to think systematically. Engineering is a process and a way of thinking, as well as a discipline. The approach I learned in this context is one that starts with problem identification and description and ends with a set of solutions or recommendations—after which, the process starts again. The second influence on my thinking regarding evaluation is a well-known evaluator named Carol Weiss. Unfortunately, I never met her, but her seminal book titled *Evaluation* (Weiss, 1998) has been a trustworthy companion for decades. Of all who have attempted to define evaluation, Weiss's definition strikes me as the most comprehensive (pp. 4–5):

Evaluation is a systematic assessment of the operations and/or the outcomes of a program or policy, compared to a set of explicit or implicit standards, as a means of contributing to the improvement of program or policy.

The primary tenets of Weiss's definition are as follows:

- Evaluation is systematic.
- Evaluation focuses on operations.
- Evaluation focuses on outcomes.
- Evaluation evidence is compared to a standard.
- Evaluation is about improving programs and policies.

Systematic: logical and organized; undertaken according to a plan.

Evaluation is a **systematic** examination of a program. It uses the scientific method. Remember that from grade school?! The scientific method has been around since the

1600s, at least, and involves asking a question, researching the question, making and testing a hypothesis, analyzing data, and documenting results. It is what engineers, psychologists, biologists, and social scientists use in their work. It is the basic science that underlies evaluation. Evaluation is a formal, logical, and organized endeavor undertaken according to a plan.

Evaluation examines the **operations** of a program. The operations of a program include both what is implemented as part of the program and how it is implemented; operations are the processes involved with implementing the activities of a program. Operations are important for two main reasons—interpretation and improvement. Understanding the state of operations allows us to document how a program is operating. Understanding how a program operates, in turn, allows us to determine whether the operations are in fact in accordance with what was intended. This is important for interpreting any results. For instance, if depressive symptoms decrease after a new medication is dispensed, one might be led to believe the medication is effective in treating depression. However, examination of operations might show that over 50% of patients did not take their medication. Without examining how a program operates in actuality, versus how it might have been planned, one can draw inaccurate conclusions. The second reason to examine operations is similar to the first, but involves using the information gathered when looking at operations to make improvements to a program while in operation. So, if it is noticed two months in that patients are not taking their medication, new interventions could be put in place to improve compliance.

Evaluation also examines the **outcomes** of a program. Outcomes are the results that occur during and after implementation of a program. Examining the outcomes of a program allows you to make determinations about the effectiveness of a program. If, for instance, we are examining the instructional program at your college or university, the operations might be the quality of teaching and the rigor of assignments, but the outcome would likely be student learning. Knowing the extent to which students are learning can help (or hurt) a college with recruitment, fund-raising campaigns with alumni, and partnerships with organizations that may be interested in hiring graduates. By measuring outcomes, we can make determinations about whether the program worked, to what extent, in what ways, and for whom.

Evaluation evidence is compared to a **standard**. A standard is a target or yardstick that informs us of the ideal state. Standards are what we use, implicitly or explicitly, to judge the merit or worth of a program. The standard directs us in making this judgment. In our examples at the start of the chapter, we mentioned purchasing a new cell phone. How much greater would it need to be than the former model for you to purchase it? Likewise, the conflict of how much time to put in on an assignment versus the value of that assignment was introduced. Do you have an implicit standard, one you may not have written down or expressed, that drives whether spending an hour on a 5-point extra credit assignment is worth it to you? People who operate programs and those who evaluate those

Operations:

processes involved with implementing the activities of a program.

Outcomes:

results that occur during and after implementing a program.

Standard: target that we use, implicitly or explicitly, to judge the merit or worth of a program.

programs wrestle with similar decisions. If a medication helps 50% of the people who take it, is that enough to continue dispensing the medication? Probably. What if it helps 25%? Or what if it helps 50%, but makes the symptoms of 25% of the people worse? There are no easy answers when it comes to standards, though we will explore this thinking further in Chapter 7 in the section on program indicators and targets.

Finally, evaluation is not performed in a vacuum and it is not simply an exercise in curiosity. Evaluation is focused on how well a program works, under what conditions, and for what people. Evaluation is intended to provide information aimed at **improving programs and policies**. Ideally, the information obtained through an evaluation will be used to create more effective and efficient programs and policies. And, I imagine for most using this textbook, the programs and policies you might examine are intended to help people. For evaluators, that is the end result. It is the reason we do what we do—to inform programs and policies that will ultimately improve the lives of people. A secondary reason we evaluate programs and policies is to contribute to the field through informing theory and practice.

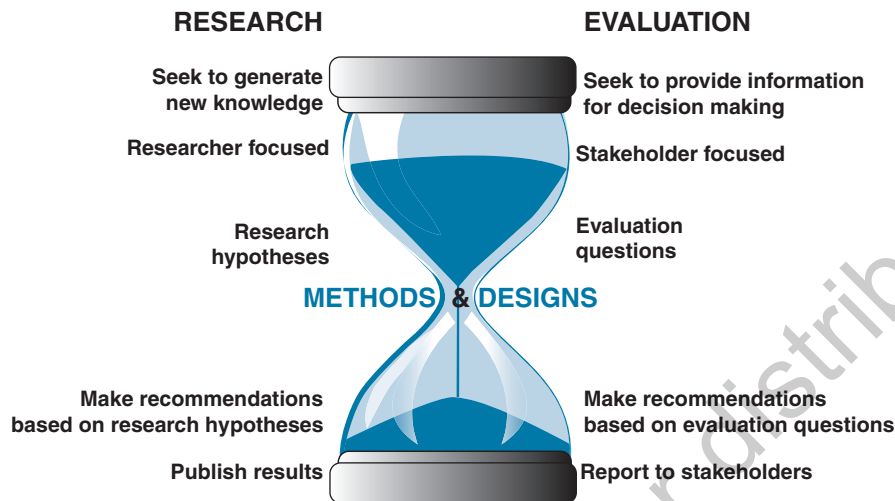
1.1.2 How Is Evaluation Different From Research?

Like evaluation, **research** is a systematic investigation in a field of study. In many ways, evaluation is a form of research. In fact, some refer to evaluation as evaluation research. Evaluation and research use the same methods and designs. The underlying science is the same. However, unlike evaluation, pure research is primarily focused on contributing to the greater body of knowledge in a certain area. This is in contrast to the primary purpose of program evaluation, which is to improve and make decisions about programs and policies.

Evaluation is more practice oriented than research. Evaluation findings are intended for use within a program or policy to effect change. In addition, while researchers often develop their own research hypotheses, evaluators typically work with program staff to develop questions to shape and focus the evaluation. In addition, as stated in the discussion of standards, evaluation intends to compare the evaluation results with what should be. That is, it is judgmental in nature and the eventual intention is to make a decision about whether a program should be continued, expanded, scaled down, or discontinued. Moreover, because evaluators are working in action settings where programs are being implemented in real time, we often face obstacles that might not be encountered in a lab or controlled research setting. For instance, in an evaluation relying on state test scores to examine the impact of curricular changes over a five-year period, policymakers discontinued use of the test and instead adopted an assessment that hindered comparisons to previous scores. Finally, evaluators depend upon people for data collection. As such, interpersonal skills, such as strong communication and listening skills, as well as flexibility and even a positive attitude, can be determinants of whether an evaluation is efficacious or unsuccessful.

Improving programs and policies: the purpose of evaluation; to create more effective and efficient programs and policies.

Research: a systematic investigation in a field of study; evaluation is a type of research.

FIGURE 1.1 Research and Evaluation

Source: Adapted from LaVelle (2010).

However, there are some elements that evaluation and research share. As stated above, program evaluation and research use the same methods and designs to frame and conduct their studies. Additionally, like researchers, evaluators have an obligation to disseminate their research. Sometimes this may be publication in peer-reviewed journals, as is common for researchers. However, for both researchers and evaluators, findings should also be shared with individuals or organizations that may benefit from understanding or adopting recommendations, as well as policymakers who are responsible for making policy that may be impacted and improved by the findings. Finally, both evaluators and researchers have ethical obligations and a code of conduct that guide how, why, from whom, and under what conditions data are collected. See Figure 1.1, adapted from a post by Lavelle (2010) on AEA365, a daily blog sponsored by the American Evaluation Association (AEA), for an illustration of some of the differences and intersections between evaluation and research.

1.2 WHY EVALUATE?

A pioneer in the field of research for effective marketing, Arthur C. Nielson based his work on the philosophy that “the price of light is less than the cost of darkness.” I know this is deep. Perhaps too deep for the hour in which you have to read this text. But it is definitely worth the time to think about the implications of his statement. If we are honest with ourselves, it is why we further our education. We go to school, we read, we study

because in the long run we believe it will make a difference in our quality of life that is worth the price we pay for our education.

In fact, it would be difficult to find an example where knowledge and truth do not matter. Yet so much *light* is ignored because of the immediate and short-term cost, resulting in a great long-term cost of managing the *darkness*. There is no place this is truer than in policy making. For instance, it is well documented that drug treatment is a more effective as well as cost-efficient solution than minimum mandatory sentences for and incarceration of drug offenders (McVay, Schiraldi, & Ziedenberg, 2004). Yet policymakers often reject the up-front costs of effective drug treatment programs, which result in a much heavier burden on society in the long run due to incarceration, recidivism, reduced productivity, and decreased safety. Similarly, the cost to treat mentally ill individuals is much less than the cost of incarceration after a crime has been committed. The deinstitutionalization of mental health treatment in the 1960s and 1970s, by the shuttering of state mental hospitals, resulted in a large increase of severely mentally ill individuals in the U.S. prison system (Collier, 2014). While state mental hospitals may not have been a humane solution, an alternative, community-based treatment for the mentally ill was not developed. Thus, prisons became the new asylum for the mentally ill. Treatment of the mentally ill in the community focusing on medication compliance, counseling, housing support, and job opportunities is a much less expensive alternative to unsafe communities and incarcerating mentally ill individuals. The National Alliance on Mental Illness (Giliberti, 2015) estimates an annual cost of \$31,000 to incarcerate an individual with mental illness, while community-based mental health care costs about \$10,000 annually.

The previous example is well documented in the literature. Another well-documented example is the Scared Straight program. See “In the Real World” for information on the program and related evaluations.

Unfortunately, not all policies and programs are as well researched as investments in community-based mental health treatment and the Scared Straight program, reinforcing the need to have data regarding policy and program effectiveness. In addition, even for programs that have been well researched, such as Scared Straight, policymakers and practitioners may still decide to use them. Both of these issues, a lack of informative research as well as an underuse of available research in decision making, are of relevance to evaluators and we hope by the end of the text, you will have the knowledge and tools to

- Design rigorous and informative evaluations;
- Collect evaluative information on programs and policies;
- Interpret evaluation data to inform policies and programs; and
- Effectively present and disseminate data to increase opportunities for use.

IN THE REAL WORLD . . .

Scared Straight was introduced in the 1970s as a program to prevent juvenile delinquency. Participants were youth at risk of becoming delinquent; the program introduced them to prisons and hardened criminals in order to deter them from continued criminal activity.

Multiple randomized trials in the United States showed the program did not work, and in fact was harmful to many youth (Aos, Phillips, Barnoski, & Lieb, 2001; Lilienfeld, 2005; Petrosino, Turpin-Petrosino, Hollis-Peel, & Lavenberg, 2012). Youth who went through the program had a higher rate of re-offending than similar youth who did not participate in the program.

Why did policymakers continue to use the program? Because it cost less than \$100 per child. It seemed like a low-risk program; if it didn't work, little money would be lost. Think again. Evaluations showed that in the long run, taxpayers and crime victims paid much more than the program costs because of the additional criminal activity of those who participated. In fact, a comprehensive cost-benefit study found that for every \$100 spent on Scared Straight, taxpayers and victims paid an additional \$10,000 in costs due to increased contact with the criminal justice system after participating in the program (Washington State Institute for Public Policy, 2007, 2018).

Source: Petrosino et al. (2012); Washington State Institute for Public Policy (2018).

1.2.1 Evaluation Is an Ethical Obligation

When presented with information regarding the costs of ineffective programs, it is not a leap to conclude that it is an ethical obligation of those who implement programs and policies to also have those programs and policies evaluated. Yet, program planning in general is often one of those areas where many opt to forego evaluation, due to the up-front costs, in favor of spending those funds on program services. While serving more people may seem noble, it is not at all noble if the program is ineffective at best, and harmful at worst. While perhaps an unpopular view, my view nonetheless is that claiming ignorance to a program or policy being ineffective or even harmful, due to a lack of available data to make a judgment or due to an unwillingness to listen to available data, is an unacceptable and unethical assertion. What is your viewpoint? There is no right or wrong answer, but certainly it is an interesting question to consider.

Some readers of this text may think it is understandable that program leaders want to maximize program funds used to deliver services. I tend to agree with you. It is a dilemma . . . spend money on program services (and serve more people) or spend money on evaluation (and serve fewer people). If you have donated money to a charity to provide a new after-school mentoring program for middle-school students, you may want the charity to put all donated funds into the mentoring. However, what if the mentoring is ineffective and a waste of your donated dollars? Would you be willing to let the charity allocate a

portion of the donation to evaluate the effectiveness of the mentoring program? Using policy and program resources to collect the necessary data to evaluate effectiveness is the only way, as Nielson might say, to live in the *light*. That is, using funds now to determine if the outcomes of a program warrant continued funding of the program in the future is a long-game mindset.

1.2.2 Evaluation Fosters Quality

It is the very nature of evaluation to increase knowledge and this knowledge can be used to improve programs. Thus, evaluation fosters quality. It provides the necessary information to improve a program continuously, allocate resources in ways that can maximize effectiveness, and refine program strategies for greater impact. A student's improvement can be facilitated with constructive teacher feedback. An employee's performance is supported when provided with ways to improve. An organization's productivity is enhanced when there is a culture of process improvement. And the quality of a program is fostered when program components are examined and sound evaluative information is made available. Thus, the premise holds for people, organizations, and programs: When good information is provided, better decisions can be made.

Have you ever heard anyone say, the more you know, the less you know? This is directly tied to one of my favorite statements: Ignorance is bliss. While it might be blissful to the ignorant, to those who have to deal with the consequences of ignorance, it can be aggravating, troubling, and costly. Yet, the more we learn, the more we understand all that we do not know. This journey of learning empowers us to make better, more informed decisions. And it also inspires us to search for greater understanding. Thus, evaluation produces knowledge that informs decisions, which in turn creates the need for more knowledge. This cycle of knowledge generation and use is a continuous improvement process that fosters informed decision making and, in turn, promotes quality in programs and policies.

1.2.3 Evaluation Is a Viable Career

If evaluation as an ethical obligation as well as fostering quality in programs and policies has not convinced you to learn all you can about evaluation, perhaps knowing that evaluation is a growing field with many job opportunities will spark your interest. A decade ago, Russ-Eft and Preskill (2009) included in their list of reasons to pursue a career in evaluation the increasing respect for evaluation experience as a skill that is highly marketable. Indeed, there is a need for trained evaluators with nonprofit organizations, corporations, and research centers. There are many opportunities for evaluators internationally. In this era of data-driven decision making, thankfully organizations are recognizing the value that data can provide to their operations. It is also a time of accountability, with many programs being required to show evidence of impact to receive continued funding.

1.3 VALUES AND STANDARDS IN EVALUATION

For evaluators, the creation of knowledge is based upon data. But how do we decide what data to collect? How do we decide what questions to ask? And once knowledge is generated, how are the data used? In evaluation, the people who use evaluation findings are called stakeholders. In fact, a **stakeholder** is anyone who has an interest in or is involved with the operation or success of a program. Key stakeholder groups often include program staff, program participants, community members, and policymakers. In what ways do different stakeholder groups use evaluation findings? How do stakeholders weigh evaluation data in making decisions? Evaluation is the activity of examining programs and collecting information, as well as the process of determining how that information will be used. Both aspects, how we collect data and how we use data, are influenced by factors related to evaluator skills and preferences, as well as stakeholder values and the context within which the program operates.

Thus, the *valuing* that is part of evaluation is influenced by context, including our own values, the values of stakeholders, as well as politics, resources, and even history. As evaluators, it is important that we are clear about the values and standards upon which our evaluative judgments are based.

An important tenet of Weiss's definition of evaluation involves the comparing of evaluation evidence to a standard, in order to make a judgment about a program or policy. Thus, the standard holds power. For instance, in your classes, you must achieve a certain grade to pass a course. That grade requirement is the standard. Likewise, states set cut scores for state achievement testing that are used to determine course placement and even graduation. The cut score is a standard that has the power to affect a student's future. How was that cut score set? The standard was likely set by a group of administrators based upon something. That *something* likely includes data, professional judgment, research, and experience—all of which are influenced by values.

Valuing is the process of estimating the importance of something. A **value** is a principle or quality that we use to estimate that importance. Value is also the estimate of importance. That is, we use values to assign value. A teacher might value effort over performance, and thus assign grades based largely on effort. A manager might value quantity of work over quality of work, and use standards based on these values for employee performance appraisals.

Earlier we mentioned that evaluation, or the process of valuing, has two components: merit and worth. The merit of an evaluation may be determined by the methods used and the rigor of the design. But what influences methods and design? An evaluator's own values often guide the choice of design and methods. What does the evaluator value? Hearing stories from participants detailing personal experiences with the program? Studying quantitative indicators of program impact? Involving stakeholders in all aspects of the evaluation? Being the expert and using that expertise to design and implement the evaluation? Ensuring all possible

Stakeholder: anyone who has an interest in or is involved with the operation or success of a program. Key stakeholder groups often include program staff, program participants, community members, and policymakers.

Value: a principle or quality used to estimate importance; an estimate of importance.

participants receive the program being evaluated? Using the most rigorous evaluation design, even if that means some potential participants do not receive the program or are delayed in receiving it? There are no right answers to these questions; all are debated among evaluators.

The worth of an evaluation is dependent upon context and who is making the judgment of worth. Worth to an evaluator may raise the same questions described earlier relating to merit. Worth to stakeholders is influenced by their own values. In order to design an evaluation that is useful to stakeholders, it is important for an evaluator to understand stakeholder values. These values will likely vary across stakeholder groups, and thus the design of the evaluation will have multiple components to address issues that allow for judgments of worth to be made.

If at this point you are ready to throw your hands up in frustration at the subjectivity inherent in valuing, instead marvel at the complexity of human thought. Okay—enough marveling. There are tools to guide evaluators in understanding not just our own values, but more important, those of our stakeholders. There are also guidelines and principles evaluators can use to conduct evaluations with objectivity.

1.3.1 Guiding Principles for Evaluators

The American Evaluation Association (AEA) provides guiding principles for evaluators. The **AEA Guiding Principles for Evaluators** (AEA, 2018) is a set of five principles that embody the values of the American Evaluation Association, an international professional association of evaluators. And yes, the guiding principles are based on values. However, these are values accepted by evaluators across disciplines and have been ratified by a membership of over 7,000 evaluators. So, they have merit in their interdisciplinary nature and worth in their widespread acceptance. The guiding principles are intended to promote the ethical behavior of evaluators, and address ideals that reach across disciplinary boundaries, such as an evaluator's obligation to be professionally and culturally competent. They include guidance in the following domains:

- systematic inquiry,
- competence,
- integrity,
- respect for people, and
- common good and equity.

Each of these five guiding principles for evaluators will be described more fully in Chapter 3. The full text of the *American Evaluation Association Guiding Principles for Evaluators* appears, with permission from the AEA, at the end of this chapter (see Figure 1.4).

AEA Guiding Principles for Evaluators: a set of five principles intended to guide the ethical behavior of evaluators; the guiding principles are systematic inquiry, competence, integrity, respect for people, and common good and equity.

1.3.2 Program Evaluation Standards

Another important resource for evaluators is a set of standards issued by the Joint Committee on Standards for Educational Evaluation. The Joint Committee is a group of representatives from multiple professional organizations, including the American Evaluation Association, that have an interest in improving evaluation quality. The **Joint Committee's Program Evaluation Standards** (Yarbrough, Shulha, Hopson, & Caruthers, 2011) is a set of 30 standards to guide evaluators in designing and implementing quality evaluations. The standards address five areas:

- utility,
- feasibility,
- propriety,
- accuracy, and
- evaluation accountability.

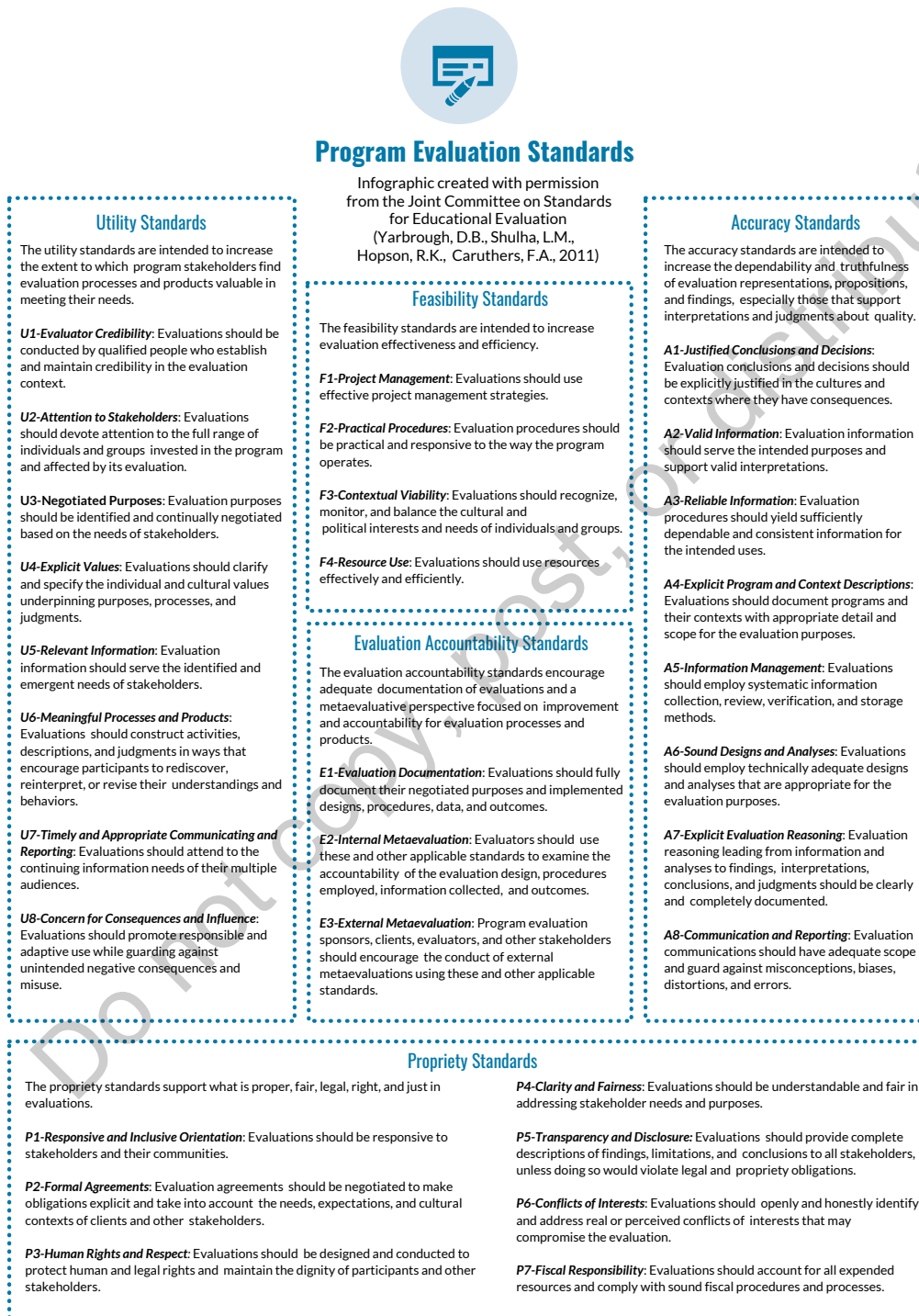
These standards provide practical guidance on how to conduct effective and equitable evaluations that produce accurate findings and promote usability. See Figure 1.2 for a list and description of the 30 program evaluation standards.

In addition to the AEA Guiding Principles and the Joint Committee's Program Evaluation Standards, experienced evaluators have also provided resources to guide evaluators in the appropriate consideration of values and standards in their evaluation work. Stufflebeam (2001) offers a checklist of values and criteria for evaluators to consider when designing and conducting evaluations. This checklist includes societal values, such as equity, effectiveness, and excellence. Also included are institutional values, such as the organization's mission, goals, and priorities. House and Howe (1999) provide a detailed look at values in evaluation in their book *Values in Evaluation and Social Research*. Should the topic of values in evaluation spark your interest, the House and Howe text is an excellent resource through which to continue your exploration.

In summary, our own values affect all aspects of evaluation, from the research design and methods we choose, to how we interact with stakeholders and the way we interpret our findings. Stakeholder values and the political context in which a program operates also affect how an evaluation is conducted and how data are used. However, knowledge is power. Understanding stakeholder values and the political context can aid you in designing an evaluation that meets stakeholder needs and is more likely to be used to influence decision making. Understanding our own values can help us to examine how they might impact our evaluations, as well as increase awareness of ways to improve our practice. Understanding and adhering to professional guidelines and standards can only serve to strengthen the

Joint Committee's Program Evaluation Standards: a set of 30 standards intended to guide evaluators in the areas of utility, feasibility, propriety, accuracy, and evaluation accountability.

FIGURE 1.2 Joint Committee's Program Evaluation Standards



**QUICK CHECK**

1. What is evaluation? How do we use evaluation in our everyday lives?
2. How are research and evaluation related? What are some ways in which research and evaluation differ?
3. Why might it be considered unethical to not evaluate a program or policy?

work that we do as evaluators. These same professional guidelines and standards can aid stakeholders and evaluators in assessing the merit and worth of evaluation findings.

1.4 TYPES OF EVALUATION

The terms evaluation, program, research, and embedded evaluation were explained in the previous section. There are several additional important evaluation terms that will be explained in this section. The terminology introduced in this section is often used in evaluation solicitations and requests for proposals. A **request for proposal (RFP)** is an announcement that an agency has funds available for specified work and an invitation for organizations to prepare a description of how they would complete that work. Some RFPs ask specifically for evaluation services and some may ask for program development or implementation, with a stipulation that an evaluation plan must be included in the proposal. RFPs will often use language indicating that formative and summative evaluation is required, or an external evaluator is preferred.

An RFP is only one method through which you might hear about the need for an evaluation. While some of the evaluations I have conducted originated with an RFP, most of my evaluation work comes when an individual or organization directly contacts our evaluation center. Sometimes we are asked to evaluate a program that is being planned or already in operation. Other times we are asked to write an evaluation plan for a project being proposed and submitted for funding, with the understanding that should the project be funded, we will conduct the evaluation. Evaluators might also be hired to be part of an organization; some larger organizations have evaluators on staff to conduct routine evaluations of their programs and policies. Regardless of whether an evaluation comes about due to an RFP, direct contact, in-house planning, or some other means, the terms presented in this section are commonly used when requesting evaluation assistance.

The framework presented in this section to introduce evaluation terminology is adapted from Trochim's (2001) *The Research Methods Knowledge Base*. He categorizes common types of evaluation within the formative and summative domains. **Formative evaluation** is evaluation aimed at providing information to program staff so they can improve the program while it is in operation; formative evaluation methods include

Request for proposal (RFP): a solicitation for organizations to submit a proposal indicating how they would complete a specified project.

Formative evaluation: evaluation aimed at providing information to improve a program while it is in operation.

process evaluation, implementation assessment, needs assessment, and evaluability assessment. **Summative evaluation** is evaluation aimed at providing information to program staff regarding effectiveness so they can make decisions about whether to continue or discontinue a program; summative evaluation methods include outcome evaluation, impact evaluation, cost-effectiveness/cost-benefit analysis, and meta-analysis. Many evaluations have both formative and summative components, with the formative component geared toward improving the impact measured by the summative evaluation.

1.4.1 Formative Evaluation

An important purpose of evaluation is to collect information that enables program staff to improve a program while it is in operation. Formative decisions are those that are intended to form, shape, and improve a program while being implemented. Thus, formative evaluation is performed to provide ongoing data to program staff for continuous improvement. Formative evaluation examines the implementation process, as well as outcomes measured throughout program implementation, to make decisions about mid-course adjustments, technical assistance, or professional development that may be needed, as well as to document the program's implementation so that others can learn from the program's operation.

Evaluators use **process evaluation** to make mid-course adjustments to shape a program. When evaluators conduct a process evaluation, they examine the output of the process of implementing a program's operations. A process evaluation might focus on the number of people trained, types of services delivered, methods of training used, and so forth.

Another form of formative evaluation is **implementation assessment**, that is, determining the degree to which a program is implemented as planned. Implementation assessment examines the fidelity with which a program's strategies or activities have been implemented. In order to assess fidelity of implementation, one must have a model of how a program would "look" if it was implemented as envisioned. Likewise, having a sense of how a program implementation is not optimal can help to establish the degree of fidelity. For instance, think about how your teachers might use a rubric to grade a written assignment. Let's suppose your rubric scores range from 0 to 10 on ten components of the paper, for example, identification of thesis, organization, and grammar. The rubric would tell you what it means to get a 10 on organization versus a 5 or 0. That way, you would know what the teacher sees as an "ideal" paper versus an average or below average paper. Similarly, implementation assessment can help evaluators understand how various activities within a program are implemented and the degree to which implementation matches the intentions of the program developers.

Prior to implementing a new program or restructuring an existing program, **needs assessment** can be used to shape a program by examining the needs of proposed participants, needs of stakeholders, and how to meet the needs of both. Needs assessment is a systematic examination of what program services are needed, who needs these services, and in what ways they need the services.

Summative evaluation:

evaluation aimed at providing information about effectiveness in order to make decisions about whether to continue or discontinue a program.

Process evaluation:

formative evaluation aimed at understanding the operations of a program.

Implementation assessment:

formative evaluation that examines the degree to which a program is implemented with fidelity (according to plan).

Needs assessment:

formative evaluation that focuses on what services are needed and who needs them.

Finally, **evaluability assessment** helps determine whether it is feasible to conduct an evaluation of a particular program (Trevisan, 2007; Wholey, 1979, 2002). It addresses whether a program or policy has clearly defined outcomes of interest; if it is feasible to attribute outcomes to the program or policy; whether data are available, reliable, and valid; whether stakeholders are identifiable and accessible; if the necessary resources are available to conduct the evaluation; and the likelihood that findings will be used appropriately. Evaluability assessment also examines how stakeholders might be used within the evaluation to shape the program and its attendant evaluation in a way that best meets the determined needs. An excellent resource on how to conduct evaluability assessments is *Evaluability Assessment: Improving Evaluation Quality and Use* by Trevisan and Walser (2014).

1.4.2 Summative Evaluation

A primary purpose of evaluation is to make summative decisions. Summative decisions are made by looking at all of the information. At the root of the word “summative” is *sum*. A “sum” is a total or a result. Thus, summative evaluation is performed to make final, outcome-related decisions about program funding. Summative decisions include whether to continue, expand, or discontinue a program based on evaluation findings.

Summative evaluation speaks to decisions about a program’s future. As such, **outcome evaluation** is summative evaluation focused on how well a program met its specified long-term goals. If a program proposes to improve learning, outcome evaluation would focus on changes in knowledge. If a program proposes to change practices related to healthy eating or medication compliance, outcome evaluation would focus on behavior change.

Impact evaluation also measures the outcomes of programs. However, **impact evaluation** is broader than outcome evaluation, as it measures all impacts of a program, both those intended as specified by a program’s goals and those unintended. For example, an impact evaluation of No Child Left Behind (NCLB) showed that principals and teachers made better use of test data after NCLB was passed and that scores on state tests had increased (Center on Education Policy, 2006). However, the study also showed that the curriculum had narrowed to focus on tested material, student creativity had declined, and flexibility within the law might account for more students being classified as proficient (Center on Education Policy, 2006). Other studies have shown that NCLB decreased the average quality of principals at disadvantaged schools due to principals seeking employment at schools less likely to experience NCLB sanctions (Li, 2010), reduced educational programming for gifted students (Beisser, 2008), and raised new challenges specific to using accommodations in high-stakes testing (Cawthon, 2007).

Many program funders request information on the efficiency of a program. That is, they want to know the value of a program, either in terms of dollars saved or benefits to participants or society. Hence, **cost-benefit/cost-effectiveness analysis** is summative evaluation that focuses on estimating the efficiency of a program in terms of dollar costs saved

Evaluability assessment: formative evaluation used to determine if an evaluation is feasible and the role stakeholders might take in shaping the evaluation design.

Outcome evaluation: summative evaluation aimed at measuring how well a program met its stated goals.

Impact evaluation: summative evaluation that measures both the intended and unintended outcomes of a program.

Cost-benefit/cost-effectiveness analysis: summative evaluation that focuses on estimating the efficiency of a program in terms of dollar costs saved (cost-benefit) or outcomes measured (cost-effectiveness).

(cost-benefit) or outcomes observed (cost-effectiveness). The amount saved by a program might differ depending on the time frame used for the analysis. For instance, recall the example about mental health treatment at the start of the chapter. Estimating the amount saved one year out would not give a full picture of the benefits of the program; the cost savings for some programs are not realized until years or decades after the program ends. The allure of funded preschool programs is not simply preparing a child for kindergarten, but rather, proponents argue, the long-term benefits of preschool programs include increased high school graduation rates, which in turn lead to increased employability and improved quality of life. Estimating the cost-benefit of a program intended to have long-term cost savings is difficult, but can be done (see “In the Real World” on the Scared Straight program). An alternative to measuring program success in terms of cost savings is calculating the benefits of a program in terms of nonmonetary outcomes. While cost-benefit is a ratio of the costs of the program to the costs saved by the program, cost-effectiveness is calculated by using a ratio of the total costs associated with program delivery to the impact of the program on a chosen outcome. For instance, a behavioral intervention program might measure cost-effectiveness as the change in behavioral outcomes for every \$1 spent on the program. A program targeting healthy eating might estimate the change in fast food consumption per dollar spent on the program or weight loss associated with each dollar invested in the program. For more information, see Cellini and Kee’s (2015) chapter on cost-effectiveness and cost-benefit analysis in the *Handbook of Practical Program Evaluation* (Newcomer, Hatry, & Wholey, 2015).

Meta-analysis is a form of summative evaluation that integrates the findings of multiple studies to estimate the overall effect of a type of program. Meta-analysis is a statistical approach that merges results across a body of research. Such analyses are also referred to as systematic reviews because the methodology is highly structured and involves defining inclusion and exclusion criteria for prospective studies, combining measures across studies, and calculating new estimates of effectiveness based on the pooled data. See the What Works Clearinghouse (<https://ies.ed.gov/ncee/wwc/>) for more information on systematic reviews of evidence.

Finally, a **meta-evaluation** is an evaluation of an evaluation (Scriven, 1969, 2009; Stufflebeam, 1978). Formative meta-evaluations provide feedback to improve an evaluation. Summative meta-evaluations assess the quality and merits of an evaluation. The Joint Committee’s Program Evaluation Standards, discussed earlier, can be used to conduct meta-evaluations.

This text will focus primarily on the process evaluation and implementation assessment components of formative evaluation and the outcome and impact evaluation components of summative evaluation. However, resources will be provided in subsequent chapters for additional information on needs assessment, evaluability assessment, cost-benefit/cost-effectiveness analysis, meta-analysis, and meta-evaluation.

Meta-analysis: summative evaluation that integrates the findings of multiple studies to estimate the overall effect of a type of program.

Meta-evaluation: an evaluation of an evaluation.

1.5 INTERNAL AND EXTERNAL EVALUATION

Because internal and external evaluation are terms commonly used by those within and outside of the evaluation field, I include them under types of evaluation. However, it should be noted that they are not types of evaluation like formative and summative evaluation, but rather a way of describing the relationship of the evaluator to the program itself.

An evaluation can be conducted by someone inside the organization within which a program operates or by someone outside of the organization. However, the optimal arrangement is often a partnership between the two, that is, forming an evaluation team that includes both internal and external evaluators.

An **internal evaluator** may be someone at the organization who is knowledgeable about the program. For evaluations that focus on program improvement and effectiveness, having an internal evaluator on the evaluation team can foster a deeper understanding of the context in which the program operates. Involving people inside the organization also helps to build capacity within the organization to conduct evaluation. However, an internal evaluator should be someone who is in a position to be objective regarding program strengths and weaknesses. For this reason, choosing an internal evaluator who is *responsible* for the program's success is not recommended and may compromise the evaluation. Likewise, any time an internal evaluator is very close to the program being evaluated, objectivity or perceived objectivity may suffer. In order to maintain objectivity, an internal evaluator should be outside of the program. However, while staff from within the program should not be part of the core evaluation team, they should certainly partner with the evaluation team to ensure that the evaluation informs the program during every phase of implementation.

An **external evaluator** is an evaluator who is employed from outside of the organization that operates the program or policy to be evaluated. It is good practice to have an external evaluator be part of your evaluation team. Using an external evaluator as a “critical friend” provides you with an extra set of eyes and a fresh perspective from which to review your design and results. Professional evaluators are trained in the design of evaluations to improve usability of the findings, and they are skilled in data collection techniques such as designing surveys, facilitating focus groups, conducting interviews, choosing quality assessments, and performing observations. An experienced evaluator can also help you analyze and interpret your data, as well as guide you in the use of your results.

Partnering with an external evaluator can improve the **credibility** of the findings, as some may question whether an evaluator from within an organization can have the **objectivity** to recognize areas for improvement and to report results that might be unfavorable to the program. This is not to imply that credibility or objectivity problems are usual or even common with internal evaluations. External as well as internal evaluations can suffer from a lack of credibility or objectivity. But issues of credibility and objectivity in internal

Internal evaluator: an evaluator employed by the organization that operates a program (but preferably not responsible for the program itself).

External evaluator: an evaluator who is employed outside of the organization in which the program operates.

Credibility: in evaluation, the degree of trust someone has that findings are reported accurately and should be believed.

Objectivity: in evaluation, the degree to which an evaluator can put aside any bias and impartially interpret and report findings.

evaluations come up because of the perceived threat to the findings. For that reason, it is important for evaluators to disclose, in a straightforward manner, any conflicts of interest or connections to the program under evaluation when reporting evaluation findings.

The choice of who conducts your evaluation should depend upon the anticipated use of the results and the intended audience, as well as your available resources. If evaluation results are to be used with current or potential funding agencies to foster support and assistance, contracting with an external evaluator would be your most prudent choice. If the evaluation is primarily intended for use by your organization to improve programs and understand impact, an evaluation team composed of an internal and an external evaluator may be preferable. Connecting with someone outside your organization to assist with the evaluation and results interpretation will likely enhance the usability of your evaluation and the credibility of your evaluation findings. Evaluation as a partnership between an internal evaluator and an external evaluator is the ideal arrangement to ensure the utility of the evaluation and its results.

An external evaluator may be a researcher or professor from your local university, a professional evaluator from a private evaluation firm, or an independent evaluation consultant. For programs where an external evaluator might be preferred, funding an outside evaluator may not be feasible. In such cases, partnering with an evaluator within your organization, yet outside your program, might work well. For instance, when evaluating education programs, staff from a curriculum and instruction office implementing a program might partner with staff from another office within the school district, such as an assessment or evaluation office, to conduct the evaluation.

If resources are not available for an external evaluator and there is no office or department in your organization that is not affected by your program, you may want to consider other potentially affordable evaluation options. You could put out a call to individuals with evaluation experience within your community who might be willing to donate time to your program; contact a local university or community college regarding faculty or staff with evaluation experience who might work with you at a reduced rate; ask your



QUICK CHECK

1. What is formative evaluation? How does formative evaluation differ from summative evaluation?
2. How can implementation assessment be used to make formative and summative evaluation decisions?
3. Why might someone be skeptical of an evaluation conducted by an internal evaluator? What can be done to strengthen the perceived objectivity when an internal evaluator is used?

local university if there is a doctoral student in evaluation who is looking for a research opportunity or dissertation project; or explore grant opportunities that fund evaluation activities. Overall, it is important to remember that both internal and external evaluations have their benefits and drawbacks. In determining the structure of who conducts an evaluation, weigh the extent to which *perceived* objectivity is a threat to evaluation credibility, as well as the ways in which different stakeholder groups might use the findings.

1.6 EMBEDDING EVALUATION INTO PROGRAMS

The resource tug-of-war between program services and program evaluation is real and has real implications. It is this dilemma that has shaped the way in which I work with my clients, so that evaluation is useful in not only determining the outcomes of their programs but also in helping to improve their programs on an ongoing basis. Embedded evaluation is an evaluation approach that can be built into programs and processes, so that it is part of everyday practice. This method recognizes the preciousness of resources and time, the need for information, and the tension between the two. The embedded approach to evaluation is not an additional step to be superimposed upon a program and the strategies it employs, but rather a way to weave evaluation into the design, development, and implementation of policies, programs, and projects.

1.6.1 Grounded in Continuous Improvement

Embedded evaluation incorporates the underlying philosophies of both total quality management (TQM) and quality improvement (QI) initiatives, in that the purpose of embedding evaluation into your programs is to create continuous improvement processes. Thus, **embedded evaluation** is a method of continuous improvement in which processes and practices are examined and refined to improve outcomes.

If you are not familiar with TQM or QI, TQM is a philosophy and an approach used to improve processes within organizations. See the American Society for Quality (ASQ) website for more information on TQM (asq.org). TQM is based on quality improvement principles. QI is concerned with improving performance in a systematic and continuous manner. Its processes are also referred to as continuous improvement (CI). The U.S. Department of Health and Human Services (Health Resources and Services Administration, 2011) has made available a resource on the principles and processes of QI. This report, titled *Quality Improvement*, explains QI and provides practical guidance in creating and implementing QI programs.

Embedded evaluation: an evaluation approach based on continuous improvement, in which program processes and practices are examined and refined to improve outcomes.

1.6.2 Theory Based and Utilization Focused

The embedded evaluation approach presented in this textbook is one of many approaches that can be used when conducting an evaluation (note that Chapter 4 provides a comprehensive review of evaluation approaches). Embedded evaluation combines elements

from several common evaluation approaches, including theory-based evaluation, logic modeling, stakeholder evaluation, participatory evaluation, and utilization-focused evaluation. Theory-based evaluation, in particular, focuses on indicators related to the logic underlying a program to guide evaluation. Utilization-focused evaluation is based on the premise that an evaluation's worth rests in how useful it is to the program's stakeholders. Both theory-based and utilization-focused evaluation approaches, as well as stakeholder and participatory evaluation, will be described in detail in Chapter 4.

1.6.3 Dynamic and Cyclical

Earlier I mentioned that evaluation is a lot like the scientific method: You define a problem, investigate the problem, document results, refine the problem based on lessons learned from the results, investigate again, and so on. So, while the steps of embedded evaluation presented in this text may appear as if they are linear rungs on a ladder culminating with the final step, they are not rigid steps. Rather, embedded evaluation steps build on each other and depend upon decisions made in prior steps, and information learned in one step may lead to refinement of another step. So, like the scientific method, embedded evaluation is cyclical. The steps of embedded evaluation are components of the evaluation process that impact and influence each other. What you learn or decide in one step may prompt you to return to a previous step for modification and improvement. Just as programs are ongoing, evaluation is dynamic.

The dynamic nature of evaluation and the interconnectedness of an embedded evaluation with the program itself may seem amiss to researchers who prefer to study a phenomenon over time and wait until a predefined time to analyze and report findings. And inarguably, having a program stay its course without mid-course refinements and improvements would make cross-site comparisons and replication easier. However, as stated previously, embedded evaluation is built upon the principle of continuous program improvement. With embedded evaluation, as information is gathered and lessons are learned, the program is improved. However, embedded evaluation goes beyond simply program monitoring. It is a way to build evaluation into a program, as well as to monitor implementation and assess effectiveness.

The focus of embedded evaluation is to enable program staff to build and implement high-quality programs that are continuously improving, as well as to determine when programs are not working and need to be discontinued. The overall purpose of designing a rigorous, embedded evaluation is to aid program staff in providing effective services to their clients.

1.6.4 Program Specific

Just as the first step in solving a problem is to understand the problem, the first step in conducting an evaluation is to *understand what you want to evaluate*. For the purposes of this textbook, what you want to evaluate is referred to as the “program.” As noted earlier, the term “program” is used broadly throughout this textbook to represent small

interventions, groups of activities, community-based services, agency-wide projects, and statewide initiatives, as well as national or international policy.

You can use the evaluation process that is presented in this textbook to define and evaluate a small project, as well as to understand and evaluate the inner workings of large programs and initiatives. Regardless of the size or type of program, understanding the program is not only the first step in evaluation; it is the most important step. Defining why a program should work and making the theory that underlies a program explicit lay the foundation upon which you can foster program improvement and measure program effectiveness. Further, understanding the program enables you to develop evaluation questions and define metrics, in collaboration with stakeholders, that are meaningful and useful to stakeholders. Understanding how the program operates can also aid you in integrating processes for the collection and use of these indicators into everyday program operation.

1.6.5 A Framework for Evaluation

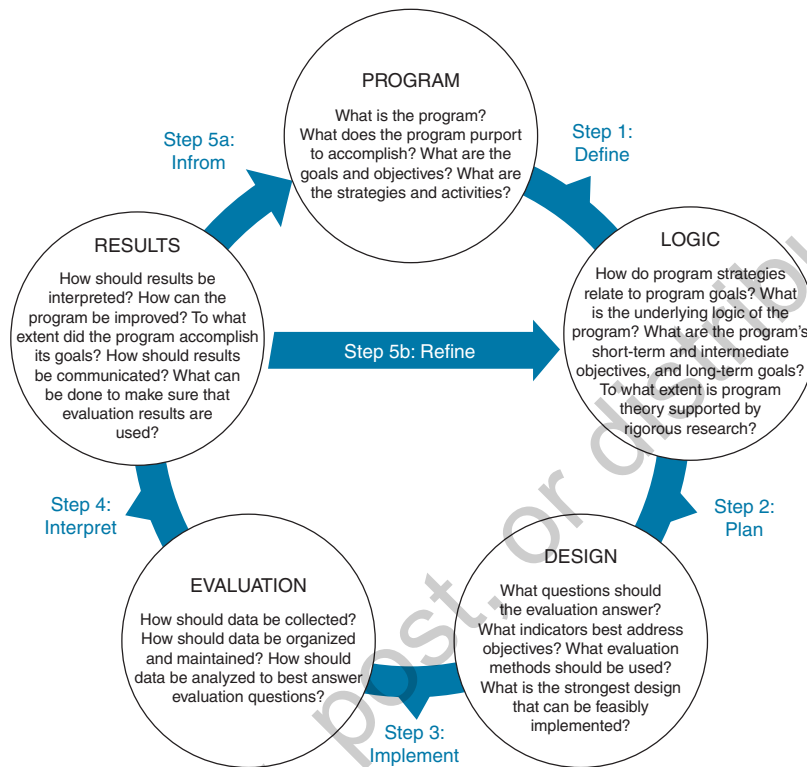
Embedded evaluation is a framework grounded in continuous improvement, based on a program's theory, focused on utilizing results, dynamic and cyclical in its operation, and built into a specific program's operations to foster data-driven decision making. Chapters 5–12 will guide you through designing and conducting an evaluation using this framework. You will be led step-by-step from documenting how and why a program works to using evaluation results. The embedded evaluation framework is presented graphically in Figure 1.3. The framework is based on the following five steps:

- Step 1. DEFINE: What is the program? (Chapters 5–6)
- Step 2. PLAN: How do I plan the evaluation? (Chapters 7–8)
- Step 3. IMPLEMENT: How do I evaluate the program? (Chapters 9–10)
- Step 4. INTERPRET: How do I interpret the results? (Chapter 11)
- Step 5. (a) INFORM and (b) REFINE: How do I use the results? (Chapter 12)

Prior to embarking upon the embedded evaluation process, Chapters 2–4 will provide a necessary foundation for future evaluators. This foundation includes a contextual understanding of the history of evaluation and its development over time; an awareness of the ethical obligations of an evaluator and the history of ethical abuses that make this awareness necessary; and a conceptual understanding of different approaches to evaluation.

Whether the program is a new program or one that has been in operation for many years, the process of embedding evaluation into your program is the same. Explicitly defining the program is a critical first step toward responsible program management, as well as program improvement. Program staff and program evaluators should have a clear understanding of the program and its intended goals, as well as how and why the strategies that the program employs relate to the program's goals.

FIGURE 1.3 The Embedded Evaluation Model



1.6.5.1 Embedding Evaluation Into Program Development

For a new program, embedding evaluation into the program development process allows data to be built into all future decision making for the program. It provides the opportunity for information to be the foundation of the program's operation from day one. Embedding evaluation during program development also provides the most flexibility with evaluation design, often allowing a more rigorous evaluation than may be feasible with an existing program. When evaluators are involved from the very beginning of a program's development, it provides an opportunity for the evaluator to work collaboratively with program staff to integrate evaluation into the program's operation and to build capacity within the program itself to use and rely upon data for decision making.

1.6.5.2 Embedding Evaluation Into Existing Programs

Existing programs with good documentation and established data management systems may find embedding evaluation into the program a relatively straightforward and educational process. Existing programs with poor documentation and little data supporting their operation may find the process similar to that of embedding evaluation into a new program.

Taking the time to document the logic of an existing program can not only clarify all aspects of a program's implementation, but also provide a good opportunity to examine existing strategies and their relation to the program's goals. The process of embedding evaluation into existing programs can also aid in developing a common understanding of program goals and help to foster buy-in among stakeholders. While examining the program's logic, you will likely uncover data needs that must be adopted by the program. Fostering broad stakeholder involvement during the embedded evaluation process often makes any additional data collection needs easier to implement. However, even if changing data collection methods is cumbersome, remember, it is the responsibility of an evaluator to provide program staff with the information necessary to ensure the program is working for program participants. Just as any organization must periodically reexamine and reaffirm its mission, all programs should routinely examine the logic underlying a program and refine that logic as necessary as lessons are learned and results are measured.

1.7 TEXTBOOK ORGANIZATION

This textbook will provide you with the tools to embed evaluation into programs to foster continuous improvement, by making information and data the basis upon which the program operates. The textbook is divided into four sections:

- **Section 1** includes general evaluation background, including information on key terms, the history of evaluation (Chapter 2), ethical considerations in evaluation (Chapter 3), and evaluation approaches (Chapter 4).



QUICK CHECK

1. Embedded evaluation is
 - a. An evaluation approach used to continuously improve the program.
 - b. An evaluation approach that focuses solely on the program's staff.
 - c. A linear approach to evaluation.
2. The first step in evaluation is to
 - a. Collect data about the program.
 - b. Decide on the methods to be used.
 - c. Understand the program.
3. What common method is the embedded evaluation approach similar to? In what ways is it similar to this method?

Answers: 1-a; 2-c; 3-scientific method

FIGURE 1.4 AEA Guiding Principles for Evaluators

AEA Guiding Principles for Evaluators

Glossary

Common Good — the shared benefit for all or most members of society including equitable opportunities and outcomes that are achieved through citizenship and collective action. The common good includes cultural, social, economic, and political resources as well as natural resources involving shared materials such as air, water, and a habitable earth.

Contextual Factors — geographic location and conditions; political, technological, environmental, and social climate; cultures; economic and historical conditions; language, customs, local norms, and practices; timing; and other factors that may influence an evaluation process or its findings.

Culturally Competent Evaluator — “[an evaluator who] draws upon a wide range of evaluation theories and methods to design and carry out an evaluation that is optimally matched to the context. In constructing a model or theory of how the evaluator operates, the evaluator reflects the diverse values and perspectives of key stakeholder groups.”¹

Environment — the surroundings or conditions in which a being lives or operates; the setting or conditions in which a particular activity occurs.

Equity — the condition of fair and just opportunities for all people to participate and thrive in society regardless of individual or group identity or difference. Striving to achieve equity includes mitigating historic disadvantage and existing structural inequalities.

Guiding Principles vs. Evaluation Standards — The Guiding Principles pertain to the ethical conduct of the evaluator whereas the Evaluation Standards pertain to the quality of the evaluation.

People or Groups — those who may be affected by an evaluation including, but not limited to, those defined by race, ethnicity, religion, gender, income, status, health, ability, power, underrepresentation, and/or disenfranchisement.

Professional Judgment — decisions or conclusions based on ethical principles and professional standards for evidence and argumentation in the conduct of an evaluation.

Stakeholders — individuals, groups, or organizations served by, or with a legitimate interest in, an evaluation, including those who might be affected by an evaluation.

¹The quotation is from the “Public Statement on Cultural Competence in Evaluation,” by the American Evaluation Association, 2011, Washington DC: Author, p. 3.

Purpose of the Guiding Principles: The Guiding Principles reflect the core values of the American Evaluation Association (AEA) and are intended as a guide to the professional ethical conduct of evaluators.

Focus and Interconnection of the Principles: The five Principles address systematic inquiry, competence, integrity, respect for people, and common good and equity. The Principles are interdependent and interconnected. At times, they may even conflict with one another. Therefore, evaluators should carefully examine how they justify professional actions.

Use of Principles: The Principles govern the behavior of evaluators in all stages of the evaluation from the initial discussion of focus and purpose, through design, implementation, reporting, and ultimately the use of the evaluation.

Communication of Principles: It is primarily the evaluator’s responsibility to initiate discussion and clarification of ethical matters with relevant parties to the evaluation. The Principles can be used to communicate to clients and other stakeholders what they can expect in terms of the professional ethical behavior of an evaluator.

Professional Development about Principles: Evaluators are responsible for undertaking professional development to learn to engage in sound ethical reasoning. Evaluators are also encouraged to consult with colleagues on how best to identify and address ethical issues.

Structure of the Principles: Each Principle is accompanied by several sub-statements to amplify the meaning of the overarching principle and to provide guidance for its application. These sub-statements do not include all possible applications of that principle, nor are they rules that provide the basis for sanctioning violators. The Principles are distinct from Evaluation Standards and evaluator competencies.

Evolution of Principles: The Principles are part of an evolving process of self-examination by the profession in the context of a rapidly changing world. They have been periodically revised since their first adoption in 1994. Once adopted by the membership, they become the official position of AEA on these matters and supersede previous versions. It is the policy of AEA to review the Principles at least every five years, engaging members in the process. These Principles are not intended to replace principles supported by other disciplines or associations in which evaluators participate.

AEA Guiding Principles

A: Systematic Inquiry: Evaluators conduct data-based inquiries that are thorough, methodical, and contextually relevant.

- A1. Adhere to the highest technical standards appropriate to the methods being used while attending to the evaluation's scale and available resources.
- A2. Explore with primary stakeholders the limitations and strengths of the core evaluation questions and the approaches that might be used for answering those questions.
- A3. Communicate methods and approaches accurately, and in sufficient detail, to allow others to understand, interpret, and critique the work.
- A4. Make clear the limitations of the evaluation and its results.
- A5. Discuss in contextually appropriate ways the values, assumptions, theories, methods, results, and analyses that significantly affect the evaluator's interpretation of the findings.
- A6. Carefully consider the ethical implications of the use of emerging technologies in evaluation practice.

B: Competence: Evaluators provide skilled professional services to stakeholders.

- B1. Ensure that the evaluation team possesses the education, abilities, skills, and experiences required to complete the evaluation competently.
- B2. When the most ethical option is to proceed with a commission or request outside the boundaries of the evaluation team's professional preparation and competence, clearly communicate any significant limitations to the evaluation that might result. Make every effort to supplement missing or weak competencies directly or through the assistance of others.
- B3. Ensure that the evaluation team collectively possesses or seeks out the competencies necessary to work in the cultural context of the evaluation.
- B4. Continually undertake relevant education, training, or supervised practice to learn new concepts, techniques, skills, and services necessary for competent evaluation practice. Ongoing professional development might include formal coursework and workshops, self-study, self- or externally-commissioned evaluations of one's own practice, and working with other evaluators to learn and refine evaluative skills and expertise.

C: Integrity: Evaluators behave with honesty and transparency in order to ensure the integrity of the evaluation.

- C1. Communicate truthfully and openly with clients and relevant stakeholders concerning all aspects of the evaluation, including its limitations.
- C2. Disclose any conflicts of interest (or appearance of a conflict) prior to accepting an evaluation assignment and manage or mitigate any conflicts during the evaluation.
- C3. Record and promptly communicate any changes to the originally negotiated evaluation plans, the rationale for those changes, and the potential impacts on the evaluation's scope and results.
- C4. Assess and make explicit the stakeholders', clients', and evaluators' values, perspectives, and interests concerning the conduct and outcome of the evaluation.
- C5. Accurately and transparently represent evaluation procedures, data, and findings.
- C6. Clearly communicate, justify, and address concerns related to procedures or activities that are likely to produce misleading evaluative information or conclusions. Consult colleagues for suggestions on proper ways to proceed if concerns cannot be resolved, and decline the evaluation when necessary.
- C7. Disclose all sources of financial support for an evaluation, and the source of the request for the evaluation.

(continued)

FIGURE 1.4 (Continued)

D: Respect for People: Evaluators honor the dignity, well-being, and self-worth of individuals and acknowledge the influence of culture within and across groups.

- D1. Strive to gain an understanding of, and treat fairly, the range of perspectives and interests that individuals and groups bring to the evaluation, including those that are not usually included or are oppositional.
- D2. Abide by current professional ethics, standards, and regulations (including informed consent, confidentiality, and prevention of harm) pertaining to evaluation participants.
- D3. Strive to maximize the benefits and reduce unnecessary risks or harms for groups and individuals associated with the evaluation.
- D4. Ensure that those who contribute data and incur risks do so willingly, and that they have knowledge of and opportunity to obtain benefits of the evaluation.

E: Common Good and Equity: Evaluators strive to contribute to the common good and advancement of an equitable and just society.

- E1. Recognize and balance the interests of the client, other stakeholders, and the common good while also protecting the integrity of the evaluation.
- E2. Identify and make efforts to address the evaluation's potential threats to the common good especially when specific stakeholder interests conflict with the goals of a democratic, equitable, and just society.
- E3. Identify and make efforts to address the evaluation's potential risks of exacerbating historic disadvantage or inequity.
- E4. Promote transparency and active sharing of data and findings with the goal of equitable access to information in forms that respect people and honor promises of confidentiality.
- E5. Mitigate the bias and potential power imbalances that can occur as a result of the evaluation's context. Self-assess one's own privilege and positioning within that context.



- **Section 2** includes a step-by-step approach to designing an embedded evaluation. It is not intended to be simply a “how to” lesson but rather a comprehensive approach to support you in planning and understanding programs, with a rigorous evaluation included as an integral part of the program’s design. The section includes understanding the program (Chapter 5), modeling the program (Chapter 6), planning the evaluation (Chapter 7), and designing the evaluation (Chapter 8).
- **Section 3** focuses on the post-design phases of evaluation, including conducting the evaluation (Chapter 9), analyzing data (Chapter 10), interpreting results (Chapter 11), and using evaluation findings (Chapter 12).
- **Section 4** provides several case studies.

1.8 CHAPTER SUMMARY

Evaluation is a method used to determine the value or worth of something. In our case, that “something” is a program. A **program** is defined broadly in this text to include a group of activities ranging from a small intervention to a national or international policy. **Program evaluation** is evaluation used to determine the merit or worth of a program.

Evaluation has the following attributes, components, and purposes:

- **Systematic:** logical and organized; undertaken according to a plan.
- **Operations:** processes involved with implementing the activities of a program.
- **Outcomes:** results that occur during and after implementing a program.
- **Standard:** the target used, implicitly or explicitly, to judge the merit or worth of a program.
- **Improving programs and policies:** the purpose of evaluation; to create more effective and efficient programs and policies.

Evaluation is a type of **research**, a systematic investigation in a field of study. **Embedded evaluation**, in particular, is an evaluation approach based on continuous improvement, in which program processes and practices are examined and refined to improve outcomes. A **stakeholder** is anyone who has an interest in or is involved with the operation or success of a program. Key stakeholder groups often include program staff, program participants, community members, and policymakers.

A **request for proposal** (RFP) is a solicitation for organizations to submit a proposal on how they would complete a specified project. Evaluation RFPs often ask for formative and summative evaluation. **Formative evaluation** is evaluation aimed at providing information to improve a program while it is in operation. Formative evaluation techniques include

- **Process evaluation:** formative evaluation aimed at understanding the operations of a program.

- **Implementation assessment:** formative evaluation that examines the degree to which a program is implemented with fidelity (according to plan).
- **Needs assessment:** formative evaluation that focuses on what services are needed and who needs them.
- **Evaluability assessment:** formative evaluation used to determine if an evaluation is feasible and the role stakeholders might take in shaping the evaluation design.

Summative evaluation is evaluation aimed at providing information about effectiveness, to make decisions about whether to continue or discontinue a program. Summative evaluation techniques include

- **Outcome evaluation:** summative evaluation aimed at measuring how well a program met its stated goals.
- **Impact evaluation:** summative evaluation that measures both the intended and unintended outcomes of a program.
- **Cost-benefit/cost-effectiveness analysis:** summative evaluation that focuses on estimating the efficiency of a program in terms of dollar costs saved (cost-benefit) or outcomes measured (cost-effectiveness).
- **Meta-analysis:** summative evaluation that integrates the effects of multiple studies to estimate the overall effect of a program.
- **Meta-evaluation:** an evaluation of an evaluation.

Finally, an **internal evaluator** is an evaluator employed by the organization that operates a program (but preferably not responsible for the program itself). An **external evaluator** refers to an evaluator who is employed from outside of the organization in which the program operates.

Reflection and Application

1. Research the Drug Abuse Resistance Education (DARE) program. What have past evaluations found regarding the effectiveness of DARE?
2. Why do you think policymakers continue to support programs even after evidence suggests they are ineffective?
3. Explain why it might be considered unethical to not evaluate a program or policy.
4. Find a program or policy where data showed it was not working as intended or had unintended consequences. Was the program continued?
5. Describe embedded evaluation and how it is used.
6. Identify the steps of embedded evaluation.