

# 2

## EVALUATION



SDI Productions/E+/Getty Images

### IN THIS CHAPTER: OBJECTIVES

1. Define evaluation and distinguish it from research in terms of purpose, method, and use.
2. Explain the history of evaluation and current theoretical models that align with the five philosophical paradigms described in Chapter 1.
3. Describe the following approaches to evaluation: the CIPP (context, input, process, product) model, responsive evaluation, theory-based evaluation, impact evaluation, participatory evaluation, utilization-focused evaluation, Real World Evaluation, developmental evaluation, empowerment evaluation, inclusive evaluation, feminist evaluation, culturally responsive evaluation, and postcolonial and Indigenous evaluation.
4. Identify the steps needed to plan an evaluation in terms of what is to be evaluated, the purpose of the evaluation, the stakeholders in the evaluation, constraints affecting the evaluation, the evaluation questions, selection of an evaluation model, data collection specification, analysis and interpretation strategies, utilization, management of the evaluation, and meta-evaluation plans.

5. Discuss the ethical guidelines from the American Evaluation Association.
6. Identify the questions needed to critically analyze evaluation studies based on the *Program Evaluation Standards* (Yarbrough et al., 2011).

## DEFINING EVALUATION

Evaluators have not been immune to the generally pluralistic and sometimes contentious spirit prevailing in the research community. This is partially because evaluations are often (but certainly not exclusively) conducted on programs designed to help oppressed and troubled people in a context of scarce resources and different opinions. The direct relationship between the evaluation of social and educational programs and access to resources sets the stage for tensions that can sometimes result in conflicts. Ernie House (1993) captures this spirit in his description of the evolution of evaluation:

Gradually, evaluators recognized that there were different interests to be served in an evaluation and that some of these interests might conflict with one another. The result was pluralist conceptions of evaluation in which multiple methods, measures, criteria, perspectives, audiences, and interests were recognized. Conceptually, evaluation moved from monolithic to pluralist conceptions, reflecting the pluralism that had emerged in the larger society. How to synthesize, resolve, and adjudicate all these multiple multiples remains a formidable question, as indeed it does for the larger society. Evaluation, which was invented to solve social problems, was ultimately afflicted with many of the problems it was meant to solve. (p. 11)

Given the tone of House's words, it should come as no surprise that even the definition of evaluation has been contested. Many definitions of evaluation have been proposed. In the *Encyclopedia of Evaluation* (Mathison, 2005), Fournier (2005) provided this as a general definition of evaluation:

Evaluation is an applied inquiry process for collecting and synthesizing evidence that culminates in conclusions about the state of affairs, value, merit, worth, significance, or quality of a program, product, person, policy, proposal, or plan. Conclusions made in evaluations encompass both an empirical aspect (that something is the case) and a normative aspect (judgment about the value of something). It is the value feature that distinguishes evaluation from other types of inquiry, such as basic science research, clinical epidemiology, investigative journalism, or public polling. (p. 140)

The terms *merit* and *worth* used in this definition of evaluation also need clarification. In the same *Encyclopedia of Evaluation* (Mathison, 2005), we find these definitions:

- Merit is the absolute or relative quality of something, either overall or in regard to a particular criterion. To determine the merit of an evaluand in regard to a particular criterion, it is necessary to collect relevant performance data and to explicitly ascribe value to it; that is, to say how meritorious the evaluand is in that particular dimension. To determine the overall merit of the evaluand, a further step is required: synthesis of performances with multiple criteria. Merit determination and synthesis are two of the core methodological tasks that distinguish evaluation from the collection and reporting of descriptive data for interpretation by others. (Davidson, 2005, p. 247)

- Worth is an outcome of an evaluation and refers to the value of the evaluand in a particular context, as opposed to the evaluand's intrinsic value, which is its merit. Worth and merit are not dependent on each other, and an evaluand (e.g., a doctor) may have merit (she is a highly skilled cardiologist) but have little worth (the hospital needs an anaesthesiologist). The opposite is also the case (the hospital has found an anaesthesiologist but not a very good one). The worth of an evaluand requires a thorough understanding of the particular context as well as the qualities and attributes of the evaluand. (Mathison, 2005, p. 452)

Michael Patton (2008) makes this distinction between merit and worth:

Merit refers to the intrinsic value of a program, for example, how effective it is in meeting the needs of those it is intended to help. Worth refers to extrinsic value to those outside the program, for example, to the larger community or society. A welfare program that gets jobs for recipients has merit for those who move out of poverty and worth to society by reducing welfare costs. (p. 113)

It is possible for a program to have merit but not worth. For example, if an agricultural program improves tobacco production, this would have merit for the farmers who increase their income. However, it would be viewed as having less worth by those who know that smoking causes cancer and increases medical costs.

In international development communities, a distinction is made between monitoring and evaluation. The United Nations Evaluation Group (2016) provides the following definitions that are relevant in evaluations undertaken in international development:

- Monitoring is “management’s continuous examination of any progress achieved during the implementation of an undertaking in order to track its compliance with the plan and to take necessary decisions to improve performance” (p. 30).
- “An evaluation is an assessment, conducted as systematically and impartially as possible, of an activity, project, programme, strategy, policy, topic, theme, sector, operational area or institutional performance. It analyses the level of achievement of both expected and unexpected results by examining the results chain, processes, contextual factors and causality using appropriate criteria such as relevance, effectiveness, efficiency, impact and sustainability. An evaluation should provide credible, useful evidence-based information that enables the timely incorporation of its findings, recommendations and lessons into the decision-making processes of organizations and stakeholders” (p. 10).

Shadish (1998) called for an expansion of the definition in terms of the purposes for which evaluations are done. His definition of evaluation included the idea that evaluation is defined in terms of its use of feasible practices to construct knowledge of the value of the evaluand that can be used to ameliorate the problems to which the evaluand is relevant. Even the part of the definition that refers to the purpose of the evaluation has been discussed and criticized. Sometimes, evaluations are done but no big decisions are made based on the results. Patton (2008) notes that evaluations can be used to reduce uncertainty about decisions that have to be made but that many other factors influence program decisions, such as availability of resources and the political climate. You can use the following list of questions to process the content of the various definitions and their implications for practice.

## EXTENDING YOUR THINKING

### Definitions of Evaluation

1. Several definitions of evaluation were presented in the preceding section. What do the definitions of program evaluation mean to you? Explain in your own words.
2. How do the concepts of merit and worth figure into your understanding of the meaning of evaluation?
3. Search for alternative definitions of program evaluation by examining other texts of web-based resources (see the American Evaluation Association's web page at [www.eval.org](http://www.eval.org) for good sources). What similarities and differences do you see between/among the definitions? What distinguishes one definition from the others?
4. Why is it important which definition of evaluation you use? What kind of power rests in the definition to tell you what to do or how to do it?
5. What are the implications of the various definitions for working in culturally diverse communities? What are the advantages and disadvantages of adopting different definitions in a culturally complex setting?
6. What are your reflections on the power related to who gets to decide which definition of evaluation is used?

## THE LANGUAGE OF EVALUATION

As with any discipline, evaluators have their own jargon. In this section, I explain some of the terms that evaluators use that are a bit different from the terms used by researchers.

When evaluators talk about the evaluand or object of the evaluation, they are talking about what it is that will be evaluated. This can include a social or educational program, a product, a policy, or personnel. Examples of the kinds of programs that are evaluated include enrichment programs for Deaf, gifted adolescents; drug and alcohol abuse programs for people experiencing homelessness; and management programs for high-level radioactive waste.

*Formative evaluations* are conducted primarily for the purposes of program improvement. Typically, formative evaluations are conducted during the development and implementation of the program and are reported to in-house staff who can use the information to improve the program. A *summative evaluation* is an evaluation used to make decisions about the continuation, revision, elimination, or merger of a program. Typically, it is done on a program that has stabilized and is often reported to an external agency.

*Internal evaluators* work within the organization that operates the program; *external evaluators* are “experts” brought in from outside the organization for the express purpose of conducting or assisting with the evaluation.

Evaluators must respond to the concerns and interests of selected members of the setting being evaluated (called the *stakeholders*). These include the program funders, the administrators, staff members, recipients of the services (and sometimes those who do not receive the services for various reasons), policymakers, and the wider community impacted by the evaluation. These are the audiences that the evaluators serve in the planning, conduct, and use of the evaluation study, compared with the scholarly, academic audience of the researcher.

As previously stated, evaluations can be conducted on social and educational policies, programs, products, or personnel. For purposes of this chapter, I focus on the evaluation of social and educational programs. References for individuals interested in personnel evaluation are provided at the end of this chapter.

## DISTINGUISHING RESEARCH AND EVALUATION

Given the definitions of evaluation already discussed, you may have an inkling of how research and evaluation differ. While there is much overlap between the worlds of research and evaluation, evaluation occupies some unique territory (Mertens, 2018; Mertens & Wilson, 2019). Greene (2000) writes about the commonalities that demarcate evaluation contexts and distinguish program evaluation from other forms of social inquiry (such as research). She argues that what distinguishes evaluation from other forms of social inquiry is its political inherency; that is, in evaluation, politics and science are inherently intertwined. Evaluations are conducted on the merit and worth of programs in the public domain, which are themselves responses to prioritized individual and community needs that resulted from political decisions. Program evaluation “is thus intertwined with political power and decision making about societal priorities and directions” (Greene, 2000, p. 982).

Trochim (n.d.) argues that evaluation is unique because of the organizational and political contexts in which it is conducted, which require skills in management, group processes, and political maneuvering that are not always needed in research. Mathison (2008) makes a strong claim that evaluation needs to be considered as a distinct discipline because of its historical emergence in the 1960s as a mechanism to examine valuing as a component of systematic inquiry as well as the ensuing development of methodological approaches that focus on stakeholder input and use of defined criteria (see the American Evaluation Association’s [2018] *Guiding Principles for Evaluators* and Yarbrough et al.’s [2011] *Program Evaluation Standards*, discussed later in this chapter).

Scriven (2003) adds a thoughtful evolution of this train of thought by describing evaluation as a transdiscipline because it is used in so many other disciplines. He writes, “Evaluation is a discipline that serves other disciplines even as it is a discipline unto itself, thus its emergent transdisciplinary status” (p. 22). He says evaluation is like such disciplines as statistics and ethics that have unique ways of approaching issues but are also used in other areas of inquiry such as education, health, and social work.

Mertens and Wilson (2019) recognize the uniqueness of evaluation as well as its overlap with applied research in education and the social sciences. While evaluation has contributed to our understanding of how to bring people together to address critical social issues, parallel developments have also been occurring in applied social research. Hence, “There is a place at which research and evaluation intersect—when research provides information about the need for, improvement of, or effects of programs or policies” (Mertens, 2009, p. 2). Thus, this provides additional rationale for including evaluation in this textbook as a major genre of systematic inquiry that borrows and enhances the methodologies developed in the research community.

### EXTENDING YOUR THINKING

#### Research Versus Evaluation

Locate a published evaluation study. Explain why that study should or should not be classified as an evaluation study (as opposed to a research study) based on the definition of *evaluation*. What do you derive from this exercise in terms of the difference between evaluation and research?

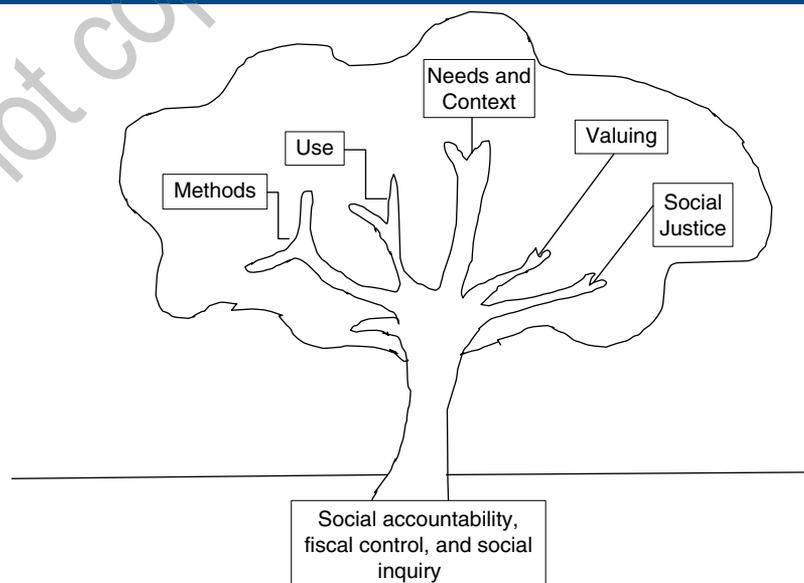
Reflecting on the quotation that appeared earlier in this chapter, you might be wondering what time period is referenced in House's (1993) remark that "gradually, evaluators recognized . . ." (p. 11). When did evaluators think differently, and what did they think? I now present you with a brief history of evaluation that provides the context for understanding theories and methods in evaluation.

## HISTORY AND MODELS OF EVALUATION

The origins of evaluation can be traced back to the 1800s, when the government first asked for external inspectors to evaluate public programs such as prisons, schools, hospitals, and orphanages (Stufflebeam & Shinkfield, 2007). However, most writers peg the beginning of the profession of evaluation, as it is now known, to the 1960s, with the passage of Great Society legislation (e.g., Head Start programs and the Elementary and Secondary Education Act) that mandated evaluations as a part of the programs. The history of evaluation is also complicated by its pluralistic disciplinary roots, with educational evaluators coming from a testing, assessment, and objectives-based evaluation background and psychologists more closely aligned with applied social research traditions (Mark et al., 2006).

Alkin (2013) depicted the historical roots of evaluation theories through a tree with three branches: use, methods, and valuing (see Figure 2.1). In a rough way, Alkin's three branches can be mapped on three of the major paradigms described in Chapter 1. The use branch equates to the pragmatic paradigm, the methods branch to the postpositivist paradigm, and the valuing branch to the constructivist paradigm. Alkin's version of the theory tree is useful in some respects; however, it has limited representation of evaluation theorists outside the United States, theorists of color, or Indigenous evaluators. To remedy these omissions, Mertens (2020) added the social justice branch that corresponds with the transformative paradigm, and Chilisa (2020)

FIGURE 2.1 ■ History of Evaluation Through the Theory Tree



Source: Chilisa (2020). Used with permission.

added the needs and context branch to represent the Indigenous paradigm. Hence, the history of evaluation is now told through five branches of theory and their associated paradigms.

### Postpositivist Paradigm: Methods Branch

Evaluation in both education and psychology began in the postpositivist paradigm. In education, evaluation emerged from a tradition of testing to assess student outcomes and progressed through an era of specification of objectives and measurement to determine if the objectives had been met. Ralph Tyler (cited in Stufflebeam et al., 2000) developed the objectives-based model for evaluation, and Malcolm Provus (cited in Stufflebeam et al., 2000) developed the discrepancy evaluation model. In psychology, the early years of evaluation were dominated by the work of Donald Campbell (see Shadish et al., 2002) in quasi-experimental design (a topic that is discussed extensively in Chapter 4 of this book). Contemporary evaluators have not abandoned the use of testing, objectives, and quasi-experimental designs, but they have modified and extended these strategies and added new approaches in the ensuing years. Lest we think that the postpositivist paradigm is a thing of the past, consider this statement from the U.S. Department of Education's (2003) *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*: "Well-designed and implemented randomized controlled trials are considered the 'gold standard' for evaluating an intervention's effectiveness, in fields such as medicine, welfare and employment policy, and psychology" (p. 1). The U.S. Department of Education also maintains a website called the What Works Clearinghouse (<https://ies.ed.gov/ncee/wwc/>) that lists educational programs that are deemed to have been rigorously evaluated through the use of randomized experimental designs. In addition, the international development community has adopted the practice of using experimental designs in evaluation; see the work of the Poverty Action Lab for multiple examples of this approach ([www.povertyactionlab.org](http://www.povertyactionlab.org)) and the International Initiative for Impact Evaluation ([www.3ieimpact.org](http://www.3ieimpact.org)).

Sample Study 2.1 used a randomized experimental design to evaluate the effectiveness of an intervention designed to increase representation of Black and Hispanic students in science, technology, engineering, and mathematics (STEM) programs (Cohodes et al., 2022).

### SAMPLE STUDY 2.1 SUMMARY OF A POSTPOSITIVIST METHODS BRANCH EVALUATION STUDY

**Evaluation Problem:** Black and Hispanic students are underrepresented in STEM programs. Pipeline programs that include summer programs for this population of high school students might result in greater representation in STEM, thus providing them with the potential for higher earnings in the future.

**Sample Evaluation Questions:** Which STEM summer intervention for Black and Hispanic high school students is effective in increasing students' likelihood of attending a highly ranked university, graduating, and earning a degree in STEM?

**Method:** The evaluation used a randomized controlled experimental design that involved random assignment of students to one of three experimental conditions or to a control group. The experiment was repeated three times in 2014, 2015, and 2016.

**Treatment:** The interventions consisted of a six-week on-site program at a highly ranked university, a one-week on-site program during the summer at a highly ranked university, a six-month online program with a short visit to the same university in the summer, and a control group.

**Participants:** The participants consisted of high-achieving, STEM-interested students with the following frequencies: one-week on-site program: 75–120 students per year; six-week on-site program: 80 students each year; six-month online program: 150–175 students each year. The report did not indicate how many students were in the control group. Almost all study participants identified as a member of a group “underrepresented in STEM fields, with 35 percent identifying as Black and 43 percent identifying as Hispanic” (Cohodes et al., 2022).

**Data Collection:** The dependent variables included college matriculation, completion, and graduation with a STEM degree. The evaluators gathered the data using program application admissions information from the highly ranked university, and college attendance and enrollment information from the National Student Clearinghouse.

**Results:** The “students offered seats in the STEM summer programs were more likely to enroll in a highly-ranked university and persist through and graduate from college. The programs also increased the likelihood that students graduate with a degree in a STEM field . . . The six- and one-week programs led to an increase in degrees from STEM fields, reflecting the overall increase in the number of degrees for STEM-interested students. The six-week program increased the rates at which students graduate in four years with the STEM degree by 13 percentage points (a 35 percent increase from a baseline of 37 percent). The shift to STEM is even larger when looking at five-year STEM degree attainment: the six-week program increases STEM degrees by 20 percentage points (a 44 percent increase from a baseline of 45 percent) and the one-week program by 15 percentage points (33 percent)” (Cohodes et al., 2022). The results for the six-month online program were also positive, but not as strong as the other two experimental conditions.

**Discussion:** The evaluators concluded that the summer programs encouraged students to apply to more universities and to more prestigious universities. This resulted in an increase in the probability that they would be accepted at such a university. Students gained knowledge about the admissions process and supportive resources. The study supports the idea that early intervention (i.e., at high school level) can increase the pipeline for students of color entering STEM programs.

---

Source: Based on Cohodes et al. (2022).

Additional contributions to postpositivist approaches to evaluation can be found in the extensions of Donald Campbell’s work from the writings on theory-based evaluation (Donaldson, 2007). Theory-based evaluation is an approach in which the evaluator constructs a model of how the program works using stakeholders’ theories, available social science theory, or both to guide question formation and data gathering. Theory-based evaluation is a way to mitigate the problems encountered in a more simplistic notion of quasi-experimental design when applied in an evaluation setting. The role of the evaluator is to bring a theoretical framework from existing social science theory to the evaluation setting. This would add insights to the structure of the program and its effectiveness that might not be available to the stakeholders in the setting. They warn against uncritically accepting the stakeholders’ viewpoints as the basis for understanding the effectiveness of the program. They acknowledge that qualitative methods can be used during the program conceptualization and monitoring, but they advocate the use of randomized experiments for assessing program impact. Because theory-based evaluation is guided by a preference to use structural modeling methods, the kinds of questions formulated for the evaluation are those that fit neatly into causal modeling.

Donaldson (2007) described program theory-driven evaluation as “the systematic use of substantive knowledge about the phenomena under investigation and scientific methods to improve, to produce knowledge and feedback about, and to determine the merit, worth, and

significance of evaluands such as social, educational, health, community, and organizational programs” (Donaldson & Lipsey, 2006, p. 67). A theory-driven program evaluation must first develop the program theory that guides the program in its definition of a solution to a social problem. The evaluator obtains information about the program and its underlying theory by reviewing prior research, making public the implicit theories held by those closest to the program, and observing the program in operation. After the program theory has been made explicit, the evaluator can then work with the stakeholders to make decisions about evaluation questions and methods.

### Constructivist Paradigm: Values Branch

Theorists such as Guba and Lincoln (1989), Patton (2002), Stake (2006), and House and Howe (1999) were influential in bringing the constructivist paradigm, along with qualitative methods, into evaluation. Stake’s work in responsive evaluation led the way to the introduction of the constructivist paradigm to the evaluation community. Guba and Lincoln (1989) acknowledge the foundation provided by Stake’s work in responsive evaluation in their development of constructivist approaches to evaluation.

Stake (2006) combined some of the elements of a utilization focus and discrepancy evaluation in his model of responsive evaluation. He includes the idea that evaluation involves comparing an observed value with some standard. The standard is to be defined by the expectations and criteria of different people for the program, and the observed values are to be based on those values actually held by the program. The evaluator’s job is to make a comprehensive statement of what the observed program values are with useful references to the dissatisfaction and satisfaction of appropriately selected people. He extended his work in the direction of case study methodology, thus further strengthening the place of qualitative methods in evaluation inquiry. Sample Study 2.2 provides an example of a constructivist, values branch evaluation of a program for youth to recover from substance abuse (Nelson et al., 2015).

Currently, the evaluation community seems to have reached a certain comfort level with a pluralistic approach to methodology. Qualitative methods for research with applicability to evaluation are discussed in Chapter 8.

## SAMPLE STUDY 2.2 SUMMARY OF A CONSTRUCTIVIST VALUES BRANCH EVALUATION

**Evaluation Problem:** Recovery support groups for youth with substance abuse problems frequently fail because the youth return to the same neighborhood and conditions that supported their drug use before the program. Very little evidence was available about the long-term effectiveness of such programs that were designed to provide ongoing support for youth through peer support.

**Sample Evaluation Question:** “What were the lived experiences of alumni clients of a recovery support group for youth who have experienced long-term sobriety?” (Nelson et al., 2015, p. 102).

**Method:** The evaluators conducted a phenomenological case study in order to understand how the program contributed to the participants’ long-term sobriety.

**Treatment:** The Palmer Drug Abuse Program (PDAP) was implemented in Houston, Texas, as a free substance abuse recovery support group specifically focused on treating adolescents. The program included peer support, 12-step meetings, social functions, individual counseling, family support, and psychosocial education.

**Participants:** Ten self-identified alumni of the program were contacted via a social media site for former clients. There were 6 females and 4 males; they had been associated with the program for 2 to 5 years. Their years of sobriety averaged 31.3 years.

**Data Collection:** Demographic data, including data about the alumni's alcohol/drug use, relapses, and years of sobriety, were collected via a survey. A structured interview protocol was used to collect qualitative data; the questions were all open-ended and designed to learn about the alumni's experiences in the program and in the ensuing years. How did the program help them? How did they sustain their sobriety?

**Results:** The results were summarized in terms of themes that emerged from the data. Participants spoke of the importance of relationships with themselves, peers, and a higher power, emphasizing the importance of hanging out with people who were nonusers of alcohol/drugs. A second theme revolved around descriptions of their life experiences before and after their participation in the program. They learned that they did have the skills necessary to recover from this disease with the help of others around them. They also expressed some resentment about the intrusiveness of the program into all aspects of their lives.

**Discussion:** The results support the importance of having the opportunity to be in contact with other teenagers to learn social skills, engage in recreation, and have a sense of belonging. As adolescents are more likely to relapse than adults after treatment, there are important implications of having a program that is not time limited but allows for ongoing support.

---

Source: Nelson et al. (2015).

## Transformative Paradigm: Social Justice Branch

Although constructivist qualitative evaluators recognize the importance that values play in the inquiry process, this paradigm does not specify any particular set of values. The transformative paradigm emerged in response to louder calls for consideration of the values of equity and justice.

*We advocate for the importance of deepening the meaning of credible evaluation practice and findings by bringing multiple philosophical and theoretical lenses to the evaluation process as a basis for the use of mixed methods in evaluation, thus providing evaluators with strategies for garnering more complex and diverse perspectives on the creation of credible evidence.*

—Mertens and Hesse-Biber, 2013, p. 5

*Without intentional and ongoing attention, both evaluation and efforts to advance equity can become one-off endeavors. Instead, both should be integrated into organizational routines. In particular, any initiative to advance equity and support underserved populations should include a program of ongoing evaluation.*

—Greyson, 2021, p. 2

Many years ago, House and Howe (1999) raised the question of what social justice and fairness mean in program evaluation. House and Howe (1999) and Howe and MacGillivray (2009) developed an approach to evaluation called deliberative democratic evaluation. They identified foundational conditions of deliberation: inclusion of all relevant interests, and dialogue so that the interests of various stakeholders can be accurately ascertained. These form the basis for a democratic deliberative evaluation with the capacity to equalize power relations in making evaluative judgments.

Hood and Hopson (2008) brought to visibility evaluators of color who made early contributions to representing this community in evaluation in their “Nobody Knows My Name” project. For example, Asa Hilliard’s work, cited by Hood and Hopson, illustrates the potency of his contribution to evaluation from an Afrocentric perspective:

In 1989, Hilliard combined his Afrocentric perspective with his observations about evaluation at the annual meeting of the American Evaluation Association in his keynote address entitled “Kemet (Egyptian) Historical Revision: Implications for Cross Cultural Evaluation and Research in Education.” In particular, he reminded evaluators that they are engaged in the “manipulation of power and therefore politics in its truest sense. Different approaches to evaluation can result in the painting of very different pictures of reality.” (p. 8)

Within the evaluation community, some evaluators have shifted the focus to prioritize social justice and fairness within evaluation, with the consequent opening of the door to the transformative paradigm of social inquiry for evaluators (Greene, 2000; Mertens, 2009). Transformative approaches to evaluation parallel many of the transformative approaches in research discussed in Chapter 1 and elsewhere in this text; hence, I offer a brief listing of approaches that reflect this paradigm, including transformative evaluation, transformative participatory evaluation, culturally responsive evaluation, feminist evaluation, and empowerment evaluation.

### Transformative Evaluation

Mertens (1999, 2020; Mertens & Wilson, 2019) developed an approach based on the transformative paradigm called inclusive evaluation, which evolved into what is now called transformative evaluation. There is a deliberate attempt to include groups that have historically experienced oppression and discrimination on the basis of gender, culture, economic levels, ethnicities/race, sexual orientation, and disabilities, with a conscious effort to build a link between the results of the evaluation and social action. To this end, transformative evaluation consciously addresses power imbalances in society by involving all relevant stakeholders in a way that is authentic and that accurately represents the stakeholders’ viewpoints. These evaluators are cognizant of issues of social justice that impact the definition of social problems. For example, they are aware that deficit models can place the blame for social problems on individuals or their culture rather than on the societal response to the individual or cultural group.

Transformative participatory evaluation is an important variation of transformative evaluation; it is akin to participatory action research, described in Chapter 8. It requires that the investigator be explicitly concerned with the intersectionality of gender, class, race, ethnicity, sexual orientation, and different abilities and the consequent meaning of these characteristics for access to power. While participatory evaluation means that evaluation is a participatory process that involves the stakeholders in the various tasks of the evaluation so that the results are fully comprehensible to project participants, not all participatory evaluations are transformative. The question of who is invited to participate in the process and the nature of their participation determines the extent to which participatory evaluation exemplifies the principles of the transformative paradigm.

The basic processes involved in the conduct of participatory evaluation provide a first step toward a transformative perspective in evaluation in that the professional evaluator works as a facilitator of the evaluation process but shares control and involvement in all phases of the research act with practitioners. In participatory evaluation, the evaluator helps train key organizational personnel in the technical skills vital to the successful completion of the evaluation project. These key organizational personnel—often administrators, counselors, or teachers—are taught sufficient technical knowledge and research skills to enable them to take on the coordinating role of continuing and new projects, with consultation with a professional evaluator as necessary.

The project participants are co-planners of the evaluation who complete the following tasks:

1. Discuss the evaluation questions that need to be addressed to determine if the project is making progress
2. Help define and identify sources of data required to answer evaluation questions
3. Participate in the data collection and analysis and in report preparation

The main goal of participatory evaluation is to provide information for project decision-makers and participants who will monitor the progress or improve their project.

### Culturally Responsive Evaluation

Hood and colleagues (2015) developed the culturally responsive evaluation approach based on their experiences with evaluation in African American communities. They place priority on understanding the cultural and historical context in which programs are situated as well as critiquing perspectives of community members that are based on deficit thinking. The culturally responsive approach to evaluation is well illustrated in the approach used to evaluate the talent development model of school reform (Thomas, 2004). These evaluators recognize the importance of matching salient characteristics of the evaluation team with the participants, establishing trusting relationships, and contributing to the overall progress of the educational enterprise in an ongoing manner. Expanding the Bench, a mentoring program designed to increase the number of evaluators of color, offered this expanded definition of culturally responsive and equitable evaluation:

Culturally responsive and equitable evaluation (CREE) requires the integration of diversity, inclusion, and equity in all phases of evaluation. CREE incorporates cultural, structural, and contextual factors (e.g., historical, social, economic, racial, ethnic, gender) using a participatory process that shifts power to individuals most impacted. CREE is not just one method of evaluation, it is an approach that should be infused into all evaluation methodologies. CREE advances equity by informing strategy, program improvement, decision-making, policy formation, and change. (Expanding the Bench Team and Advisory Team, 2019)

Sample Study 2.3 illustrates a transformative culturally responsive evaluation study conducted to support students who had been involved in the criminal justice system in their return to school (Drill et al., 2017).

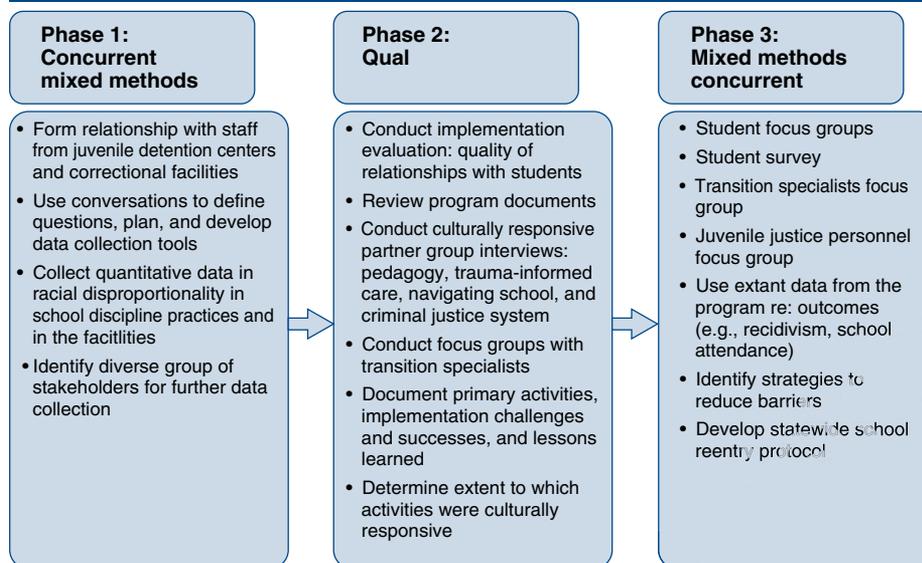
### SAMPLE STUDY 2.3 SUMMARY OF A TRANSFORMATIVE SOCIAL JUSTICE BRANCH MIXED METHODS EVALUATION STUDY

**Evaluation Problem:** Young people who have been detained by law enforcement face challenges when they try to move from the legal system back into the education system. The state of Oregon recognized this problem and provided funds for a program to support the successful transition from the legal system to education and community connections.

**Sample Evaluation Questions:** To what extent is the program being implemented as intended? How does the program remove barriers for returning students? To what extent are the program outcomes being achieved?

Copyright ©2024 by Sage Publications, Inc.

**FIGURE 2.2** ■ Transformative Mixed Methods Evaluation Study Using a Culturally Responsive Approach



Source: Drill et al. (2017).

**Method:** The evaluators used a culturally responsive approach to the evaluation that began with a contextual analysis inclusive of the history of the community and identification of the major stakeholders and their positions of power. They were attentive to language issues and cultural competency in this specific context.

**Treatment:** The Bars to Bridges program was developed through consultation with the members of the law enforcement community, educators, students, and parents. They had many activities that were adjusted as new information came to light about what was needed in this context. For example, groups were formed to support successful transition, meetings were held as youth exited the detention system, school officials discussed barriers, and law enforcement personnel were contacted to determine their ability to support reentry and crime prevention.

**Participants:** Participants consisted of detained African American, Black, and multiracial youth aged 11 through 21 in a school district in Oregon. As of March 2017, the program had served 131 students.

**Data Collection:** The data collection is complex and is portrayed in Figure 2.2.

**Results:** Overall, the students who worked with transition specialists in this program were able to return to school. The transition specialists were needed to provide many services for the youth, from basic needs to support for academic growth.

**Discussion:** The positive outcomes were appreciated by the students, their families, and all the other stakeholders. The transition specialists expressed frustration with some aspects of the program, such as not having adequate transportation services. They also felt that the program management data collection system was inefficient and not able to provide accurate information.

Source: Drill et al. (2017).

In addition to transformative evaluation and culturally responsive evaluation, other theoretical approaches that align with the transformative paradigm have been developed, including feminist evaluation and empowerment evaluation.

### Feminist Evaluation

Feminist evaluation includes judgments of merit and worth, application of social science methods to determine effectiveness, and achievement of program goals as well as tools related to social justice for the oppressed, especially, although not exclusively, women. Its central focus is on gender inequities that lead to social injustice. It uses a collaborative, inclusive process and captures multiple perspectives to bring about social change. As noted in Chapter 1, there is no single definition of feminism, nor a single definition of feminist evaluation. However, there are principles that have been identified across studies that claim to have a feminist orientation that are listed in Chapter 1 of this text.

### Empowerment Evaluation

Fetterman and Wandersman (2007) developed an approach called **empowerment** evaluation, defined as “the use of evaluation concepts, techniques, and findings to foster improvement and self-determination” (p. 186). Fetterman and Wandersman acknowledge that not all empowerment evaluations are meant to be transformative. Empowerment evaluation can be practical or transformative, much the same as Whitmore (1998) describes these two perspectives for participatory evaluation. Practical empowerment evaluation focuses on program decision-making and problem-solving. Transformative empowerment evaluation focuses on psychological transformation as well as political transformation. An underlying principle of this approach is that program participants conduct their own evaluations with an outside evaluator who often serves as a coach or facilitator, depending on internal program capacities. “Empowerment evaluation is an approach that aims to increase the probability of achieving program success by (1) providing program stakeholders with tools for assessing the planning, implementation, and self-evaluation of their program and (2) mainstreaming evaluation as part of the planning and management of the program/organization” (Wandersman et al., 2005, p. 28).

Transformative evaluators reflect on the meaning of *participatory* and *empowerment*:

When people use the words “participatory” and “empowerment,” not a lot of people get below the surface of talking about, well, what does that mean in terms of having a true understanding of the group you are working with? What does it mean in terms of appropriate ways that are considered to be culturally comfortable to establishing the questions for collection of data? What are the variations within that group? I do think there is danger in thinking I can take a strategy and just walk into a group and use a cookbook approach on how to do participatory or how to do empowerment without thinking about what does it mean that I am the researcher or evaluator in this process? There needs to be that sensitivity to the way you interact and how you engage people in a respectful way (Mertens, quoted in Edno et al., 2003).

Proponents of a transformative approach to evaluation argue that working within this paradigm can lead to more appropriate interventions and more judicious distribution of resources. The transformative paradigm in evaluation follows the same principles outlined in Chapter 1 for transformative research. Within the field of evaluation, these approaches explicitly address issues of power and representativeness of groups that have been traditionally pushed to the margins. Using the transformative paradigm as a base, the evaluator views each step in the evaluation process as an opportunity to raise questions about social justice, challenge the status quo, and bring in the voices of those who have been marginalized or inaccurately represented in previous evaluation studies.

## Pragmatic Paradigm: Use Branch

When evaluators discovered that “objective social science methods” were not sufficient to ensure that their work would have an effect on public policy and social program decisions, they shifted their focus to more decision-based models of evaluation. As evaluators gained experience with trying to improve social programs, other models of evaluation were developed that tried to address some of the shortcomings of traditional educational assessment or experimental designs. This branch of the evaluation tree roughly parallels Alkin’s (2013) use branch; the focus is on making the information from the evaluation useful to stakeholders. Stufflebeam (1983) was instrumental in extending the definition of evaluation beyond the achievement of objectives to include the idea that it was a process of providing information for decision-making. From his efforts with the Ohio State Evaluation Center and under the auspices of Phi Delta Kappa, Stufflebeam worked with other pioneers in the field of evaluation (including Egon Guba) to develop the CIPP (context, input, process, product) model of evaluation (Stufflebeam & Zhang, 2017). Thus, the CIPP model tries to incorporate many aspects of the program that were not considered under earlier models. The components of the model can be explained by the nature of the evaluation questions asked for each component (see Table 2.1).

The context and input phases represent a needs assessment function that evaluation sometimes plays to determine what is needed in terms of goals and resources for a program. For more extensive information about needs assessment, refer to Altschuld and Kumar’s (2010) work on this topic.

Other approaches that fit on the use branch include evaluations that focus on organizational culture (Preskill & Torres, 1999), Patton’s (2008) utilization-focused evaluation, and the international development model of evaluation called Real World Evaluation (Bamberger et al., 2006).

## Learning Organizational Evaluation

Preskill and Torres (1999) provide this definition situated in organizational culture: “We envision evaluative inquiry as an ongoing process for investigating and understanding critical organization issues. It is an approach to learning that is fully integrated with an organization’s work practices, and as such, it engenders (a) organization members’ interest and ability in exploring critical issues using evaluation logic, (b) organization members’ involvement in evaluative processes, and (c) the personal and professional growth of individuals within the organization” (pp. 1–2). Preskill (2008) expands this approach to evaluation by discussing methods for engaging organizational management in the cultivation of a culture that seeks

**TABLE 2.1 ■ CIPP Evaluation Questions**

| Component | Evaluation Questions  |
|-----------|---|
| Context   | What are the program’s goals? Do they reflect the needs of the participants?  |
| Input     | What means are required to achieve a given set of goals in terms of schedules, staffing, budget, and the like?  |
| Process   | How were the participants informed of the process? How were the resources allocated? How were the materials adapted?  |
| Product   | What is the evidence of outcomes? Should we terminate, continue, or revise this program? Should we decrease or increase funds? Should we merge it with another program? |

and uses information for improvement. She emphasizes the need to build capacity in organizations to participate in and learn from evaluations through such mechanisms as training, technical assistance, written materials, establishing communities of practice, providing internships and apprenticeships, coaching and mentoring, engaging in appreciative inquiry activities, and using technology.

### Utilization-Focused Evaluation

Patton (2008) developed an approach known as *utilization-focused evaluation* (UFE), defined as “evaluation done for and with specific intended primary users for specific, intended uses” (p. 37). Patton first published the UFE book in 1978, and this approach has been widely used. He clearly states that utilization should be considered from the very beginning of an evaluation study and that the quality of the evaluation is dependent on the actual use made of its findings. This was considered to be a radical idea in 1978; however, it has been accepted as common sense by many in the evaluation community today. According to Patton, the evaluator has a responsibility to facilitate the planning and implementation of evaluations to enhance the use of the findings. This of course necessitates the identification of the intended users (stakeholders) with whom the evaluator negotiates the type of information that the client would find useful. Sample Study 2.4 provides an example of a UFE with some aspects of developmental evaluation based on using the assumptions of the pragmatic paradigm in the use branch (Lawrence et al., 2018).

### Developmental Evaluation

Developmental evaluation is defined as evaluation that “supports innovative development to guide adaptation to emergent and dynamic realities in complex environments” (Patton, 2011, p. 1). Patton recognizes that this conceptualization of evaluation might be problematic for others in the evaluation community who view evaluation as a process of rendering judgment about whether a program’s goals have been met. He argues that this is one option for evaluators who wish to be valuable partners in the design process of programs for which the goals are emergent and changing and for which the purpose is learning, innovation, and change. Patton also places his own philosophical position in pragmatism and the use branch, so developmental evaluation can be implemented according to the assumptions of either the transformative paradigm or the pragmatic paradigm.

## SAMPLE STUDY 2.4 SUMMARY OF A PRAGMATIC USE BRANCH MIXED METHODS EVALUATION STUDY

**Evaluation Problem:** Students with disabilities (SWDs) and English language learners (ELLs) exhibit lower achievement than other children in many schools. In Massachusetts, this problem was addressed by developing a charter public school network. An evaluation was needed to guide development and determine effectiveness.

**Sample Evaluation Questions:** What are the appropriate systems, structures, and professional instructional capacities needed to better serve SWDs and ELLs? How many schools participated in the training programs? How did teams view their own school capacity as evidenced by class records?

**Method:** The evaluators used a pragmatic mixed methods design combined with a developmental evaluation approach with a focus on utilization. The evaluation was designed to support the program personnel’s decisions about what the program should look like by providing data throughout the development of the intervention. “The evaluation team worked with the directors (*co-creation and timely feedback*) to support adaptation and innovation

(developmental purpose, systems thinking, complexity), bringing our data collection and analysis tools (evaluation rigor), with a focus on actions (utilization)” (Lawrence et al., 2018, p. 76).

**Treatment:** The Massachusetts Charter Public School Association Capacity Building Network was funded by a grant from the U.S. Department of Education “to increase charter school capacity for English language learners (ELLs) and students with disabilities (SWDs)” (Lawrence et al., 2018, p. 71). The program included activities such as professional development, creation of a network of experts to provide necessary professional services, and provision of opportunities for charter schools to network amongst themselves. In addition to providing guidance in the development of the interventions, the evaluation focused on two program components: (1) Cluster Working Groups (CWGs) that consisted of representatives who met monthly to identify school strengths and needs for additional resources and expertise, and (2) Model Demonstration Schools (MDSs).

**Participants:** Fifty-six (70%) charter schools participated in at least one of the activities. The schools were quite diverse in terms of organizational structure, philosophy, geographic location, and student body. Data were collected from the three program directors, the four members of the evaluation team, MDS coaches, and individual school teams.

**Data Collection:** Participant feedback forms were used to collect data from the members of the CWGs, along with observations of the meetings and case studies conducted in a sample of the schools that included observation, interviews, and review of documents. Data on the MDSs were collected through review of documents, end-of-year summaries of learning implementation, feedback from the coaches, case study interviews, observations at the schools, and notes from evaluation/project director team meetings. Quantitative data were also obtained through state-run databases.

**Results:** Data were collaboratively analyzed periodically throughout the course of the evaluation in order to provide feedback about the emerging interventions. The program was adapted in response to the evaluation findings through collaborative development of solutions to identified problems. The program directors were able to question their own assumptions about what would work and what was needed to change in order to achieve the goals.

**Discussion:** Initially, members of the CWGs thought that they would be able to develop separate programs to serve SWDs and ELLs. As this was not the intent of the program, many of the school representatives dropped out of the CWG. The problem was addressed by providing professional development opportunities for school leaders who worked with SWDs and ELLs. This approach to evaluation is time intensive, and so evaluators had to be sensitive to the many demands that were placed on the school personnel through their regular responsibilities. Program directors actively sought evaluative information from the evaluation team as a basis for decision-making.

---

Source: Based on Lawrence et al. (2018).

## Real World Evaluation

*Real World Evaluation* is an approach that emerged from the world of international development evaluation in which, for better or worse, funding agencies often impose severe constraints in terms of time and money (e.g., an evaluator might be brought into a country for a short period of time to conduct an evaluation; Bamberger et al., 2006). Bamberger et al. (2006) offer strategies to adapt evaluation designs to meet the constraints imposed by the funding agency while still trying to be responsive to the cultural complexities in the evaluation context. For example, they suggest that a local consultant can be engaged to collect background data and conduct exploratory studies prior to the arrival of the outside evaluator. This might include preparation of reports on the social and economic characteristics of the targeted community, describing key features of the program to be evaluated, and a list of potential key informants and potential participants in focus groups. Other design features are similar to those discussed in the methodological chapters of this text.

## Indigenous Paradigm: Needs and Values Branch

Postcolonial and Indigenous approaches were developed by members of postcolonial and Indigenous communities, reflecting the cultural roots of American Indians (LaFrance & Crazy Bull, 2009); the Maori in New Zealand, also known as Aotearoa (Cram, 2009; L. T. Smith, 2012); and Africans (Chilisa, 2020). Evaluation protocols focus both inward and outward; they specify how Indigenous and non-Indigenous persons need to proceed when working with Indigenous groups as well as guide the way the evaluation work is done based on the values and traditions of the community. Human relations based on trust and respect are considered to be paramount, especially given the history of violence against Indigenous peoples.

As was explained in Chapter 1, Indigenous studies reflect tribal epistemologies and tribal affiliations. Indigenous evaluations need to be based on a “holistic worldview, a relational orientation to understanding the world, and valuing cultural protocols and norms” (Cram et al., 2018, p. 10). Indigenous evaluators seek to work collaboratively with communities while prioritizing inclusiveness, respect, and self-determination.

Indigenous evaluation “reflects traditional, cultural, and spiritual ways of knowing as a process where we respectfully listen and seek to understand first, and then decide on best pathways together” (Bremner & Bowman, 2021, p. 109). There is a need to respect cultural and intellectual influences of the study by means of “oral agreements, treaty rights, Indigenous theories and frameworks or ethics, nation-to-nation agreements, pre-contact practices, and so on” (p. 110). Investigative strategies need to be culturally specific and responsive to the needs of the community. The traditional wisdom of ancestors needs to provide the grounding for understanding the learnings achieved through the evaluation. Sample Study 2.5 illustrates the application of the Indigenous paradigm through an evaluation of a Native American–owned nonprofit (Native PRIDE) that addresses issues such as substance abuse and suicide (Kelley et al., 2022).

### SAMPLE STUDY 2.5 SUMMARY OF AN INDIGENOUS MIXED METHODS EVALUATION STUDY

**Evaluation Problem:** The COVID-19 pandemic disproportionately affected American Indian communities socially, culturally, and economically in ways that impacted their physical and mental well-being. There was an increased need to address issues such as “substance abuse, bullying, negative peer pressure, unhealthy relationships, suicidal ideation, and fatalities” (Kelley et al., 2022, p. 51) among American Indian youth.

**Sample Evaluation Questions:** How did COVID-19 affect the culturally based project (Native PRIDE) and tribal communities? How did the staff and tribal members adapt services to meet the program goals? What was felt, observed, and known during the COVID-19 pandemic?

**Method:** The evaluators used an Indigenous evaluation framework over a five-year period that included building community capacity for evaluation through training student interns and community members in research methods, survey development, data collection, data analysis, and dissemination of findings. They used a mixed methods culturally centered approach that involved a team approach for the development of the data collection instruments and interpretation of the data.

**Treatment:** The nonprofit had developed the Intergenerational Connections Project (ICP) to provide connections between elders and youth through participation in cultural and spiritual-based programs. The interventions took place with four tribes in Montana and South Dakota located in two schools, one recovery center, and one youth-serving organization. The

goal of the program is to increase cultural resilience and cultural participation with youth and elders. The program included mentoring, peer counseling, creating safe and sacred opportunities to facilitate changing unhealthy cycles of historical trauma from colonization, and strengthening the youth's skills. One important part of the treatment was the Good Road to Life (GRL) training that included both elders and youth "to strengthen cultural resilience, increase social support, develop leadership skills, and promote healthy relationships" (Kelley et al., 2021, p. 52). During the pandemic, the services were delivered by webinars, digital stories, and videos.

**Participants:** The authors do not report the number of Native PRIDE staff who were involved. They reported that survey data were collected from 76 people who completed the GRL training. Their tribal affiliations were the Ogalala Sioux Tribe and the Fort Peck Tribe. The respondents included youth who were trained, youth mentors, clan leaders, and mental health professionals.

**Data Collection:** The evaluator and the trained youth worked with the Native PRIDE team to develop data collection instruments that reflected the desired program impacts, tribal values, connection, and meaning. They focused on strategies used to rise above the impacts of COVID-19. "Qualitative measures included informal interviews conducted during site meetings with Native PRIDE staff, tribal site coordinators, and the evaluation team. In these interviews, staff reflected on the COVID-19 pandemic, lessons learned, and recommendations for future work . . . Quantitative measures included an online survey, sent to GRL program participants, that explored the impact of COVID-19 on tribal communities. The online survey link was disseminated by email to site coordinators who then forwarded it to youth and elders involved in the ICP and GRL trainings" (Kelley et al., 2021, pp. 53–54). The survey had 15 questions and was administered online.

**Results:** The quantitative data indicated that youth felt that the training and program participation increased their ability to make healthy choices, participate in cultural activities, and seek help when they needed it. The youth reported isolation, depression, anxiety, and frustration. The program helped them rise above these feelings through connections and cultural ceremonies. The evaluation team concluded that cultural and resilience-based mentoring and training can be delivered in a virtual format, but it is not as effective as in-person gatherings.

**Discussion:** The pandemic required adjustments in program delivery and in evaluation activities. They were able to continue with activities and data collection through virtual means, but both delivery of the program and conduct of the evaluation were diminished as compared to in-person activities. By focusing on resilience and strategies for rising above the adversity created during the pandemic, they were able to enhance skills and relationships between elders and youth and their communities.

---

Source: Kelley et al. (2022).

## EXTENDING YOUR THINKING

### Evaluation Approaches

Review the historical and current models of evaluation presented in this chapter. Select five models (one from each paradigm, perhaps). For each model, determine the theorist's viewpoint regarding the following:

- a. The purpose(s) of evaluation
- b. The role of the evaluator in making valuing judgments
- c. The role of the evaluator in making causal claims
- d. The role of the evaluator in accommodating to the political setting

Copyright ©2024 by Sage Publications, Inc.

- e. The role of the evaluator in providing information for decision-making
- f. The perception of the theorist about who is the primary audience for the evaluation
- g. The perception of the theorist about the appropriate role of stakeholders in the evaluation
- h. The perception of the theorist about the role of other parties affected by or interested in the evaluation
- i. The most appropriate way to train an evaluator based on that theoretical perspective

## RESOURCES AND PROCESSES FOR CONDUCTING EVALUATIONS

A general outline for steps in conducting evaluations is presented in this section. The student who is seriously considering conducting an evaluation study can find a list of helpful resources that expands their understandings at the end of this chapter.

## STEPS IN PLANNING AN EVALUATION

In some respects, the steps for conducting an evaluation parallel those used to conduct any research project. However, variations occur in terms of the process of inquiry because this is an *evaluation* study and because of such factors as the status of the program being evaluated and the model of evaluation chosen for use. The steps listed here are generic, and they will need to be adapted based on the paradigmatic framework used in the study. For example, the transformative paradigm and the Indigenous paradigm would both begin with phases of relationship building and contextual analysis. The steps include the following:

### Focusing the Evaluation

- A description of what is to be evaluated
- The purpose of the evaluation
- The stakeholders in the evaluation
- Constraints and opportunities associated with the evaluation
- The evaluation questions
- Selection of an evaluation model

### Planning the Evaluation

- Data collection specification, analysis, interpretation, and use strategies
- Management of the evaluation
- Meta-evaluation plans

### Implementing the Evaluation

- Completing the scope of work specified in the plan

What follows is a general description of the steps for focusing and planning an evaluation. I have integrated ideas from all five paradigms. Before launching into a discussion of the focusing stage, I want to acknowledge that all evaluators would not necessarily move through the

process in exactly the same way. Nevertheless, if these steps are viewed as a nonlinear, iterative framework for planning an evaluation, they should provide a helpful guide to that end for all evaluators, no matter their orientation. Furthermore, my remarks are probably biased toward the perspective of an external evaluator because that has been my perspective for the last 40 years or so; however, I did work as an internal evaluator for about 8 years prior to assuming the external status. I attempt to write information that would be pertinent and helpful to both internal and external evaluators.

## Steps for Planning an Evaluation Study

### Focusing Stage

During the focusing stage, the evaluator needs to determine what is being evaluated, the purpose of the evaluation, the stakeholders in the evaluation, and the constraints within which the evaluation will take place. In one sense, the evaluator is stuck with a “which came first, the chicken or the egg” dilemma, even in the first stage of the evaluation, in that just learning about what is to be evaluated implies contact with at least one group of stakeholders. Typically, the evaluator is contacted to perform an evaluation by some individual or agency representing one group of stakeholders in the program. Often, this first contact is initiated by a program director, policy-maker, or funding group. When listening to the description of the evaluand, purpose, stakeholders, and constraints within this initial context, the evaluator can gain valuable information by asking the right kinds of questions. These questions can provide a sufficient knowledge base to direct further planning efforts and to alert the initial contact person of things that might not have been thought of, such as theoretical perspectives or groups that need to be included because they will be affected by the program (Mertens, 2009; Mertens & Wilson, 2019).

**Description of the Evaluand.** The evaluand, you will recall, is what is being evaluated (e.g., a professional development initiative, a substance abuse prevention program, a multicultural education program, or a state agency policy). The evaluator needs to determine the status of the evaluand: Is it a developing, new, or firmly established program? If it is developing, it is possible that the evaluator will be asked to play a role in the evolution of the evaluand. Thus, an evaluator who is using the developmental evaluation model (Patton, 2011) might find that a considerable amount of time is spent collaboratively designing the evaluand. If it is new or firmly established, the program is probably described in printed documents. It is always a good idea to ask for whatever printed documents are available about the program in advance of your meeting with the initial stakeholders, if possible. Reading documents such as an annual report, accreditation agency report, previous evaluation reports, or a proposal can give you a good background about the evaluand and the context within which it functions. Questions to start with include the following:

- Is there a written description of what is to be evaluated?
- What is the status of the evaluand? Relatively stable and mature? New? Developing? How long has the program been around?
- In what context will (or does) the evaluand function?
- Who is the evaluand designed to serve?
- How does the evaluand work? Or how is it supposed to work?
- What is the evaluand supposed to do?

- What resources are being put into the evaluand (financial, time, staff, materials, etc.)?
- What are the processes that make up the evaluand?
- What outputs are expected? Or occur?
- Why is it being evaluated?
- Whose description of the evaluand is available to you at the start of the evaluation?
- Whose description of the evaluand is needed to get a full understanding of the program to be evaluated?

You should read available documents and hold initial discussions with the conscious realization that things are not always being played out exactly as they are portrayed on paper or in conversation. Therefore, it would behoove you to alert the client to the fact that you will want to observe the program in action (in its development, during implementation, etc.) to get a more accurate picture of what is being evaluated. You should also let the client know that you are aware that programs are generally not static (i.e., you will expect changes in the program throughout the duration of the evaluation) and that multisite programs will probably not be implemented in exactly the same way from site to site.

You should be aware that different individuals with different relationships with the program may view the program quite differently. Explain to your client that this often happens in evaluations and that you want to build in a mechanism to discover diverse views about the program and to explore these diversities within a public context. In this way, you can increase the chances that the program you think you are evaluating is the one that is actually functioning or developing. Without this assurance, you may find at the end of the evaluation that your results are challenged because you did not adequately represent different perceptions of the program.

Part of the description of the evaluand should include ascertaining diverse views of the program's purpose. For example, the evaluator can help program staff members, advisory committees, and partnership teams make "claims" statements regarding the program—that is, statements that represent the participants' perceptions of what would change as a result of the program activities. The evaluator leads the meeting by asking the stakeholders to comment on what they are committed to changing—and asks them to voice their real expectations for change. In this way, the evaluator hopes to increase the ownership by those responsible for the program activities for the intended changes. Sample Study 2.5 provides an excellent example of ways to enhance ownership of an evaluation through employing training strategies for a wide range of stakeholders (Kelley et al., 2022).

A common tool that evaluators use to depict the evaluand is called a logic model (W. K. Kellogg Foundation, 2004, 2017). A logic model is a graphic technique that allows the explicit depiction of the theory of change that underlies a program. It should reflect a model of how the program will work under certain conditions to solve an identified problem. Logic models can be quite simple, including three major components: the program that is delivered, the people who receive the program, and the results that are expected. The logic model is generally arranged in columns in chronological order, indicating a logical flow from one element to the next, but it is not unusual to have to revisit some of the earlier elements during the course of the evaluation. Evaluators in international development also use graphics similar to logic models, but they call them logframes—a type of table-based logic model (Davies, 2004). Many evaluators find logic models or logframes useful; however, they also acknowledge the limitations in trying to depict a complex, dynamic program in terms of a linear, two-dimensional graphic. A sample

**FIGURE 2.3 ■ Logic Model Bars to Bridges**

| Inputs/Resources   | Activities   | Outputs  | Outcomes   | Impact  |
|--|--|--|--|---|
| <ul style="list-style-type: none"> <li>Legislative authorization</li> <li>Funding</li> <li>Oregon Leadership Network Equity Plan</li> <li>Students and schools</li> <li>Community organizations and law enforcement</li> <li>Families of students</li> <li>Transition specialists</li> </ul> | <ul style="list-style-type: none"> <li>With students               <ul style="list-style-type: none"> <li>Lead groups to promote successful transition</li> <li>Meet as exiting detention system</li> </ul> </li> <li>With law enforcement               <ul style="list-style-type: none"> <li>Maintain contact</li> <li>Provide reentry and outreach information</li> </ul> </li> <li>With schools               <ul style="list-style-type: none"> <li>Identify barriers and opportunities</li> <li>Provide professional development</li> </ul> </li> <li>With families               <ul style="list-style-type: none"> <li>Coordinate outreach</li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li># of students served</li> <li># of law enforcement contacts</li> <li># of school personnel contacts and meetings held</li> <li># of families engaged in transition</li> </ul> | <ul style="list-style-type: none"> <li>Increase rate of reentry</li> <li>Increase school engagement</li> <li>Decrease absenteeism</li> <li>Decrease status violations</li> <li>Decrease recidivism</li> <li>Increase connections between justice system workers and schools</li> </ul> | <ul style="list-style-type: none"> <li>Increase access to path for graduation and entry into post-secondary education</li> <li>Increase trauma-informed pedagogy</li> <li>Decrease exclusionary discipline</li> <li>Raise awareness of disproportionality of discipline by race</li> <li>Increase family access to supports and prevention</li> </ul> |

Source: Adapted from Drill et al. (2017).

logic model based on the Drill et al. (2017) evaluation (Sample Study 2.3) can be found in Figure 2.3. Additional logic model examples can be found at the [website for this textbook](#).

**Purpose of the Evaluation.** Evaluations can be conducted for multiple purposes. As you discuss the evaluand, the purpose of the evaluation may begin to emerge. However, it is important to directly address this issue, within the context of what stimulated the need for the evaluation and the intended use of the evaluation results. The purpose of the evaluation needs to be distinguished from the purpose of the program.

One purpose of evaluation might be to provide information useful in improving the program; this is called **formative evaluation**. Another purpose might be to provide judgments about the program's worth or merit at the end of the program; this is called **summative evaluation**. Developmental evaluations are typically designed to provide systematic data in a fluid, dynamic, changing context of **program development**. Although summative evaluations tend to focus more on program impact, formative and developmental evaluations can include program impact data that are viewed as a barometer for program changes. Transformative evaluations are conducted to **address specific inequities and increase justice**. Indigenous evaluations are conducted to **increase justice, to promote healing from the effects of colonization, and to strengthen tribal relations with the land and each other**.

Questions can be asked to determine the purposes of the evaluation; the questions should be asked with an understanding of diversity of viewpoints and the necessity of representation of appropriate people in the discussions. Possible questions include these:

- What is the purpose of the evaluation?
- What events triggered the need for the evaluation?
- What is the intended use of the evaluation findings?
- Who will use the findings and for what purpose?
- Whose views should be represented in the statement of evaluation purpose?

Mertens and Wilson (2019) identify a number of possible purposes for evaluations:

- To gain insights or to determine necessary inputs before a program is designed, implemented, or revised (context evaluation, needs assessment, organizational assessment)
- To find areas in need of improvement or to change practices in terms of program implementation (process evaluation)
- To assess program effectiveness (impact evaluation)
- To address issues of human rights (transformative approaches)

Evaluators should be on the alert for evaluations conducted without a clear commitment to use the results, for which decisions have already been made and the decision-maker is looking for justification, or cases in which the evaluation is looked at only as a public relations activity (Patton, 2008). Of course, decision-makers are unlikely to characterize the purpose of the evaluation in such blunt terms, so the evaluator must be aware of establishing a clear purpose with intentions for use prior to commencing on the evaluation proper. Evaluators should also be aware that the purpose of the evaluation is something that may change as the project evolves.

**Identification of Stakeholders.** Of course, by now, you should know who the primary players in the program are. However, you should ask specific questions to ascertain whether all of the appropriate stakeholders have been identified. It is not uncommon in evaluations to address the concerns of the funding agency, the policymakers, or the program managers. However, this does not really cover all the people who are or will be affected by the program. Therefore, it is incumbent on you as the evaluator to raise questions about who should be involved in the evaluation process and to raise the consciousness of the powerful people to include those who have less power.

Possible questions at this stage include the following:

- Who is involved in the administration and implementation of the program?
- Who are the intended beneficiaries?
- Who has been excluded from being eligible for participation in the program?
- Who stands to gain or lose from the evaluation results?
- Which individuals and groups have power in this setting? Which do not?
- Did the program planning and evaluation planning involve members of marginalized communities?
- Who is representative of the people with the least power?
- What opportunities exist for mutual sharing between the evaluators and those without power?
- What opportunities exist for the people without power to criticize the evaluation and influence future directions?
- To what extent is there appropriate representation on the basis of gender, ethnicity, disability, and income levels?

- Who controls what the evaluation is about and how it will be carried out?
- Who has access to the program resources?
- What are the dominant cultural values and assumptions in the program?
- Who are the supporters of the program? Who are its opponents?
- Possible stakeholders in evaluations include the following:
  - Sponsors and funders
  - Governing and advisory boards
  - State- and federal-level agency representatives
  - Policymakers
  - Administrators (at various levels)
  - Staff members (e.g., regular education teachers, special education teachers, resource personnel, psychologists)
  - Service recipients (clients, students, etc.)
  - Nonrecipients (who may or may not be eligible but are not receiving services)
  - Parents of service recipients
  - Grassroots organization representatives in the community
  - Public community representatives

This list is meant not to be exhaustive but, rather, to give you an idea that the breadth of the impact of an evaluation can extend far beyond the individuals with whom you first discuss the evaluation project. Various strategies can be used to identify stakeholders. You might start with the people who invited you to do the evaluation and then ask them to nominate others. You might ask for people to become involved by some type of organizational or public announcement of the evaluation. You can often identify many stakeholders by reviewing the organizational chart of the program and asking questions of your early contacts about who is represented in the charts and who is not and why that is so.

**Constraints of the Evaluation.** Evaluations occur within specific constraints:

- Money: “We have budgeted this much money for the evaluation.”
- Time: “We need the results of the evaluation before the board meets six months from now.”
- Personnel: “The teachers will be given four hours per week release time to assist with data collection.”
- Existing resources: “Data program records are digitized, and they contain information about all recipients and their characteristics.”
- Politics: “If this program were reduced, it would free up additional funds to respond to other needs.”

Politics are integrally involved in the evaluation process. Therefore, the evaluator must be aware at the start—and sensitive throughout the process—of who supports or opposes the program; who would gain or lose if the program was continued, modified, reduced, or eliminated; who sanctions the evaluation, and who refuses to cooperate; who controls access to information; and who needs to be kept informed as the evaluation progresses. This would be a good time in the evaluation planning to bring up the issue of communication lines and mechanisms so that

you are sure there is a formalized understanding of who needs to be informed of what and by what means.

Also, it is important to note that people move in and out of programs, so it is not safe to assume that the understandings engendered by your initial efforts will be shared by those who enter the program later. You will need to orient those who enter the program later about the purpose, process, and so on of the evaluation. If you do not do it yourself, you might establish a mechanism to have other knowledgeable participants take on the task of orienting newcomers.

It is usually a good idea at this point to prepare a synthesis of your current understanding of a program, in terms of the evaluand, purpose of the evaluation, stakeholders, and constraints. You can then share this synthesis with the stakeholder groups, characterize it as your current understanding, and ask for their feedback.

### Evaluation Questions

Evaluation questions can be derived from the statement of purpose for the evaluation, expanded on by holding brainstorming sessions with stakeholder groups, borrowed from previous evaluation studies, or generated by a theoretical framework that is relevant to the study. The U.S. General Accounting Office, now known as the Government Accountability Office, examined categories of evaluation questions that would provide useful information for Congress (Shipman et al., 1995). I share these with you, as they seem to have generic relevance for evaluators in local settings as well.

*Descriptive questions* are those that tell what the program is and what it does:

- What activities does the program support?
- Toward what end or purpose?
- Who performs these activities?
- How extensive and costly are the activities, and whom do they reach?
- Are conditions, activities, purposes, and clients fairly similar throughout the program, or is there substantial variation across program components, providers, or subgroups of clients?

*Implementation questions* are those that tell about how and to what extent activities have been implemented as intended and whether they are targeted to appropriate populations or problems:

- To what extent are mandated or authorized activities actually being carried out?
- To what extent are the activities in accordance with the purpose of the law and implementing regulations?
- How do activities conform to the intended program model or to professional standards of practice, if applicable?
- What evidence is there that program resources are efficiently managed and expended?

*Impact questions* illuminate the program effects:

- To what extent is the program achieving its intended purposes or outcomes? What is the aggregate impact of the program? How did impact or outcomes vary across participants

and approaches? How did impact vary across providers? Specifically, what type of program support was given to providers whose performance was consistently weak?

- What other important effects relate to the program (side effects)? What unforeseen effects (either positive or negative) did the program have on the problems or clients it was designed to address? How did the program affect other programs aimed at a similar problem or population?
- How does this program compare with an alternative strategy for achieving the same ends? To what extent are the effects gained through the program worth its financial and other costs? Taking both costs and effects into account, what evidence is there that the current program is superior to alternative strategies for achieving the same goals?

If evaluation questions are prepared within the transformative framework, they exemplify an understanding of the power relationships that need to be addressed (Mertens, 2020). Examples of evaluation questions from this perspective might include the following:

- How can we teach and counsel students and clients so that they **do not continue** to be oppressed?
- To what extent are the resources equitably distributed?
- How has the institution or agency been responsive/**unresponsive** in meeting the needs of people with disabilities?

The focus of these questions is to place the **problem in an unresponsive system with power inequities** rather than in the individual **without power**.

No matter what the initial evaluation questions, the evaluator should always be sensitive to emerging issues that necessitate a revision of these questions. This might be especially important in responsive and transformative evaluations.

### Selection of an Evaluation Model

Evaluators carry an inclination, based on their worldview, to use specific models in their evaluation work. However, the needs of the evaluation will determine the appropriateness and feasibility of using a specific model. At this point, evaluators must ask themselves questions such as these: Can I conduct this evaluation using the model that seems most appropriate to me, given my view of the world? Can I modify, adjust, or adapt my way of thinking to be responsive to the needs of this client in this setting? Can I use my way of thinking to help the client think about the problem in ways that are new and different and, ideally, more constructive and productive? I suspect that older, more experienced evaluators can make choices based on compatibility with their worldviews more readily than newer, less-experienced evaluators. Perhaps newer evaluators would find it easier to adjust their model selection to the needs of the client. For example, Patton (2008) declares himself to be a pragmatist in his choice of models and methods, asserting that sensible methods decisions should be based on the purpose of the inquiry, the questions being investigated, and the resources available.

The reader should review the major models of evaluation as they are presented earlier in this chapter. If one model seems to address the needs of the evaluation more than another, consider if that is a model that aligns with your worldview. If you think that you can “enlighten” a client to a different approach to evaluation, give it a try: One of the roles that evaluators fulfill is educating the client about evaluation. Just as the implicit or explicit model that you entered the situation

with influenced your decisions up to this point, the model that you choose to operationalize influences your next steps in planning the evaluation.

### Planning the Evaluation

**Data Collection Decisions.** Data collection decisions basically answer these questions:

- What data collection strategies will you use? Who gets power to decide on the type and nature of the data collection?
- Who will collect the data? The evaluator or a team or community members?
- From whom will the data be collected (sampling)?
- When and where will the data be collected?
- How will the information be returned to you?

Your basic choices for data collection strategies are the same as those outlined in Chapters 11 and 12 (on sampling and data collection, respectively) in this text, as well as those discussed in Chapter 6 on survey research and Chapter 8 on qualitative methods. Your major concern is choosing data collection strategies that provide answers to your evaluation questions within the constraints of the evaluation study that are culturally and linguistically appropriate and that satisfy the information needs of your stakeholders. The constraints of data collection in public institutions and agencies sometimes include specific policies about who can have access to information. When the data collection plan is being developed, the evaluator should inquire into such possible constraints. Stakeholders should participate in the development of the data collection plan and should reach agreement that the data collected using this plan will satisfy their information needs. Of course, you need to stay flexible and responsive to emerging data collection needs.

**Analysis, Interpretation, and Use.** Issues related to analysis, interpretation, and use planning are directly affected by your data collection choices. These topics are discussed in Chapter 12. Nevertheless, as a part of the evaluation plan, the steps that you will take for analysis, interpretation, and use of the evaluation data should be specified. Your model of evaluation will determine how interactive and iterative the analysis and interpretation phases are. For example, in transformative, constructivist, and Indigenous evaluations, you would expect to have fairly constant interaction with the stakeholder groups. You would share preliminary results and consult on the use of the findings for the purpose of program modification as well as for directions for further data collection and analysis. If you are functioning within a postpositivist model, you would be less likely to have such a high level of interaction with your stakeholder groups. Rather, you would attempt to maintain distance so that your presence did not unduly influence the effects of the program. In terms of interpretation, evaluators have emphasized the development of a standard for use in making judgments about a program's merit or worth. For example, if a program is designed to reduce the dropout rate of high school students, is a 50% reduction considered successful? How about a 25% reduction? Such standards are appropriate for studies that focus on

impact evaluation, although process-oriented standards could be established in terms of number of clients served or number of participants at a workshop.

### Management of the Evaluation

The management plan for the evaluation should include a personnel plan as well as a budget (see Chapter 13 for more on this topic). The personnel plan specifies the tasks that will be done as well as how, when, and by whom they will be done. The cost plan specifies the costs of the evaluation in such categories as personnel, travel, supplies, and consultants. Together, these two parts of the management plan can be used as the basis of a formal contract between the evaluator and the sponsoring agency.

**Meta-Evaluation Plan.** The meta-evaluation plan specifies how the evaluation itself will be evaluated. Typically, the meta-evaluation specifies when reviews of the evaluation will be conducted, by whom, and with reference to what standards. In evaluation work, three time points often seem appropriate for meta-evaluation to occur: (a) after the preliminary planning is finished, (b) during the implementation of the evaluation, and (c) after the evaluation is completed. The meta-evaluation can be accomplished by asking a person outside of the setting to review the evaluation planning documents, the progress reports, and the final report. It can also involve feedback from the stakeholder groups. One source of standards that the evaluation can be measured against is *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users* (3rd ed.; Yarbrough et al., 2011). These standards are discussed in the next section.

## EXTENDING YOUR THINKING

### Evaluation Planning

Find a published evaluation study. On the basis of information given in the article, create an evaluation plan (in retrospect) using the guidelines for planning provided in this chapter. Your plan should include the following:

- a. Description of what is to be evaluated
- b. The purpose of the evaluation
- c. The stakeholders in the evaluation
- d. Constraints affecting the evaluation
- e. The evaluation questions
- f. Description of the evaluation model
- g. Data collection specifications
- h. Analysis, interpretation, and use strategies
- i. Management plan
- j. Meta-evaluation plans

If you are unable to locate information from the published article that would provide a complete retrospective plan, make a note of the missing elements. Add a section to your plan indicating what you think needs to be added to improve the plan. This will require some creative thinking on your part.

## STANDARDS FOR CRITICALLY EVALUATING EVALUATIONS

The *Program Evaluation Standards* (referred to earlier and hereafter referred to as the *Standards*; Yarbrough et al., 2011) were originally developed by a joint committee that was initiated by the efforts of three organizations: the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education. Representatives of these three organizations were joined by members of 12 other professional organizations (e.g., American Association of School Administrators [now known as AASA—The School Superintendents Association], Association for Assessment and Research in Counseling, and the National Education Association) to develop a set of standards that would guide the evaluation of educational and training programs, projects, and materials in a variety of settings. The *Standards* provide one comprehensive (albeit not all-encompassing) framework for examining the quality of an evaluation.

The *Standards* are organized according to five main attributes of evaluations:

- *Utility*—how useful and appropriately used an evaluation is
- *Feasibility*—the extent to which the evaluation can be implemented successfully in a specific setting
- *Propriety*—how humane, ethical, moral, proper, legal, and professional an evaluation is
- *Accuracy*—how dependable, precise, truthful, and trustworthy an evaluation is
- *Evaluation Accountability*—the extent to which the quality of the evaluation itself is assured and controlled

The main attributes are defined and guidelines and illustrative cases are included in the *Program Evaluation Standards* text itself. The illustrative cases are drawn from a variety of educational settings, including schools, universities, the medical and health care field, the military, business and industry, the government, and law. A summary of the standards is presented here:

### Utility

#### U1 Evaluator Credibility

Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.

#### U2 Attention to Stakeholders

Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.

#### U3 Negotiated Purposes

Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.

#### **U4 Explicit Values**

Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.

#### **U5 Relevant Information**

Evaluation information should serve the identified and emergent needs of stakeholders.

#### **U6 Meaningful Processes and Products**

Evaluations should construct activities, descriptions, findings, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors.

#### **U7 Timely and Appropriate Communication and Reporting**

Evaluations should attend in a continuing way to the information needs of their multiple audiences.

#### **U8 Concern for Influence and Consequences**

Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse. (Yarbrough et al., 2011, p. 3)

### **Feasibility**

#### **F1 Project Management**

Evaluations should use effective project management strategies.

#### **F2 Practical Procedures**

Evaluation procedures should be practical and responsive to the program operates.

#### **F3 Contextual Viability**

Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.

#### **F4 Resource Use**

Evaluations should use resources effectively and efficiently. (Yarbrough et al., 2011, p. 71)

### **Propriety**

#### **P1 Responsive and Inclusive Orientation**

Evaluations should include and be responsive to stakeholders and their communities.

#### **P2 Formal Agreements**

Evaluations should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other parties.

#### **P3 Human Rights and Respect**

Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.

#### **P4 Clarity and Fairness**

Evaluations should be understandable and fair in addressing stakeholder needs and purposes.

#### **P5 Transparency and Disclosure**

Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.

#### **P6 Conflicts of Interests**

Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.

#### **P7 Fiscal Responsibility**

Evaluations should account for all expended resources and comply with sound fiscal procedures and processes. (Yarbrough et al., 2011, p. 105)

### **Accuracy**

#### **A1 Trustworthy Conclusions and Decisions**

Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.

#### **A2 Valid Information**

Evaluation information should serve the intended purposes and support valid interpretations.

#### **A3 Reliable Information**

Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.

#### **A4 Explicit Program and Context Descriptions**

Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.

#### **A5 Information Management**

Evaluations should employ systematic information collection, review, verification, and storage methods.

#### **A6 Sound Designs and Analyses**

Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes.

#### **A7 Explicit Evaluation Reasoning**

Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented.

#### **A8 Communication and Reporting**

Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors. (Yarbrough et al., 2011, p. 157)

## Evaluation Accountability

### M1 Evaluation Documentation

Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.

### M2 Internal Meta-Evaluation

Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.

### M3 External Meta-Evaluation

Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external meta-evaluations using these and other applicable standards. (Yarbrough et al., 2011, p. 225)

## Multicultural Validity

Mertens (2009) and Kirkhart (2005) recognize that concerns about diversity and multiculturalism have pervasive implications for the quality of evaluation work. Recall that, in Chapter 1, you read about Kirkhart's term *multicultural validity*, which she defined as that type of validity that indicates that we have established correct understandings across multiple cultural contexts. She identified five types of validity: theoretical, experiential, consequential, interpersonal, and methodological (see Chapter 1). Three types of threats to establishing multicultural validity are listed here:

1. It takes time to reflect multicultural perspectives soundly. Many evaluations are conducted in compressed time frames and on limited budgets, thus constraining the ability of the evaluator to be sensitive to the complexity of multicultural dimensions.
2. Cultural sophistication needs to be demonstrated on cognitive, affective, and skill dimensions. The evaluator needs to be able to have positive interpersonal connections, conceptualize and facilitate culturally congruent change, and make appropriate cultural assumptions in the design and implementation of the evaluation.
3. The evaluator must avoid cultural arrogance that is reflected in premature cognitive commitments to a particular cultural understanding as well as to any given model of evaluation.

## ETHICS AND EVALUATION: THE GUIDING PRINCIPLES

---

Evaluators need to be aware of the same ethical guidelines that researchers are, such as the APA's (2016) *Ethical Principles*. Another important resource for designing high-quality evaluations is the AEA's (2018) *Guiding Principles for Evaluators* ([www.eval.org](http://www.eval.org)). There are five guiding principles:

- *Systematic Inquiry*. Evaluators conduct data-based inquiries that are thorough, methodical, and contextually relevant.
- *Competence*. Evaluators provide skilled professional services to stakeholders.

- *Integrity.* Evaluators behave with honesty and transparency in order to ensure the integrity of the evaluation.
- *Respect for People.* Evaluators honor the dignity, well-being, and self-worth of individuals and acknowledge the influence of culture within and across groups.
- *Common Good and Equity:* Evaluators strive to contribute to the common good and advancement of an equitable and just society.

The *Program Evaluation Standards* and the *Guiding Principles for Evaluators* should be used by the evaluator in developing and implementing the evaluation study. Conducting the meta-evaluation at the design stage not only ensures that a worthwhile evaluation has been constructed that is likely to produce the information needed by users but also will increase the confidence of those associated with the evaluation. Conducting the meta-evaluation across the life cycle of the evaluation will improve the evaluation.

## QUESTIONS FOR CRITICALLY ANALYZING EVALUATION

---

The following questions are designed to parallel the *Standards* and to address issues raised by the construct of multicultural validity as it was described by Kirkhart (2005).

### Utility

1. *Evaluator credibility:* How did the evaluator establish their credibility, and how did the stakeholders perceive the evaluator's credibility?
2. *Attention to stakeholders:* What was the range of stakeholders included in the evaluation and were any stakeholders excluded?
3. *Evaluation purpose:* What were the purposes of the evaluation? How were they negotiated throughout the process of the evaluation?
4. *Explicit values:* What cultural and individual values were identified, and how did these relate to the purposes, processes, and judgments of the evaluation?
5. *Relevant information:* To what extent were the information needs of stakeholders served?
6. *Meaningful processes and products:* To what extent did the evaluation activities, descriptions, and judgments yield changes in understanding and actions for participants?
7. *Timely and appropriate communication and reporting:* What was the schedule of communication and reporting? How did it serve the needs of the stakeholders?
8. *Concern for consequences and influence:* What use was made of the evaluation findings? Was there evidence of misuse?

### Feasibility

1. *Project management:* What evidence is there that the evaluation used effective project management strategies?

2. *Practical procedures:* What evidence is there that the evaluation used practical procedures that were responsive to the program operation?
3. *Contextual viability:* How did the evaluation recognize, monitor, and balance the cultural and political interests and needs of individuals and groups?
4. *Resource use:* How did the evaluation use resources equitably, efficiently, and effectively?

## Propriety

1. *Responsive and inclusive orientation:* How did the evaluation include and be responsive to stakeholders and their communities?
2. *Formal agreements:* What evidence is there that the evaluation was based on negotiated and renegotiated formal agreements, taking into account the contexts, needs, and expectations of clients and other parties?
3. *Human rights and respect:* How did the evaluations protect human and legal rights and respect the dignity and interactions of participants and other stakeholders?
4. *Clarity and fairness:* Was the evaluation understandable and fair in addressing stakeholder needs and purposes?
5. *Transparency and disclosure:* How did the evaluation make complete descriptions of findings, limitations, and conclusions available to all stakeholders, unless doing so would violate legal and propriety obligations?
6. *Conflicts of interests:* What evidence is there that the evaluators identified and addressed real or perceived conflicts of interest that might compromise processes and results?
7. *Fiscal responsibility:* How did the evaluations for all expended resources comply with sound fiscal procedures and processes?

## Accuracy

1. *Justified conclusions and decisions:* What evidence is there that the conclusions and decisions were justified in the cultures and contexts where they have consequences?
2. *Valid information:* What evidence is there that the information served the intended purposes and supported valid interpretations? How does the evaluator establish internal and external validity?
3. *Reliable information:* What evidence is there that the evaluation information is dependable and consistent for the intended purposes?
4. *Explicit program and context descriptions:* How did the evaluation document programs and their contexts with appropriate detail and scope for the evaluation purposes?
5. *Information management:* What information collection, review, verification, and storage methods were used, and were they systematic?

6. *Sound designs and analyses*: What evidence is there that the evaluation employed technically adequate designs and analyses that are appropriate for the evaluation purposes?
7. *Explicit evaluation reasoning*: How was the evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments clearly documented?
8. *Communication and reporting*: What evidence is there that the evaluation communications were adequate in scope and as free as possible from misconceptions, distortions, and errors?

### Evaluation Accountability

1. *Evaluation documentation*: What documentation is available that shows the negotiation of purposes and implementation of designs, procedures, data, and outcomes?
2. *Internal meta-evaluation*: How did evaluators use standards to examine the accountability of the evaluation design, procedures, information collected, and outcomes?
3. *External meta-evaluation*: What evidence is there that a meta-evaluation was supported and conducted using appropriate standards?

### Interpersonal Validity

1. *Personal influences*: What influences did personal characteristics or circumstances, such as social class, gender, race and ethnicity, language, disability, or sexual orientation, have in shaping interpersonal interactions, including interactions between and among evaluators, clients, program providers and consumers, and other stakeholders?
2. *Beliefs and values*: What are the influences of the beliefs and values of the evaluator and other key players in filtering the information received and shaping interpretations?

### Consequential Validity

1. *Catalyst for change*: What evidence is there that the evaluation was conceptualized as a catalyst for change (e.g., shift the power relationships among cultural groups or subgroups)?
2. *Unintended effects*: What evidence is there of sensitivity to unintended (positive or negative) effects on culturally different segments of the population?

### Multicultural Validity

1. *Time*: Were the time and budget allocated to the evaluation sufficient to allow a culturally sensitive perspective to emerge?

2. *Cultural sophistication*: Did the evaluator demonstrate cultural sophistication on the cognitive, affective, and skill dimensions? Was the evaluator able to have positive interpersonal connections, conceptualize and facilitate culturally congruent change, and make appropriate cultural assumptions in the design and implementation of the evaluation?
3. *Avoidance of arrogant complacency*: What evidence is there that the evaluator has been willing to relinquish premature cognitive commitments and to be reflexive?

## EXTENDING YOUR THINKING

### EVALUATION STANDARDS

1. Find a published evaluation study. Critically analyze the study using the questions for critical evaluation of evaluation studies listed in this chapter. Summarize the strengths and weaknesses of the study. Suggest possible changes that might strengthen the study.
2. Find a second published evaluation study that exemplifies a different evaluation model or paradigm. Critically analyze it using the questions for critical evaluation of evaluation studies listed in this chapter. Suggest possible changes that might strengthen the study.
3. Contrast the types of strengths and weaknesses that emerge from the two studies that were conducted from a different evaluation model or paradigm. How did the differences in the studies affect your use of the evaluation standards that were presented in this chapter (the *Program Evaluation Standards* [Yarbrough et al., 2011] and Kirkhart's [2005] multicultural validity standards)?
4. Because the *Standards* were developed within the context of educational evaluations, there has been some discussion about their applicability to evaluations in other settings. Review an evaluation study based in another type of program (e.g., drug abuse, clinical services). Critically comment on the usefulness of the *Standards* within this context. Use the points presented in the previous question to structure your response.
5. Brainstorm ideas for an evaluation study. Working in a small group, design an evaluation plan following the steps outlined in this chapter. Include in your plan all the components listed in Chapter 2. Critique your plan using the questions for critical analysis listed in this chapter.

## REFERENCES FOR PERSONNEL EVALUATION

### Books

- Grissom, J. A., & Youngs, P. (2016). *Improving teacher evaluation systems*. Teachers College, Columbia University.
- Joint Committee on Standards for Educational Evaluation. (2013). *The personnel evaluation standards* (2nd ed.). Corwin.
- Marzano, R. J., Rains, C. L., & Warrick, P. B. (2020). *Improving teacher development and evaluation*. Marzano Resources.

### Journals

*Counselor Education and Supervision*

*Educational Assessment, Evaluation and Accountability* (formerly the *Journal of Personnel Evaluation in Education*)

*Educational Leadership*

*Journal of Applied Psychology*

*Journal of Counseling Psychology*

*Journal of Occupational and Organizational Psychology*

*Journal of Personnel Psychology*

*Professional School Counseling*

## EVALUATION RESOURCES FOR PROGRAM AND POLICY EVALUATIONS

### Books and Monographs

Alkin, M., & Christie, C. A. (Ed.). (2023). *Evaluation roots* (3rd ed.). Guilford Press.

Brisolera, S., Seigart, D., & SenGupta, S. (Eds.). (2014). *Feminist evaluation and research*. Guilford Press.

### Online Resources

- The website of the American Evaluation Association ([www.eval.org](http://www.eval.org)) includes many links to evaluation resources that are available online.
- The *Step-by-Step Guide to Evaluation* (W. K. Kellogg Foundation, 2017) outlines a blueprint for designing and conducting evaluations, either independently or with the support of an external evaluator/consultant; the W. K. Kellogg Foundation (2004) also publishes the *Logic Model Development Guide*, available on its website (<https://wkkf.isu.edu/elab.org>).
- The National Science Foundation ([www.nsf.gov](http://www.nsf.gov)) offers many resources that discuss designing and conducting evaluations that integrate quantitative and qualitative techniques and address issues of including members of underrepresented groups on the basis of race/ethnicity, gender, and disability.
- William M. Trochim's (n.d.) *Research Methods Knowledge Base* is available at <https://conjointly.com/kb/>. Trochim developed an online course in research and evaluation methods that is accessible at this website; it contains many links to other resources relevant for evaluators.
- The U.S. Government Accountability Office ([www.gao.gov](http://www.gao.gov)) has many online resources that discuss evaluation synthesis, designing evaluations, case study evaluation, and prospective evaluation methods.

- The Evaluation Center at Western Michigan University (<https://wmich.edu/evaluation>) maintains a website devoted to evaluation resources. One of the resources is a database of checklists for various approaches to evaluation, such as utilization-focused evaluation (<https://wmich.edu/evaluation/checklists>).
- Free Resources for Program Evaluation and Social Research Methods (<https://gsociology.icaap.org/methods/index.html>) offers links to free resources. The focus is on “how to” do program evaluation and social research: surveys, focus groups, sampling, interviews, and other methods. There are also links to sites with information about how to do statistics, how to present data, and links to sites with subject areas evaluators and social researchers need to know, such as organizations and social psychology. This site also links to free online training classes and videos about research methods and evaluation. Finally, this site has the basic guides to evaluation.
- [Transformativeresearchandevaluation.com](http://transformativeresearchandevaluation.com) is a website that I maintain and contains links to publications, videos, and other resources for those interested in transformative research and evaluation.

## Evaluation Journals

*American Journal of Evaluation*

*Educational Evaluation and Policy Analysis*

*Evaluation and the Health Professions*

*Evaluation and Program Planning*

*Evaluation Review*

*Journal of Mixed Methods Research*

*Journal of MultiDisciplinary Evaluation*

*New Directions for Evaluation*

*Studies in Educational Evaluation*

## Professional Associations

The American Evaluation Association (AEA) is the primary professional organization for practicing evaluators in the United States and has a strong representation of evaluators from other countries. Other countries and nations also have professional organizations, such as the Canadian Evaluation Society, the African Evaluation Association (AfrEA), and the Australian Evaluation Society. The AEA publishes two journals, the *American Journal of Evaluation* and *New Directions for Evaluation*.

The International Organization for Cooperation in Evaluation (IOCE) was established in 2003 through a worldwide collaboration of evaluation organization representatives (Mertens, 2005). The IOCE maintains a website that provides links to many evaluation organizations around the world ([www.ioce.net](http://www.ioce.net)). It now lists over 100 different international, regional, and national evaluation organizations, up from 5 organizations in 1995.

## SUMMARY OF CHAPTER 2: EVALUATION

Evaluation and research share many things; however, the differences between the two genres of systematic inquiry should be apparent to you after studying this chapter. Evaluation has its own jargon, including the evaluand (what is evaluated), stakeholders (people who have a stake in the outcomes of the evaluation), and other terms reviewed in this chapter. It also has a history of development from the early 1960s to the present; it is a dynamic field with new developments occurring in response to challenges and political factors. The five major paradigms can also be used to frame evaluation studies, just as they can be for research. The planning process of an evaluation is premised on ongoing involvement of stakeholder groups. Professional associations and standards for quality in evaluation are useful resources for anyone interested in pursuing an evaluation study.

## IN THE NEXT CHAPTER

The process of focusing on a specific topic for your research is explored by means of literature review and community engagement.

Do not copy, post, or distribute