

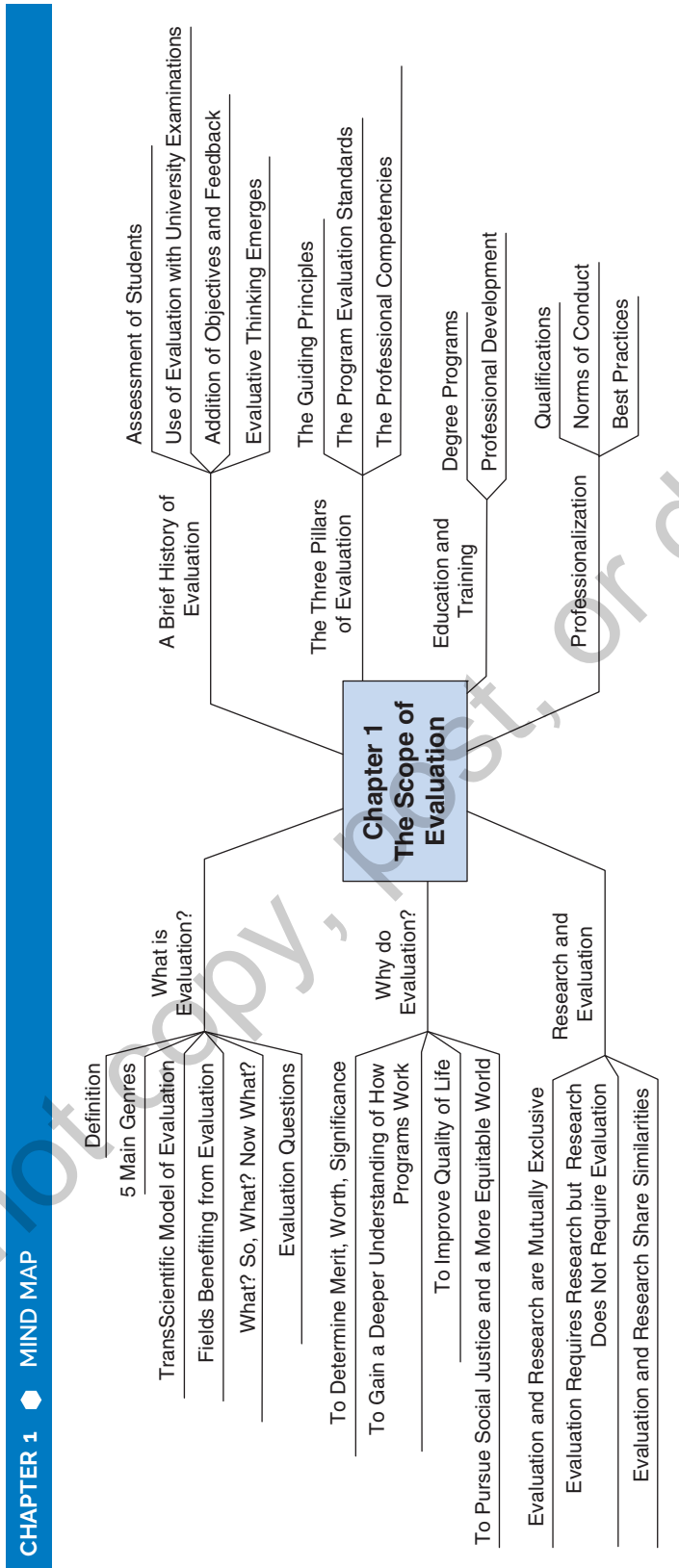


# THE SCOPE OF EVALUATION

## LEARNING OBJECTIVES

1. Define program evaluation.
2. List three basic evaluation questions.
3. Identify four main reasons for doing the evaluation.
4. Distinguish between research and evaluation.
5. Recognize the breadth of the evaluation discipline today.
6. Explain how an evaluator can use the three pillars of evaluation.

Do not copy, post, or distribute



## INTRODUCTION

Programs and policies are designed to make positive changes in the world. They affect the lives of people as well as the physical and social environments in which they live. By extension, the evaluations of these programs and policies have intrinsic power, because they serve as a bridge between the organizations that sponsor them and the individuals most affected by the programs and policies in place (McBride et al., 2020). Further, evaluators have extrinsic influence because decision makers and leaders may make significant changes based on the evaluation findings, for example, to expand, revise, replicate, reduce, or terminate the program or policy. It's a big responsibility, and it is one that evaluators willingly accept every day. This chapter will provide you with an overview of evaluation, what it is, and why it is important. The rest of the book will help you learn how to do it.

You may be surprised to learn that you are already an evaluator! You evaluate what car to buy, what school to send your kid to, what to have for dinner. You certainly evaluated which program to enroll in to meet your current educational goals. As a result, because you make decisions every day, you may think that evaluation as a formal decision-making process is unnecessary. It turns out that the personal methods you use to make your decisions are very similar to the ones that organizations use, ranging from thoughtful, evidence-based decision-making processes to the old dartboard approach. Which makes you feel more confident? Which is more likely to produce an effective outcome?

Program evaluation offers a toolbox full of tried-and-true methods and strategies to help programs make better decisions and determine where they need to make improvements. We use the term “program evaluation” throughout the book as an umbrella term, but the processes we describe are just as useful to evaluate projects, services, activities, policies, interventions, and products. Evaluation goes beyond improving how things are done to determining if the program's objectives have been achieved. It can also look at longer-term outcomes to determine impact.

Examples of typical social interventions an evaluator might encounter include an employee mentoring program, a medication compliance intervention, a wellness program for employees, a counseling program for new parents, a no-smoking ordinance, a microloans agency in a developing country, a new graduate program in environmental science, a housing program for homeless youth, or an intervention to reduce falls in the elderly. Other types of evaluations focus on program monitoring, fidelity to an established program model, the potential for scaling an innovation, tracking research dissemination and outcomes, and exploring lessons learned. Finally, evaluation can look at the findings of other evaluations to determine overall quality and best practice.

We need to know that the program has solved a **problem**. Funders want to know if their money has been well spent by supporting this initiative, agencies want to know if they should continue to pursue their current direction or change course, staffers want to know if their work has helped program participants make improvements in their lives.

This chapter provides an overview of program evaluation, what it is and why it is important. It also looks at the similarities and differences between research and evaluation. A brief history of evaluation is provided, looking across its development over the last

70 years to become what is now a worldwide discipline. Because evaluation is grounded in the unpredictable context of people, events, and politics, challenges and ethical dilemmas are commonplace and so the chapter reviews the three pillars of ethical conduct that support evaluators in their work. Last of all, an overview is provided of some of the professional education and training opportunities available to evaluators and those interested in pursuing an evaluation career.

You will also enjoy the resources in this chapter including the Spotlight on Equity, which presents the Graduate Education Diversity (GED) Program, our expert Dr. Jean King, Key Terms, Main Ideas, Critical Thinking Questions, Student Challenges, and Additional Readings and Resources.

## WHAT IS EVALUATION?

Evaluation is an applied inquiry process used to collect and synthesize information that can then be used to draw conclusions about the situation, significance, or quality of a program or other entity under review. Unlike other types of research, evaluation conclusions encompass not only the evidence obtained but also assign a value or judgment to those findings (Fournier, 2005). It combines (Donaldson & Christie, 2006):

*systematic inquiry and analysis techniques with an eye toward answering important and fundamental questions about programs, policies, and interventions such as: does it work, why does it work, for whom does it work best, and how do we make it work better? (p. 249)*

Many definitions of **program evaluation** exist, but the one we use most often is by Rossi et al. (2004). It states:

*Program evaluation is the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions. (p. 16)*

This definition addresses how evaluation borrows methods from social science research to conduct studies. It describes evaluation as systematic, meaning that it is not just an informal activity and that there are theories, frameworks, methods, and processes which have been explored and improved by many evaluators over time. The definition mentions the setting of the program and, as we will see, context is critical to any evaluation. Finally, it provides a call to action as evaluators seek to find ways to foster social justice in our communities and around the world.

However, not all evaluations focus on social justice. Greene sorted evaluation approaches into five main genres, based on the interests they serve, the values they advance, and the needs of the client (Tarsilla, 2010, p. 211):

1. The efficiency interests of policymakers
2. The accountability and ameliorative interests of on-site program managers

3. Learning, understanding and use
4. Understanding and development interests of direct service staff and affiliates
5. Democratic and social change interests of program beneficiaries and their allies.

In 1991, Scriven described evaluation as a transdiscipline (1991a). He suggests this because it is both extraordinarily multidisciplinary and multirole, as the evaluator typically performs tasks often found in other professions. These include research, instruction, therapy, public relations, administration, entrepreneurship, management, as well as the roles of “arbitrator, scapegoat, trouble-shooter, inventor, conscience, jury, judge, or attorney” (pp. 363–364). In addition, the service role is critical because evaluators have clients, not just readers. He concludes, “evaluation has a nature, a flavor, a gestalt of its own. It is idiosyncratic and complex to the extent that it requires a special kind of paradigm.” At an expert lecture at the American Evaluation Conference in 2021, he presented a paper calling for the **TransScientific Model of Evaluation** (Scriven, 2021), where he concluded that “Evaluation is the driving logic of science and of the disciplines outside science (e.g., law, drama, medical practice, journalism, education).”



Source: Chris Lysy/FreshSpectrum.com

Many fields benefit from having evaluation as a professional competency, even if it is not their main focus. Examples of fields benefitting from evaluation are listed in [Table 1.1](#).

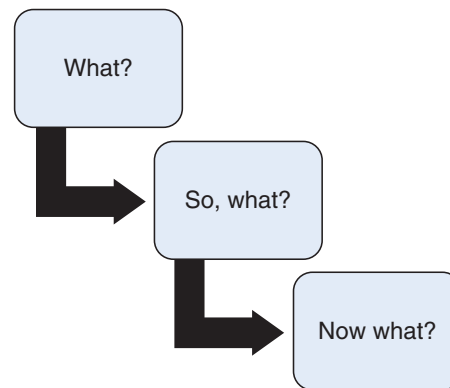
**TABLE 1.1** ■ EXAMPLES OF FIELDS BENEFITTING FROM EVALUATION

<ul style="list-style-type: none"> <li>• Child and Family Studies</li> <li>• Counseling</li> <li>• Community Development</li> <li>• Criminal Justice</li> <li>• Education</li> <li>• Emergency Management</li> <li>• Foundations</li> <li>• Government</li> <li>• Health Sciences</li> </ul>	<ul style="list-style-type: none"> <li>• Human Resource Management</li> <li>• Human Services</li> <li>• Nonprofit Organizations</li> <li>• Policy Analysis</li> <li>• Psychology</li> <li>• Public Administration</li> <li>• Public Health</li> <li>• Social Work</li> </ul>
--	--

Michael Patton (2008, p. 5) points out that in the simplest terms, evaluations answer three basic questions ([Figure 1.1](#), p. 8): What? So what? Now what?

This list of questions provides a way to understand problems and discover new solutions. “What?” asks for the story of the program so that it can be understood, and any issues and problems identified. “So what?” tries to make sense of the facts, examines assumptions

FIGURE 1.1 WHAT? SO, WHAT? NOW WHAT?



Source: Patton, M. Q. (2008). *Utilization-focused evaluation*. SAGE Publications.

made, looks at who is affected, and explores implications. “Now what?” identifies a course of action and identifies what is needed to make informed decisions (McDowell, 2017; Razzetti, 2019; Tenney & Pew, 2006).

Let’s test these basic questions using data literacy as an example.

### What?

Data literacy is “understanding what data mean, including how to read charts appropriately, draw correct conclusions from data and recognize when data are being used in misleading or inappropriate ways” (Carlson et al., 2011, p. 5). When someone is data literate, they are confident about their use of data and apply their skills. They need to know where to retrieve data, assess its quality, and interpret it to illuminate the problem they are studying.

### So what?

If someone is not data literate, we can assume that they do not know what data to use. They may lack computer skills, or an awareness of what information is available, or be unable to interpret its meaning, or they may experience a fear of numbers in general. If computer literacy is determined to be a barrier, what can be done?

### Now what?

A solution could focus first on reducing the amount of anxiety experienced by the individual. Then their computer skills could be assessed and upgraded if needed. They could access online training or work with a trainer or mentor to learn how to use tools and interpret data. Then they could learn how to apply their new skills using mini-cases or scenarios.

Typical evaluation questions might look like those listed in [Box 1.1](#) (Rossi et al., 2004, p. 3):

### BOX 1.1 TYPICAL EVALUATION QUESTIONS

1. What is the nature and scope of the problem? Where is it located, whom does it affect, how many are affected, and how does it affect them?
2. What is it about the problem or its effects that justifies new, expanded, or modified social programs?
3. What feasible interventions are likely to significantly ameliorate the problem?
4. What are the appropriate intended populations for interventions?
5. Is an intervention reaching its intended population?
6. Is the intervention being implemented well? Are the intended services being provided?
7. Is the intervention effective in attaining the desired goals or benefits?
8. Is the program cost reasonable in relation to its effectiveness and benefits

*Source:* Rossi, P., Lipsey, M., & Freeman, H. (2004). *Evaluation: A systematic approach*. SAGE Publications.

These critical questions, coupled with the extensive challenges of doing good research in real-world settings, make program evaluation one of the most fascinating fields in social science research. If you like people, love research, are comfortable with constant change, and always need a challenge, then this career is for you.

## WHY DO EVALUATION?

Interventions and their subsequent evaluations have the potential to improve lives by telling us what works, what doesn't, and what to do next. For example, a social service agency, government department, school district, public health unit, nonprofit organization, foundation, or other group may need to establish the status of a program to make decisions about its future. They may want to know more about what problems affect their community so that a new intervention can be developed, or more about how a program is working now and how it can be improved, or what difference a current program has made in the community and whether it should be continued. Usually, organizations also want to know what to do next, so most evaluations include recommendations, based on the evidence obtained and feedback provided by stakeholders.

When talking about evaluation, the term **evaluand** is frequently used. It is a term coined by Michael Scriven (2003) to apply to the object of any evaluation or “whatever is being evaluated” (p. 139). For example, a program, service, policy, or product could be an evaluand.

There are four main reasons to conduct an evaluation: to determine merit, worth, and significance; to gain an understanding of how programs work and the difference they can

make to stakeholders; to improve quality of life; and to pursue social justice and a more equitable world.

### To Determine Merit, Worth, and Significance

The terms merit, worth, and significance are used to attach value to something. According to Greene (1997), evaluators' work, including their findings, is heavily influenced by values, both their client's and their own. Evaluators need to understand what is meant by "good quality."

**Merit** relates to the intrinsic, context-free attributes and properties of an evaluand. If you are buying an apple, you might prefer one that is sweet, fresh, and symmetrical. Those are intrinsic characteristics of a particular apple. However, if you are planning to make a pie, you may want apples that are tart, not sweet. Or if you are making applesauce, you may not care if the apple is nicely formed. What we value as desirable attributes, therefore, influences the merit of the apple you seek (Lam, 2013).

**Worth** has extrinsic meaning and so it relates to the value attached to an evaluand in a specific context. Worth requires a thorough understanding of that specific situation. A good apple might be worth \$2.00 but one with a large bruise on it might only be worth \$0.50. Worth usually has a value attached to it that is equivalent to something else (in this case, money). Thus, it is generally possible to measure worth, whereas merit is more difficult to pin down because it is intrinsic.

**Significance** depends on the values and meanings ascribed to an evaluand by specific individuals and may have great symbolic value. If the apple comes from a tree that you grew from seed, it has great significance for you and will be treasured and admired. However, the neighbor next door may consider an apple from your tree garbage if it drops onto their carefully manicured lawn.

While the concepts merit, worth, and significance are critical values in evaluation, they are seldom measured directly. Harkreader and Henry (2000) studied a performance measurement system to assess the merit and worth of school reforms made by educational leaders. They found it to be a painstaking endeavor, working their way back and forth "between developing expectations based on the reform, to checking those expectations with data, then unpacking our assumptions and going back to data again" (p. 167). While it was important to investigate plausible, rival explanations before making judgments about merit and worth, in the end, they found it was impossible to link performance measures directly to them as there were too many intervening, unmeasured variables.

Martens (2018) also explored the concepts of merit, worth, and significance in the use of rubrics in evaluation studies, finding that only 20 articles out of 239 mentioned these terms. The authors of these articles "described criteria of *merit* (quality) as opposed to worth (e.g., cost-effectiveness or some sort of value for the money) or significance (importance)" (p. 34) but only about half of them used the terms to reach evaluative conclusions (p. 40). To Martens, it appears that making decisions about merit, worth, or significance is not a simple process. Sironnik and Oakes (1990) do not view evaluation as simply a way to implement research methods. They depict evaluation "as an ongoing, collaborative, value-driven project of organizational change and improvement" (p. 54).



Scriven (1994) believes that the results are the most frequently desired outcome of an evaluation; however, the findings are the catapult to determine the merit, worth, and significance of the program.

### To Gain an Understanding of How Programs Work and the Difference They Can Make to Stakeholders

Evaluations can focus on the activities and events that occur as a program is being delivered (Mathison, 2005, p. 327). They can determine what results are being produced by the operations and if the resources provided to implement the program support the functions the program performs. Evaluation methods dig deeper into how the intervention works and why it works.

Program **stakeholders** are people who have a vested interest in the evaluand and so also in its evaluation (Greene, 2006, p. 397). Typically, they are clustered into four groups:

1. People with decision making authority over the program, such as funders and board members
2. People with direct responsibility for running the program, such as planners, administrators, and staff
3. People who are the intended beneficiaries of the program, as well as their families and communities
4. People disadvantaged by the program, such as those who did not obtain funding for their program because this one was funded instead, or those who could not attend for some reason, or were not selected.

For example, youth from kindergarten through Grade 12 across the United States often participate in out-of-school time (OST) programs, both after school and during the summer months (McCombs et al., 2017). OST programs can offer a variety of options (e.g., after-school clubs, YMCA, Boys & Girls Club), can be academically oriented (e.g., homework support), or can relate to special interests (e.g., sports clubs, theater programs). These large-scale programs are typically funded through a variety of mechanisms, including both public support such as federal, state, and local grants, and private support from tuition and donations. Many stakeholders are involved from policymakers to parents.

When McCombs et al. (2017) conducted a meta-analysis of the many large-scale, rigorous, experimental and quasi-experimental evaluations that had studied these programs, they found that programs were often grouped together without regard for differences among program goals, activities, or the quality of the content provided. They were often judged by a common metric such as achievement test scores. Connecting with stakeholders can help evaluators understand how programs work so they can design effective studies and interpret study results appropriately. These authors recommend that researchers should measure outcomes that align with specific program content and should also explore other factors deemed important to youth development so that policymakers can consider a broader range of outcomes.

## To Improve Quality of Life

**Quality of life** or QOL is a multifaceted concept that is used in many fields that look at outcomes. Wenger et al. (1984) define QOL as “an individual’s perceptions of his or her functioning and well-being in different domains of life” (p. 908). It has been adopted by clinicians, researchers, economists, and managers (Carr et al., 1996) as well as by evaluators. It was coined in the United States in the post-war period to describe the effect of material affluence (such as cars, houses, and consumer goods) but has been broadened to also encompass education, physical and mental health, the environment, recreation and leisure time, social belonging, religious beliefs, safety, security, and freedom.

QOL can be thought of as the sum of a range of objectively measured life conditions which can be determined numerically and then compared to a larger population, or as a subjective reaction to life conditions (life satisfaction), measuring the significance an individual places on each domain or subscale (Celestine, 2021).

Many instruments have been designed to measure QOL, for example population surveys and screening tools. Evaluators need to be aware of its limitations, and select a scale suited to the goal of the research rather than choosing one that is popular in the literature. Its use as a generic measure, however, does eliminate more useful information about the actual experiences of the individual. There is good information available on how to select an appropriate QOL questionnaire. For a comprehensive guide on selecting the best QOL questionnaire for your needs, check the Additional Readings and Resources Section for Hyland’s (2003) *Brief Guide*.

Purcell et al. (2021) examined the impact of a pilot Whole Health Coaching program for the Veterans Health Administration to determine how the program helped veterans improve their health and quality of life. The program engaged veterans across multiple dimensions of wellbeing and provided coaching to help them develop and implement personal health plans. Using a mixed methods approach, they combined pre-and post-coaching surveys with follow-up qualitative interviews and found that, “although self-reported health goals varied widely, veterans were largely satisfied with their progress toward their goals, often describing that progress as incremental and ongoing after coaching” (p. 9).

## To Pursue Social Justice and a More Equitable World

The idea of **social justice** is underpinned by two concepts: **equality** and **equity**. Equality means that all people are entitled to the same rights, freedoms, and opportunities to make the most of their lives and talents. It means that everyone has the same amount of benefit regardless of their existing needs or assets. Equity is the fair treatment of every individual, with access to opportunity, networks, resources, and supports so that they get what they need to survive and thrive. Braverman et al. (2011) define equity in relation to health in the following quotation, but it is just as applicable to any human service:

*...is the value underlying a commitment to reduce and ultimately eliminate health disparities. Health equity means social justice with respect to health and reflects the ethical and human rights concerns.... Health equity means striving to equalize opportunities to be healthy. In accord with the other ethical principles of beneficence (doing good) and non malfesance (doing no harm), equity requires concerted effort to achieve more rapid improvements among those who were worse off to start, within an overall strategy to improve everyone's health. Closing health gaps by worsening advantaged groups' health is not a way to achieve equity. Reductions in health disparities (by improving the health of the socially disadvantaged) are the metric by which progress toward health equity is measured. (p. S151)*

An infographic from the Robert Wood Johnson Foundation (2017) provides a visual comparison of equality and equity in a health context (Figure 1.2).

**FIGURE 1.2** VISUALIZING HEALTH EQUITY: ONE SIZE DOES NOT FIT ALL



*Source:* Copyright 2017. Robert Wood Johnson Foundation. Used with permission from the Robert Wood Johnson Foundation.

The infographic depicts equality by providing all the cyclists with bikes of the same size; however, for a variety of reasons related to their personal situations, they may not be able to use their bike effectively. In the second image, equity is depicted as each person accessing and using the specific type of bike which best meets their needs.

This then brings us to the idea of fairness in a social context. Social justice refers to the just distribution of wealth, opportunity, and privilege in society. In evaluation terms, it means assessing “whether the distribution of benefits and burdens among members (or groups) of a society are appropriate, fair, and moral” (House, 2005, p. 393). Social justice is directly linked to evaluation because programs and policies directly affect the distribution of

benefits and burdens. Despite this conceptual link, House maintains that social justice concerns are often omitted from evaluation discussions, either because evaluators are not well enough versed in the concepts or because adherence can be politically risky.

The dominance of values-free social research in the twentieth century meant that a utilitarian frame prevailed and that overall benefits should be increased as much as possible so that everyone could have more, but how those benefits were distributed was not a major issue (House, 2005, p. 394). However, views on the concept have shifted over time. Eventually, the way that social benefits were distributed became important for evaluators and led to the inclusion of multiple methods and multiple stakeholders in evaluation studies. Even more recently, diverse identities have begun to be considered, giving all stakeholders an effective voice in defining their own needs and negotiating their own benefits (House, p. 395). Several evaluation approaches now give stakeholders roles to play in the evaluation itself, which is a more democratic and participatory approach, but evaluators still differ on what level of involvement is necessary. Social justice continues to be a controversial topic in evaluation although there is growing attention being paid to breaking barriers, creating safety nets, and ensuring economic justice.

In the evaluation of a university-sponsored parent education program (Cooper & Christie, 2005), rather than designing an approach that would be useful to all identified stakeholder groups, the authors shifted their focus to the least powerful stakeholder group, namely the low-income Hispanic parent participants (p. 2249) to emphasize the gap between parents' and educators' notions of empowerment. They cited the admonishment by House (1991) that, "evaluation be socially just as well as true, that it attend to the interests of everyone in society and not solely the privileged" (p. 244). Using a qualitative case study approach, they determined that parents' views were being unwittingly or inappropriately overlooked. They found that educators and administrators needed to share power, validate parents' perspectives, and show sensitivity to culturally relevant values that influenced parents' educational priorities.

## RESEARCH AND EVALUATION

How is evaluation different from **research**? This perennial question generates a lot of debate about the differences and similarities of these terms. Research is referred to as the "systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge" (Steneck, 2007, p. 39).

There are three ways to view this issue.

### Evaluation and Research Are Mutually Exclusive

The first perspective is that research produces generalizable knowledge, is more theoretical, and is more firmly controlled by the researchers; evaluation produces specific, applied knowledge, and is more controlled by those funding or commissioning the evaluation (Rogers, 2014c). The distinction is described by Patton (2017) in [Table 1.2](#), p. 15.

**TABLE 1.2** ■ DIFFERENCES BETWEEN RESEARCH AND EVALUATION

Research	Evaluation
<ul style="list-style-type: none"> <li>• Purpose is testing theory and producing generalizable findings.</li> <li>• Questions originate with scholars in a discipline.</li> <li>• Quality and importance judged by peer review in a discipline.</li> <li>• Ultimate test of value is contribution to knowledge.</li> </ul>	<ul style="list-style-type: none"> <li>• Purpose is to determine the effectiveness of a specific program or model.</li> <li>• Questions originate with stakeholders and primary intended users of evaluation findings.</li> <li>• Quality and importance judged by those who will use the findings to take action and make decisions.</li> <li>• Ultimate test of value is usefulness to improve effectiveness.</li> </ul>

Source: Patton, M. Q. (2017). Evaluation flash cards: Embedding evaluative thinking in organizational culture. *Otto Bremer Foundation*. [ottobremer.org](http://ottobremer.org)

### Evaluation Requires Research but Research Does Not Require Evaluation

The second stance suggests that research is empirical, involves factual description, and generally does not include a judgment about quality. Evaluation is also empirical, but its purpose is to determine the merit, worth, or significance of something—essentially to make a value judgment (Wanzer, 2019). However, evaluation is about more than just conducting the research, its purpose includes learning and capacity building, informing decision making, and improving programming.

Wanzer studied how evaluators and researchers defined program evaluation and differentiated evaluation from research. “Evaluators were more likely to think research and evaluation intersect whereas researchers were more likely to think evaluation is a sub-component of research.” According to Wanzer (2020), evaluators saw more differences than researchers in how a study was initiated (purpose, questions, audience) and how it ended (rendering value judgments, disseminating results).

Evaluators rely on good research every time they conduct a literature review so they can base their evaluation design on a foundation of well-supported evidence. When the knowledge is not there, evaluation becomes more developmental and exploratory because the research foundation is weak (Patton, 2017, p. 7). Bloom (2010) states that “evaluation, in contrast to research, is very much influenced by its participants, which includes on-the-dime changes of direction when local evidence supports such change” (p. 2).

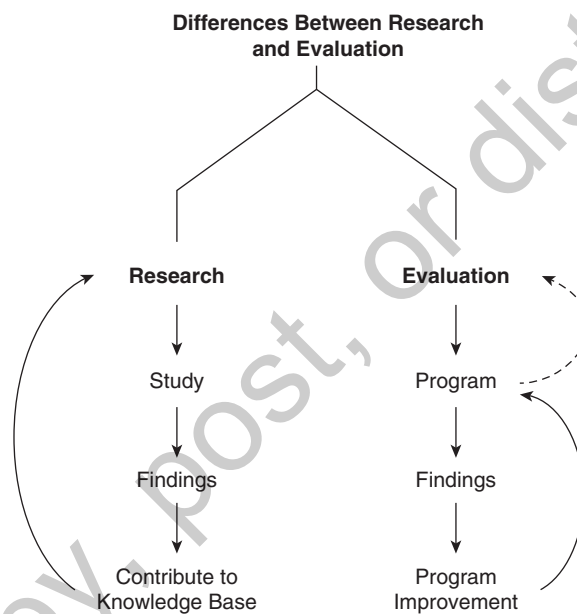
Chapel (2012) humorously illustrates how evaluation and research are different:

*Researchers must stand back and wait for the experiment to play out. To use the analogy of cultivating tomato plants, researchers ask, “How many tomatoes did we grow?” Evaluation, on the other hand, is a process unfolding in real time. In addition to determining numbers of tomatoes, evaluators also inquire about related areas like,*

*“how much watering and weeding is taking place?” “Are there nematodes on the plants?” If evaluators realize that activities are insufficient, staff are free to adjust accordingly. (p. 1)*

In [Figure 1.3](#), we see that research contributes to a knowledge base and evaluation contributes to program improvement. The dotted line returning from Program to Evaluation suggests that evidence of best practice needs to inform the field.

**FIGURE 1.3** THE DIFFERENCES BETWEEN RESEARCH AND EVALUATION



### Evaluation and Research Share Similarities

A third view is that the disciplines of research and evaluation share the same overall goal of making life better for people and use many of the same tools, strategies, and methods. Both are based on answering a question, but the evaluation tends to focus on the intervention, processes, impacts, and outcomes. Both start with a design and plan (a research plan or an evaluation plan), both work with data (e.g., qualitative, quantitative, or a combination of the two), both employ approved analysis techniques (e.g., statistics, content analysis), and the results of both types of studies can be submitted for publication in academic journals.

Which perspective resonates with you?

Next we introduce you to our Spotlight on Equity, a feature in every chapter. This first Spotlight is about the Graduate Education Diversity (GED) Program, a program of the American Evaluation Association.

## SPOTLIGHT ON EQUITY

### The Graduate Education Diversity (GEDI) Program: A Pipeline of Emerging Evaluators of Color and Those Underrepresented and Underserved in the Field of Evaluation



*Source:* American Evaluation Association/eval.org

The Graduate Education Diversity (GEDI) Program grew out of the recommendations of the American Evaluation Association (AEA) Building Diversity Initiative (BDI) (1999–2001). The initiative brought together the vision of several diverse groups of evaluators and reflected the views of those AEA members who were becoming aware of the importance of culturally responsive evaluation (CRE) and the need to address diversity issues. The mission was clear: “to develop a cadre of evaluators that were responsive to the needs of underrepresented and underserved groups of color in the United States” (Symonette et al., 2014, p. 12). While first conceived as a fellowship, after discussions with the AEA Board, the concept was expanded to become an internship program that recruited students of color.

Sponsored by the AEA and the W. K. Kellogg Foundation, the first GEDI cohort was welcomed at Duquesne University in 2004. The program’s current host site is Community Science.

Program goals include (<https://www.eval.org/gedi>) the following:

- a. Recruit graduate students of color from diverse fields to extend their research capacities in evaluation.
- b. Stimulate evaluation thinking concerning diverse communities and persons of color.
- c. Deepen the evaluation profession’s capacity to work in racially, ethnically, and culturally diverse settings.

Program participants are graduate students of color whose academic focus was not evaluation but who could see a career fit with the internship goals. They are mainly comprised of African American, Hispanic, Native American, and Asian American predoctoral students (Collins et al., 2014, p. 31) and come from a variety of disciplines

including public health, education, political science, anthropology, psychology, sociology, social work, and the natural sciences. They share a strong background in research skills, an interest in extending their skills in the field of evaluation, and a commitment to thinking deeply about culturally responsive evaluation practice (<https://www.eval.org/Education-Programs/Graduate-Education-Diversity-Internship/Application>).

The interns work approximately two days per week at an internship site for a ten-month period from September to June, gaining real-world experience to build their skills and confidence. The GEDI program also provides them with additional support through four multiday intensive trainings, monthly webinars, paid attendance at the AEA Annual Conference and the AEA Summer Evaluation Institute, mentoring, and many networking opportunities. Interns receive a stipend, and their travel expenses are reimbursed. GEDI curriculum and pedagogy is informed by a multidimensional theoretical framework across individual, organizational, community, and professional outcomes (Collins et al., 2014). Each intensive multiday training is associated with specific overarching themes, learning topics, associated readings, and reflective writing assignments. Topics are identified through a needs assessment conducted at the start of the program. Pedagogical practices include one-on-one coaching on culturally responsive and equitable evaluation theory and practice across academic institutions and internship sites. The program prioritizes equity and social justice and emphasizes peer relations and the construction of a familial environment.

At the 10-year mark, Stafford Hood reflected on the development of the program (Hood, 2014, p. 118). He saw the need to increase the number of culturally responsive evaluators worldwide to improve the lot of those who are traditionally disenfranchised. In his view, the GEDI Program has the potential to serve as a model for other programs that wish to target diversity and culturally responsive evaluation.

Now, in its eighteenth year (2021–2022), the program boasts over 130 alumni of the program. Most of them work in major philanthropic, governmental, educational, and nonprofit organizations in the United States. Their hope is to inspire and catalyze a deeper and fuller understanding of the intersection of culture with equity and justice.

The program has empowered students to become adaptive leaders. A case study on leadership, conducted by the sixth cohort, captured the perspectives of 32 former

participants as well as from seven members of their own cohort (Aponte-Soto et al., 2013, p. 40). Technical evaluation training was central to their learning, but leadership skills were also prominent. Due to their interactions with influential evaluation mentors and thought leaders in CRE, they have gained the confidence they need to support vulnerable populations. As one GEDI graduate commented (Aponte-Soto et al., 2014):

*After participating in GEDI, I have been sought after as an evaluator by organizations in need of evaluation at universities and foundations. In sum, not only do I feel more credible as an evaluator after GEDI, but others also recognize this increased credibility and potential to be a leader in evaluation. (p. 41)*

This Spotlight on Equity is based on the contributions of Dr. Sául Isaac Maldonado, Anisha Lewis, Dr. Rodney Hopson, Zachary Grays, and Dr. Brandi Gilbert.

## A BRIEF HISTORY OF EVALUATION

To get a sense of the field of evaluation, it is useful to look at how the profession started, how it has continued to grow, and the forces that have influenced it. Understanding changing politics and values over time can provide a useful backdrop for shifts in evaluative thinking and studying the past can provide inspiration and direction for the future and the work that remains. [Figure 1.4](#), p. 19 provides a timeline for the development of the evaluation profession from 1792 to the beginning of the twenty-first century based on the work by several theorists (American Evaluation Association, 2018b; Hogan, 2007; Hoskins, 1968; Mathison, 2005; Reiser, 2001; Stufflebeam et al., 2000; Worthen et al., 1997).

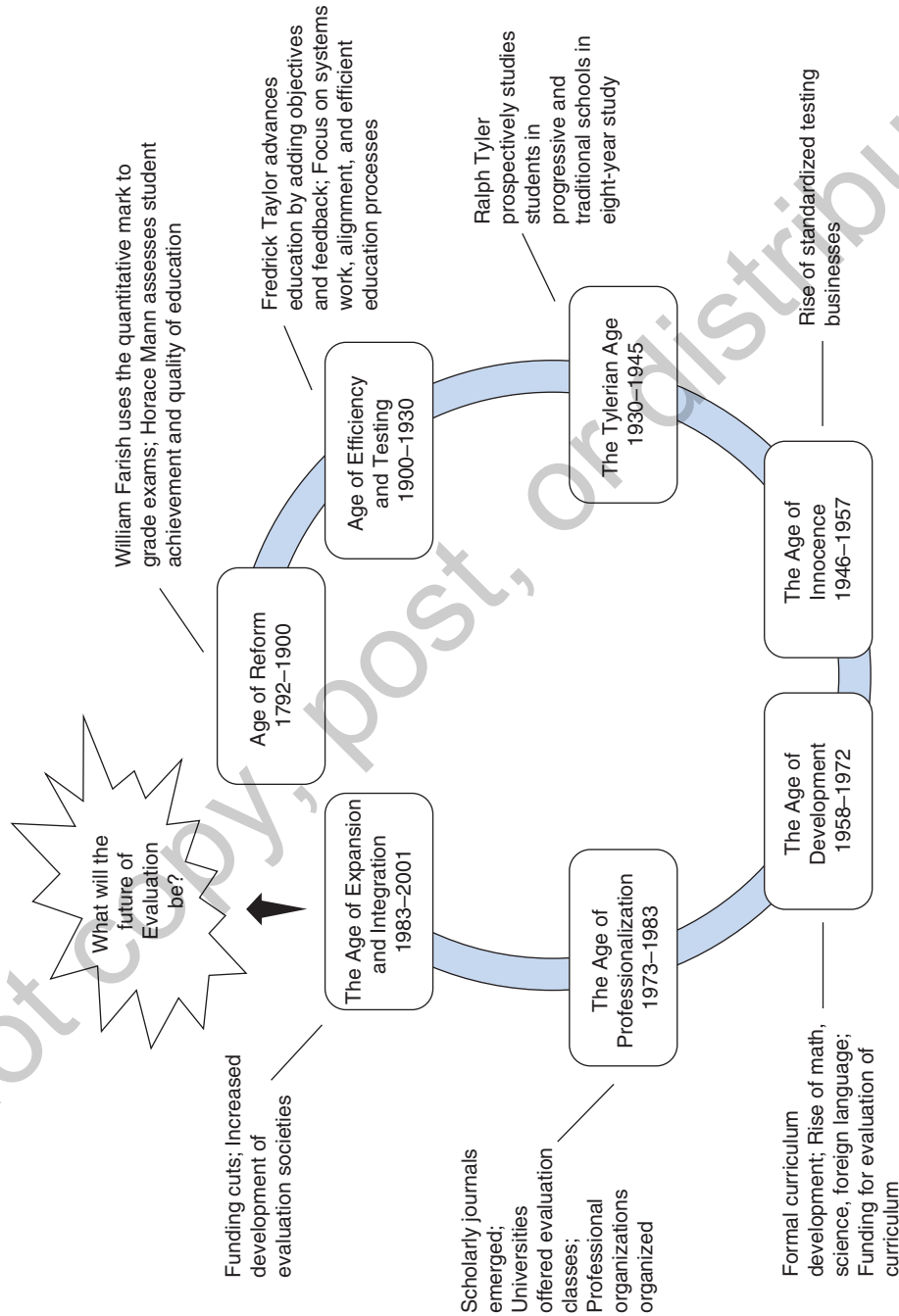
The early roots of evaluation in the United States began in the eighteenth century in 1792, when William Farish invented the first written university examination, using quantitative grades to rank students on their grasp of the practical applications of chemistry (Hoskins, 1968). Horace Mann, the father of modern education, systemized early colonial education by requiring printed tests to determine student achievement. In 1845, under his leadership, the Massachusetts Board of Education employed a comprehensive assessment of student achievement to assess a large school system (Stufflebeam et al., 2000).

In the first half of the twentieth century, Kurt Lewin, a pioneer in action research, contributed to our understanding of how to solve real-world problems with research. Famous for force field analysis, he believed that behavior is a function of the person and their environment.

During the Great Depression in the 1930s, social programs expanded. Unemployment was high and impoverished families struggled. President Franklin D. Roosevelt's New Deal came to their aid through public works projects intended to offer employment assistance (Berkin, 2011) and eventually social programs would become a mainstay in the



**FIGURE 1.4** EVALUATION PROFESSION TIMELINE



Source: Adapted from Stofflebeam, D. L., Madaus, C. F., & Kellaghan, T. (2000). *Evaluation models*. Kluwer Academic.

evaluation field. At the time, however, the main developments in evaluation went hand in hand with changes in education.

Fredrick W. Taylor advanced scientific thinking by focusing on observation, measurement, analysis, and efficiency. Educators added objectives to testing to determine the quality of instruction and the effectiveness of school districts. The goal was to add student grades to provide feedback to the student (Worthen et al., 1997).

Ralph Tyler, in the eight-year study, assessed the outcomes of programs in 15 progressive high schools and 15 traditional high schools. He revised educational objectives to include behaviors. After World War II, student evaluation became more sophisticated, with broadening objectives to include the types of learning teachers expected (Reiser, 2001).

When the Russians launched Sputnik, the first Earth satellite, in 1957, a national discussion of preparedness for war began and questions were asked about whether education was adequately preparing American youth (National Aeronautics and Space Administration, 2018). With national defense and security as the top priority, Congress focused on how students were performing in mathematics, science, and foreign languages. Funding focused on educational outcomes. Low-income children became a focal point, and the expectation was that curriculum evaluation would show what worked and how educational outcomes were improved, especially for low-income children (Weiss, 1972).

In the 1960s, after the assassination of President John F. Kennedy, President Lyndon Johnson continued the war on poverty, introducing the age of the Great Society. As a result, the government “began to take on more responsibility for the general welfare of its citizens...” (Mathison, 2008, p. 184). Not surprisingly, with the investment, concern grew about program outcomes and their return on investment. Program evaluation as a profession grew out of this need. Evaluators soon found work in contract research firms, universities, and public sector offices (Mathison, 2008). In addition, universities began to train students interested in social serving professions.

In the 1970s and 1980s, evaluators began to be professionalized. Examples of early peer-reviewed evaluation journals include *Educational Evaluation and Policy Analysis*, *Studies in Educational Evaluation*, *CEDR Quarterly*, *Evaluation Review*, *New Directions for Program Evaluation*, *Evaluation and Program Planning*, and *Evaluation News* (Stufflebeam et al., 2000).

Two American professional evaluation societies were founded in the 1970s: the Evaluation Research Society (ERS) and the Evaluation Network (ENet). They aimed to foster evaluation as both a profession and a science. In 1986, they merged to form the American Evaluation Association (AEA). Meanwhile, responding to similar societal and governmental pressures, the Canadian Evaluation Society (CES) incorporated in May 1981 (Canadian Evaluation Society, 2018). In 1986, the new AEA held its first evaluation conference in Kansas City, Missouri (American Evaluation Association, 2018b).

Evaluation has grown exponentially as the need for evaluative thinking and accountability has become mainstream around the world. The International Organization for

Cooperation in Evaluation (IOCE), founded in 2001, includes many groups involved in advancing evaluation as a profession. It represents international, national, sub-national, and regional Voluntary Organizations for Professional Evaluation or VOPEs (“vo-pee’s”). Its goal is to strengthen international evaluation by exchanging evaluation methods and promoting good governance while recognizing the value of evaluation in people’s lives. There are currently over 170 VOPEs, large and small, across the globe that are recognized by the IOCE, ranging from the Afghan Evaluation Society (AfES) to the Zimbabwe Evaluation Society (ZES). In total, they represent a membership of almost 52,000 evaluators (International Organization for Cooperation in Evaluation, 2018). As of 2018, the American Evaluation Association (AEA) had approximately 7,300 members representing all 50 states and more than 80 other countries (American Evaluation Association, 2018b).

As Shadish and Luellen commented, social and political climates have influenced evaluation. Funding and types of questions evaluators ask often change with the political tides. However, as they say, “...the need for evaluation seems to be here to stay, as citizens and their representatives prioritize programs and the social problems addressed, as existing programs continue to be scrutinized, and as new and modified programs are proposed” (Mathison, 2005).

Since 2001, there have been several developments in the field of evaluation. Cultural humility, identity and the subjective paradigm, data visualization, and the international growth of evaluation are topics that have blossomed and will continue to influence the work of evaluators.

## THE THREE PILLARS OF EVALUATION

Considering the real-world context in which evaluation occurs, where politics, social need, and financial obligation intersect, the evaluator must resolve challenges, conflicts, and dilemmas. To find guidance for completing our work in the best way possible, evaluators rely on the **three evaluation pillars** (Canadian Evaluation Society, 2010) that support reasoned and ethical conduct:

1. The Guiding Principles (the evaluator’s conduct)
2. The Program Evaluation Standards (the evaluation work)
3. The Professional Competencies (the evaluator’s capacity to do the work)

We will look at each of these important topics.

### The Guiding Principles

An essential component for evaluators is to have a clear ethical stance. **Ethics** is defined as norms for conduct, distinguishing between acceptable and unacceptable behavior (Resnik, 2020). Most people learn ethical norms at home, school, religious institution, or other social setting and ideas about right and wrong are so ubiquitous, one might think they are

simply commonsense. However, while people may recognize common norms, how they interpret, apply and balance them varies widely. Many disciplines and professions have standards for behavior that suit their particular aims and goals. Researchers in general is guided by ethical norms because of the need to promote knowledge and truth, to minimize error and misrepresentation, to instill values of collaboration and trust among co-workers, to be held accountable to the public, to build public trust in findings, and to promote important social and moral values such as social responsibility, human rights, public health, and safety (Resnik, 2020).

Because evaluation takes place in a political and organizational context, general research norms are expanded to include a focus on stakeholders and decision makers, as well as study participants, and to acknowledge the power, values and norms that underlie their relationships during the evaluation process (Barnett & Camfield, 2016). Morris (2008) states that ethical conflict can arise at any point in an evaluation, from project entry, design, data collection, data analysis, interpretation, and communication of findings, to the ultimate use of the evaluation. However, he continues that a significant minority of evaluators claim to have never encountered an ethical conflict. While this is surprising, perhaps it has to do with the lens through which they view their work. Some may interpret an issue as an ethical challenge, and others may consider it a political, philosophical, or methodological dispute.

As evaluators, we view our work through multiple ethical lenses (Barrington, 2012c):

#### ***Values Lens***

We form the values lens early. Our culture, race, religion, beliefs, values, and morals shape our worldview and govern our day-to-day behavior. The evaluator's values at the beginning of a project must be acknowledged. At times these may conflict with the values of the program.

#### ***Methods Lens***

The methods lens results from our research and evaluation training and guides our work processes throughout the life cycle of the evaluation project. Supported by institutional review boards or IRBs, well-designed research protects our human subjects. Three overarching principles include respect for persons, concern for welfare, and justice (Canadian Institutes of Health Research, 2014; U.S. Department of Health and Human Services, 2018).

#### ***Conduct Lens***

The conduct lens relates to how we do our work and involves the interpersonal realm. All our activities must lead to fostering social equity, communicating effectively, and respecting differences. How we behave is how our clients will remember us.

#### ***Business Lens***

The business lens is essential in any evaluation project. Whether a formal contract exists or not, the evaluator expects compensation for their work. Regardless of if

they are a contractor or an employee, each evaluator must manage the evaluation project, reach a mutual understanding about the scope of work, avoid conflict of interest, respect proprietary and confidential information, and be accountable for project requirements.

The American Evaluation Association (AEA) adopted a set of **Guiding Principles for Evaluators** in 1994. These have been reviewed and updated several times, with member consultation, as part of an evolving process of self-examination by the profession, incorporating member feedback (through surveys and town hall meetings), stories evaluators tell about ethical challenges, and changing societal perspectives. For example, topics related to cultural competence were added in 2004. It is the policy of AEA to review the Principles at least every five years, engaging members in the process (American Evaluation Association, 2018a).

As was stated by AEA in 2004:

*The principles are intended to guide the professional practice of evaluators, and to inform evaluation clients and the general public about the principles they can expect to be upheld by professional evaluators. Of course, no statement of principles can anticipate all situations that arise in the practice of evaluation. However, principles are not just guidelines for reaction when something goes wrong or when a dilemma is found. Rather, principles should proactively guide the behaviors of professionals in everyday practice. (p. 3)*

Five interconnected principles, as are shown in **Box 1.2** should govern the behavior of an evaluator at all stages of the evaluation (American Evaluation Association, 2018a).

#### BOX 1.2 ■ AMERICAN EVALUATION ASSOCIATION

##### Evaluators' Ethical Guiding Principles

- A: Systematic Inquiry:** Evaluators conduct data-based inquiries that are thorough, methodical, and contextually relevant.
- B: Competence:** Evaluators provide skilled professional services to stakeholders.
- C: Integrity:** Evaluators behave with honesty and transparency to ensure the integrity of the evaluation.
- D: Respect for People:** Evaluators honor the dignity, well-being, and self-worth of individuals and acknowledge the influence of culture within and across groups.
- E: Common Good and Equity:** Evaluators strive to contribute to the common good and advancement of an equitable and just society.

*Source:* American Evaluation Association. (2018a). *Guiding principles for evaluators and code of ethics*. <https://www.eval.org/p/cm/ld/fid=51>

Each principle is accompanied by several sub-statements to amplify the meaning of the overarching principle and to provide guidance for its application. The Canadian Evaluation Society also has an Ethics Statement which was first published in 1995. It has three statements about competence, integrity, and accountability, each with a subset of statements about acceptable conduct.

### The Program Evaluation Standards

While the Guiding Principles relate to *evaluator conduct*, the **Program Evaluation Standards** are about the *evaluation work* itself and the key attributes that support its quality. In 1975, a broad-based coalition of professional organizations in the United States and Canada created the Joint Committee on Standards for Educational Evaluation (JCSEE), an incorporated public charity (Joint Committee on Standards for Educational Evaluation, 2011). Each of the 15-member groups sent a representative to the Committee to discuss, interpret and determine evaluation standards resulting in *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users* (1981, 1994, 2011). Like the Guiding Principles, the number of editions speaks to evaluators' dedication to and need for these practical guidelines. In addition, scholars and practitioners have scrutinized the standards and input solicited and obtained from national and international reviews, field trials, and national hearings (Yarbrough et al., 2011). Today, the Joint Committee is supported by 17 sponsoring organizations and is a member of the American National Standards Institute. During its 35-year history, its mission has been “to develop and implement inclusive processes producing widely used evaluation standards that serve educational and social improvement” (Yarbrough et al., 2011, p. xviii).

There are five categories of standards and each is accompanied by statements that deal with critical practice issues (see [Table 1.3](#)).

TABLE 1.3 JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION		
Standards	Type	
<b>Utility</b> Ensure that an evaluation will serve the information needs of intended users.	<b>U1</b> Evaluator Credibility <b>U2</b> Attention to Stakeholders <b>U3</b> Negotiated Purposes <b>U4</b> Explicit Values <b>U5</b> Relevant Information	<b>U6</b> Meaningful Processes and Products <b>U7</b> Timely and Appropriate Communicating and Reporting <b>U8</b> Concern for Consequences and Influence
<b>Feasibility</b> Ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.	<b>F1</b> Project Management <b>F2</b> Practical Procedures	<b>F3</b> Contextual Viability <b>F4</b> Resource Use
<b>Propriety</b> Ensure that an evaluation will be conducted legally, ethically,	<b>P1</b> Responsive and Inclusive Orientation <b>P2</b> Formal Agreements	<b>P4</b> Clarity and Fairness <b>P5</b> Transparency and Disclosure

**Table 1.3** (Continued)

Standards	Type	
and with due regard for the welfare of those involved in the evaluation as well as those affected by its results.	<b>P3</b> Human Rights and Respect	<b>P6</b> Conflicts of Interests <b>P7</b> Fiscal Responsibility
<b>Accuracy</b> Ensure that an evaluation will reveal and convey technically adequate information about the features that determine the worth or merit of the program being evaluated.	<b>A1</b> Justified Conclusions and Decisions <b>A2</b> Valid Information <b>A3</b> Reliable Information <b>A4</b> Explicit Program and Context Descriptions	<b>A5</b> Information Management <b>A6</b> Sound Designs and Analyses <b>A7</b> Explicit Evaluation Reasoning <b>A8</b> Communication and Reporting
<b>Evaluation Accountability</b> Ensure the responsible use of resources to produce value by documenting and improving evaluation accountability through internal and external metaevaluation and reflection.	<b>E1</b> Evaluation Documentation <b>E2</b> Internal Metaevaluation	<b>E3</b> External Metaevaluation

Source: Yarbrough, D. B., Shula, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Corwin Press.

Table 1.4 outlines the application of the standards and questions to judge the evaluator's work.

TABLE 1.4 ● APPLYING THE PROGRAM EVALUATION STANDARDS		
Standard	Key Phrase	Questions
Utility	Evaluation has value for stakeholders	<ul style="list-style-type: none"> <li>• Are those evaluated included in the decision making and design of the evaluation?</li> <li>• Is there a good fit between the evaluators and the needs of the evaluation?</li> <li>• Is the interpretation of findings appropriate?</li> <li>• Does the evaluator prepare clear and useful documents?</li> <li>• Are findings disseminated to a broader audience?</li> <li>• Do stakeholders use the evaluation to make decisions related to whether the program should continue?</li> <li>• Is the evaluation transformative for the organization?</li> </ul>
Feasibility	Evaluation is innovative	<ul style="list-style-type: none"> <li>• Does the evaluation not interfere with the usual routine?</li> <li>• Is the evaluator considering the different lenses through which all those concerned may be viewing the issue?</li> <li>• Does evaluation use resources wisely?</li> </ul>

(Continued)

**Table 1.4** (Continued)

Standard	Key Phrase	Questions
Propriety	Evaluation sets high standards	<ul style="list-style-type: none"> <li>• Is there a formal agreement of methodology and operations?</li> <li>• Are human subjects treated ethically?</li> <li>• Do the evaluator and stakeholder interact throughout the evaluation?</li> <li>• Is the evaluation complete and fair?</li> <li>• Are results revealed?</li> <li>• Are conflicts of interest shared?</li> <li>• Is evaluation conducted with fiscal care?</li> </ul>
Accuracy	Evaluation is right on target	<ul style="list-style-type: none"> <li>• Is the evaluation explained with the context in mind?</li> <li>• Does the evaluation serve its purpose?</li> <li>• Are procedures used reliable?</li> <li>• Is evaluation described well?</li> <li>• Does the evaluator employ the use of information management to run the evaluation?</li> <li>• Is evaluation reasoning used?</li> <li>• Does the evaluator protect against misconceptions, biases, distortions, and errors in the documentation?</li> </ul>
Accountability	Evaluation is defensible	<ul style="list-style-type: none"> <li>• Do the evaluators document their work?</li> <li>• Do evaluators consult other sources to support their work?</li> </ul>

*Source:* Adapted from Yarbrough, D. B., Shula, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Corwin Press.

Evaluators often work alone or in small groups and can find themselves on an uneven playing field where the client often wields access, resources, and power. On the other hand, evaluation users or clients often have multiple responsibilities and demands placed on them by their organization and can feel torn serving various interests. As a result, both parties can experience role confusion, bias, and poor communication. The Program Evaluation Standards provide meaningful discussion on each topic, offer recommendations for implementation, identify possible hazards, and present a relevant scenario for discussion. When project difficulties arise, this critical book is a great resource.

Many other organizations and government departments have developed norms of evaluation practice. Examples include the United Nations (2019) Development Program Evaluation Policy and the United Kingdom's Magenta Book (HM Treasury, 2020). Ethical protocols and frameworks have also been developed for planning and evaluating in specific contexts, such as in public health in the United States (Kass, 2001). Others address specific populations such as indigenous peoples, for example, the Aotearoa New Zealand Evaluation Standards (Aotearoa New Zealand Evaluation Association, 2015) that reflect New Zealand's bicultural context, and in Canada, the Tri-Council statement for conducting research involving the First Nations, Inuit, and Métis peoples (2018).



## The Professional Competencies

With all the focus on ethics and standards, it is not surprising that evaluators turned toward the defining the characteristics of a good evaluator. In 2001, Jean King, Laurie Stevahn, and their colleagues began to explore the possibility of a taxonomy of essential evaluator **professional competencies** (King et al., 2001), and by 2005 they had a validated list of competencies. They define competencies as “the background, knowledge, skills, and dispositions program evaluators need to achieve standards that constitute sound evaluations” (Stevahn et al., 2005).

The American Evaluation Association struck a task force between 2015 and 2018 and engaged AEA members in discussing what makes evaluators distinct as practicing professionals. They developed a list of proposed domains and discussed the pros and cons of evaluator competencies, gathering feedback through a survey, World Café-style listening posts, presentations at conferences, focus groups, and requests for feedback on the AEA website. In 2018, the AEA Board adopted the final list of AEA Evaluator Competencies. **Box 1.3** outlines the five domains and the competencies for each (American Evaluation Association, 2018c).

### BOX 1.3 ■ THE 2018 AEA PROFESSIONAL COMPETENCIES

- 1.0 **Professional Practice** focuses on what makes evaluators distinct as practicing professionals. Professional practice is grounded in AEA’s foundational documents, including the Program Evaluation Standards, the AEA Guiding Principles, and the AEA Statement on Cultural Competence.
- 2.0 **Methodology** focuses on technical aspects of evidence-based, systematic inquiry for valued purposes. Methodology includes quantitative, qualitative, and mixed designs for learning, understanding, decision making, and judging.
- 3.0 **Context** focuses on understanding the unique circumstances, multiple perspectives, and changing settings of evaluations and their users/stakeholders. Context involves site/location/environment, participants/stakeholders, organization/structure, culture/diversity, history/traditions, values/beliefs, politics/economics, power/privilege, and other characteristics.
- 4.0 **Planning and Management** focuses on determining and monitoring work plans, timelines, resources, and other components needed to complete and deliver an evaluation study. Planning and management include networking, developing proposals, contracting, determining work assignments, monitoring progress, and fostering use.
- 5.0 **Interpersonal** focuses on human relations and social interactions that ground evaluator effectiveness for professional practice throughout the evaluation. Interpersonal skills include cultural competence, communication, facilitation, and conflict resolution.

*Source:* American Evaluation Association. (2018c). *The 2018 AEA evaluator competencies*.

Evaluator competencies provide a common language and set of criteria to clarify what it means to be an evaluator. The competencies will (American Evaluation Association, 2018c):

- Serve as a roadmap for guiding evaluator education and training
- Encourage members to engage in critical self-reflection about strengths and limitations and find appropriate ways to expand and improve their practice
- Identify ways to improve practice in the field
- Reflect the services evaluators are called upon to perform in multiple contexts
- Recognize the interdependence and overlap of the domains.

They will play a role in moving the field of evaluation toward increased professionalization and improved practice (King, 2020, pp. 7–8). An issue of *New Directions in Evaluation* has documented their development process, but as its editor, Jean King, remarks, “the real work of using the competencies has only just begun” (p. 10).

The Canadian Evaluation Society also began a process to tailor evaluator competencies to the Canadian experience, and through research, member consultation, and expert validation, prepared The Canadian Evaluation Society Competencies for Canadian Evaluation Practice in 2010. Acknowledging that the skills and knowledge for any discipline or profession grows and evolves, influenced by new research and changing environmental circumstances, the CES updated the Competencies in 2018 (Canadian Evaluation Society, 2018). They form the basis of the Credentialed Evaluator (CE) designation.

## EDUCATION AND TRAINING

---

Part of what makes evaluation such an interesting profession is that practitioners come from a wide range of fields and work in an expanding set of contexts such as education, health care, public health, government, policy, development, social services, environment, foundations, and both for-profit and not-for profit organizations.

Stufflebeam (2001) and other writers have suggested that the reputation, impact, and longevity of evaluation as a profession is dependent on evaluation practitioners participating in and applying the lessons learned from high-quality educational experiences (LaVelle, 2018, p. xiv). A recent directory of evaluator education programs in the United States revealed that there are 71 master’s and doctoral programs currently operating in the U.S. along with 42 certificate programs (LaVelle, 2018, p. xvi). An inventory conducted in Canada found programs with an emphasis in evaluation are offered by 54 departments in 27 universities across eight Canadian provinces (Hunter & McDavid, 2019, p. 213). American programs are largely offered by university departments of education, educational psychology, and psychology—79.5% of programs were located in these three disciplines (LaVelle, 2014). By contrast, only 21% of evaluation programs were housed in similar departments in Canada; many others were offered in medicine, health, public policy/

administration, management, and social work (Hunter & McDavid, 2019, p. 215). In addition, many disciplines in both countries offer a single course in program evaluation.

Evaluators draw heavily from research methods and could not practice their craft without this methodological expertise (LaVelle & Donaldson, 2015, p. 41) yet despite much being written about the value of qualitative and mixed methods, most programs focus largely on measurement, assessment, and statistics. While methodological and technical expertise are important, evaluators need additional skills in organizational theory and behavior because they spend so much time working in an organizational context. They also need to learn about situational analysis, relationship building, communication, capacity building, negotiation and conflict management, visual data presentation, project and contract management, and reflexivity. Few if any institutions offer courses that focus directly on culturally responsive evaluation (LaVelle, 2018) and other ways of knowing such as Indigenous knowledge, the Black experience, and gendered perspectives.

Many evaluation professionals receive professional development through short courses and workshops offered by professional organizations such as AEA and CES. The Claremont Evaluation Center and The Evaluators' Institute offer short courses and certificate programs. One other popular professional development opportunity, organized by AEA, is held in Atlanta each year and is designed for evaluators, applied researchers, grant makers, foundation program officers, nonprofit administrators, and others, especially in public health.

## PROFESSIONALIZATION

Emerging from the discussion on evaluator competencies is the recent global trend toward the professionalization of evaluation. What are acceptable qualifications, group norms of conduct, best practices, and how do we differentiate between the qualified and the unqualified?

While professionalization has long been a topic of discussion among evaluators, Canada was the first to solidify the process. The CES Credentialed Evaluator (CE) designation was initiated in 2010 and there are now over 500 CEs. The holder of the CE designation must provide convincing evidence of requisite skills, knowledge, and practical experience in five competency areas. Applicants must have a graduate degree or equivalent, two years (or equivalent) of evaluation experience, references, and a portfolio which they complete online that describes their use of the competencies. Their submission is reviewed by members of the CES Credentialing Board, comprised of senior evaluation professionals with at least 25 years of evaluation experience. Once the CE is granted, the evaluator must accumulate and report at least 40 hours of continuing education every three years to maintain the designation. A public registry lists the CEs (Canadian Evaluation Society, 2018).

Other VOPEs have begun to formalize the status of evaluators, including Japan, South Africa, and the United Kingdom, while the European Evaluation Society is promoting a

Voluntary Peer Review process. The IOCE is beginning to provide its members with tools for building professionalization.

As Tucker and King (2020) remind us, focusing on short-term, disconnected approaches to professional development does not leverage the collaborative potential available to evaluators working together across organizational barriers and geographic boundaries. “While university evaluator education programs typically focus on motivated, early career evaluators, VOPEs need to articulate and leverage evidence-based career-long training to support members in continuously developing competency skill sets” (p. 161). We need to build multiple pathways to competency development together.

Next, we introduce you to our Expert Corner in which we interview a distinguished evaluator or theorist. Our first expert is Dr. Jean King.

Do not copy, post, or distribute

## EXPERT CORNER

## Dr. Jean King



Dr. King is Distinguished Teaching Professor Emerita in the Department of Organizational Leadership, Policy, and Development at the University of Minnesota where she served as Director of the Minnesota Evaluation Studies Institute for over 20 years. From 2015 to 2018 she chaired the Task Force that developed AEA's Program Evaluator Competencies.

1. *What attracted you to the topic of evaluator competencies and when did you begin your research on this topic?*  
In the late 1990s, I was teaching a doctoral seminar that reviewed the literature on evaluator competencies. We discussed the difficulties of creating a single set for the field. One evening after class, three students approached me, asking why it was so difficult to develop them since their own complex fields (early childhood education, special education, and teacher education) all had long-standing and well-accepted competencies. They wondered what the challenges were in program evaluation. Building on existing lists of competencies and using the latest editions of the Program Evaluation Standards, AEA's Guiding Principles, and the Canadian Evaluation Society's Essential Skills Series as references, our group began a two-year unfunded effort to create and initially validate a set of evaluator competencies. These were ultimately published in 2001 in an article entitled "Toward a Taxonomy of Essential Evaluator Competencies" (King et al., 2001). After that, I was hooked on competencies.
2. *Since that time, of the five competency domains, namely (1) professional practice, (2) methodology, (3) context, (4) planning & management, and (5) interpersonal, which one has evolved the most? How has it changed and why?*  
The competencies needed for effective practice change constantly; they evolve because the contexts in which we work change. Professional practice evolves as theorists develop new approaches, professional associations revise foundational documents, and the role of evaluation in social justice becomes increasingly visible. Methodology changes as both quantitative and qualitative researchers add new techniques and expand existing methods repertoires. Planning and management skills develop as strategic planners and project managers shape their craft in innovative ways, and the need to interact successfully with a growing number of cultures requires enhanced interpersonal competence. To my mind, a better question is which domain has *not* changed as much as the others. I believe that context has remained the most consistent over time, probably because it relates to "understanding the unique circumstances, multiple perspectives, and changing settings of evaluations and their users/stakeholders" (American Evaluation Association, 2018c). The need for this understanding has not changed.
3. *The topic of evaluator competencies has grown worldwide. What has surprised you the most?*  
I am most surprised by the fact that the competencies movement (and professionalization of our field more generally) has only lately taken off globally. Almost 20 years ago, the research on our initial set of competencies showed that participants agreed on 78% of the proposed competencies, with almost unanimous support for "ethical conduct" and "framing evaluation questions." With such a high level of agreement suggesting that it really was possible to come to consensus on a general set of competencies, I wonder why it has taken nearly two decades for this topic to take hold in a practical way. It is true that several VOPEs (including the Canadian Evaluation Society and the European Evaluation Society) have built voluntary credentialing or review processes that use a set of competencies, but the viable possibility of common competencies for all program evaluators—adapted, of course, to specific settings—remains in the process of becoming.
4. *Finally, do you have any words of wisdom for our readers who are just beginning to explore the field of evaluation?*  
As an evaluator I have found it helpful to keep in mind what I call my four rules of life: (1) never panic; (2) work to solve the problem; (3) always keep the big picture in mind; and (4) as my mother taught me—be nice.

## Key Terms

Evaluand 9	Problem 5	Significance 10
Equality 12	Professional competencies 27	Social justice 12
Equity 12	Program evaluation 6	Stakeholders 11
Ethics 21	Program evaluation standards 24	Three evaluation pillars 21
Guiding principles for evaluators 23	Quality of life 12	TransScientific model of evaluation 7
Merit 10	Research 14	Worth 10

## The Main Ideas

1. In the TransScientific Model of Evaluation, Scriven posits that evaluation is the driving logic for disciplines outside science, such as law, medical practice, journalism.
2. Patton suggests that evaluations answer three questions: What? So what? Now what?
3. Evaluation is an inquiry into a program to answer questions about its merit, worth, significance.
4. Evaluation has three main purposes: (1) to gain an understanding of how program work and the difference they can make to stakeholders, (2) to improve quality of life, (3) to pursue social justice and a more equitable world.
5. Typically, research is used to contribute to a knowledge base; evaluation is used to improve a program.
6. The Guiding Principles cover the evaluator's conduct during their work.
7. The Evaluation Standards set the criteria for the quality of the work.
8. The Professional Competencies lay the groundwork for the evaluator's capacity to do the work.

## Critical Thinking Questions

1. Think of a program you know well (e.g., at your child's school, at work) and look at the questions that evaluation should answer. Which of these questions is the most critical one right now? Why do you feel this way? How can an evaluation help to improve the situation for both individuals and the community?
2. What is the role of a person's environment in health equity?
3. In what type of program evaluation dilemma would you seek out the Guiding Principles? When might you need the Program Evaluation Standards for support?
4. Think of a time in your life where you evaluated a problem. What was the topic or issue? How did you know you needed to make a change? What questions did you ask? Who was involved? What did you do to intervene? What were the results of your change? How do you know your life improved?
5. When is the last time you encountered a situation where you questioned someone else's ethics or your own? What was the setting? Why did you have a concern? Was it an apparent ethical concern, or was there a fine line between right and wrong? How did you settle the issue? Do you have a frame or series of questions by which you judge situations for possible ethical problems?

## Student Challenges

1. **Livability.** Go to the American Association of Retired Person's (AARP) Livability Index website (<https://livabilityindex.aarp.org/>). Enter the zip code of your hometown and see the score. What are the strengths and weaknesses that make up your score? What policies in your hometown are based on evidence-based evaluations? Are your findings surprising or what you expected?
2. **Find an Evaluator.** Go to [www.linkedin.com](http://www.linkedin.com) and search for "evaluators" or "evaluation." Review your results. How many did you find? What credentials do they have? Where in the world do they practice? In which sectors do they work (e.g., epidemiology, education, business, foundations)? Do their backgrounds interest you?
3. **Evaluation Journals.** The next time you need to do a literature review or write a paper for any class, search for articles within journals dedicated to evaluation. Journals include *African Evaluation Journal*, *Journal of MultiDisciplinary Evaluation*, *Canadian Journal of Program Evaluation*, *Evaluation Journal of Australasia*, *American Journal of Evaluation*, *Evaluation: The International Journal of Theory, Practice and Policy*.
4. **Poverty Clock.** Go to the world poverty clock (<https://worlddata.io/portfolio/world-poverty-clock>). This tool provides real-time data on how many residents escape poverty or fall into it in a given time period. Review the heat maps by country. Filter by geography, gender, and age.

## Additional Readings and Resources

1. Bechar, S., & Mero-Jaffe, I. (2013). Who is afraid of evaluation? Ethics in evaluation research to cope with excessive evaluation anxiety: Insights from a case study. *American Journal of Evaluation*, 35(3), 364–376. Read about a case study where the client did not want the evaluator to write an evaluation report.
2. Donaldson, S., Gooler, L., & Scriven. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation. *American Journal of Evaluation*, 23(3), 261–273. The authors focus on the concept of anxiety induced by being part of an evaluation. First, the authors define the concept and discuss the fear of a negative evaluation and give some common signs and consequences of excessive evaluation anxiety. Finally, the authors pose strategies for managing evaluation anxiety.
3. LaVelle, J. M. (2018). *2018 directory of evaluator education programs in the United States*. University of Minnesota Libraries Publishing. [https://conservancy.umn.edu/bitstream/handle/11299/200790/Directory\\_full.pdf?sequence=5&isAllowed=y](https://conservancy.umn.edu/bitstream/handle/11299/200790/Directory_full.pdf?sequence=5&isAllowed=y) This directory provides all the academic evaluation programs in the United States. Included are certificate, masters, and doctoral-level education programs with an evaluation emphasis. Provides required credits, required course with "Evaluation" in the title, required entry courses, and other requirements.
4. Mason, S., & Hunt, A. (2018). So what do you do? Exploring evaluator descriptions of their work. *American Journal of Evaluation*, 40(3), 395–413. This article shares the way evaluators tell others what they do. Describes the reactions by others when evaluators share their work.
5. Pattyn, V. (2014). Why organizations (do not) evaluate? Explaining evaluation activity through the lens of configurational comparative methods. *Evaluation*, 20(3), 348–367. The author focuses on the reasons why some organizations evaluate policy, and some do not.
6. Wanzer, D. L. (2021). What is evaluation? Perspectives of how evaluation differs (or not) from research. *American Journal of Evaluation*, 42(1), 28–46. doi:10.1177/1098214020920710
7. Patel, M. (2013). African evaluation guidelines. *African Evaluation Journal*, 1(1), 1–5. <https://aejonline.org/index.php/aej/rt/prINTERfriendly/51/67> Fourteen African evaluation associations met to discuss the standards by which they would conduct evaluation. Review their standards. Although they used the American Evaluation Association Programme Evaluation Standards, they worked to "develop a checklist for quality evaluation that as suited to African conditions and culture.

Do not copy, post, or distribute