Chapter 1

Setting the Stage

Supporting the Diverse Knowledge and Skills Evaluators Need in an Ever-Changing Context

Debra J. Rog
Leonard Bickman

TODAY'S EVALUATION CONTEXT

Evaluation is a very diverse profession; evaluators work in a variety of capacities in many different organizations and in a wide range of topical areas. As active evaluation practitioners for more than 40 years, we have watched our field grow and change. We have seen the growth of the number of people engaged in evaluation, the types of organizations that commission and use evaluation, and the kinds of questions addressed by evaluation. In addition, we have witnessed an increasing institutionalization of evaluation in both the United States and globally, and in turn, greater needs for professional development and training.

Today, in the United States, evaluators work in all levels of government (federal, state, local, and regional), contract research organizations, universities, nonprofit and for-profit organizations, and foundations and other philanthropic organizations. For many of these organizations, evaluators are hired internally to conduct evaluations as well as serve as contract officers for evaluations that are conducted by external evaluators. Two examples in the United States that illustrate evaluation's deepening and growing influence in guiding social practice and policy come from both the federal government and philanthropy, two of the largest supporters of evaluation in the United States.

Within the federal government, the Foundations for Evidence-Based Policymaking Act of 2018 (commonly known as **the Evidence Act**) has even more firmly cemented evaluation as a critical part of federal government stewardship. The Evidence Act mandates federal agencies to implement several actions and activities to strengthen its capacity to build and use evidence, including establishing the position of an evaluation officer in the agency, engaging in a capacity assessment to evaluate the agency's ability to implement and use the data from rigorous evaluations, create learning agendas, and develop evaluation plans that include the design and implementation of evaluations (Evidence Act toolkits/Office of Evaluation Sciences; *gsa.gov*). The Evidence Act, including the standards and practices promoted with it, call for a broad range of understanding of **evaluation theory**, designs and methods, and foundational principles (such as ethics).

The federal government has had prior evaluation and data policies (e.g., **Government Performance and Results Act**, the **Program Assessment Rating Tool** review; Hart & Newcomer, 2018) and the practice of evaluation has continued to grow in federal agencies (American

Evaluation Association [AEA] Evaluation Policy Task Force, 2022). However, the Evidence Act distinguishes itself from prior evaluation movements and policies in calling for a strengthening of the government's infrastructure and its workforce to conduct evaluation and other strategies for developing and using data. As the Evidence Act continues to unfold, the federal government is expected to continue to grow and strengthen its evaluation workforce, and as noted by Epstein and colleagues (2022) from the U.S. Office of Management and Budget, "professional development and training will assume even more importance in the coming years" (p. 96).

Although not directly tied to the Evidence Act, additional guidance on developing, analyzing, and using equitable data was issued by the federal government through Executive Order 13985 on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government ("Equity EO"). Three recommendations for equitable data included disaggregating survey data to understand historically underserved groups more fully, increasing public access to disaggregated data, and conducting robust **equity** assessments of federal programs.

Within philanthropy, Long (2017) notes several trends in evaluation's growth and expansion in various roles. One trend in foundations is moving toward a greater focus on accountability and learning, both within an equity framework. Long also reports a trend of foundations embedding evaluation into strategy to inform the evaluability of programs, their implementation, and impact.

Much like the federal government has explicitly described in the Evidence Act, foundations are using evaluation to provide insights for continuous improvement and learning. Of those responding to the Center for Evaluation Innovation (2023) benchmarking survey, two out of three respondents had "learning" in their job title, and over half of respondents had "evaluation" in their job title. Both measures were substantially higher than earlier surveys in 2015 and 2019. In fact, of the responsibilities that learning and evaluation staff prioritize, the top responsibility prioritized by most respondents (90%) is designing and/or facilitating learning processes or events within the foundation.

Within recent years, philanthropies have also taken a stronger equity foothold, assessing not only whether they are being successful but whether they are reaching disadvantaged and marginalized populations and addressing inequities and disparities in the system. Seventy-eight percent of the foundations surveyed in 2023 have organizationwide diversity, equity, and inclusion efforts, up from 56% of the foundations in 2019. Long (2017) notes that a growing focus is infusing into investments the perspectives of community members and other interested parties. Finally, Long notes the stronger relationship between communication and evaluation and the need for evaluators "to create a compelling narrative about *why* an investment worked, *how* it worked, and the tangible *difference* it made in people's lives."

Globally, evaluation has experienced explosive growth. As one measure of growth, over 41,500 individuals are now members of over 220 **voluntary organizations for professional evaluation** (VOPEs; national and subnational associations and societies across 129 countries) and identify as evaluators, either as practitioners, academics, or government or other officials with evaluation-related responsibilities (Rugh, 2018; International Organization for Cooperation in Evaluation [IOCE]). VOPEs help bring together evaluators across these different roles to contribute to advancing the field of evaluation.

PREPARING TODAY'S EVALUATORS

The variety of roles evaluators play and the knowledge and skills they need requires the field of evaluation to ensure that evaluation courses and training programs adequately prepare current and future evaluator generations. A cluster of recent research studies and articles (LaVelle, 2014, 2020; LaVelle et al., 2020; LaVelle & Donaldson, 2021) demonstrate that evaluation training is increasingly available, albeit not as widespread as more traditional disciplines, such as psychology. A review of university-based programs across the world (LaVelle, 2020) found, in 2017, that 87 colleges and universities in the United States were offering evaluation-specific education (two or more courses with "evaluation" in the title) across master's, doctoral, and certificate programs. Fewer programs (27) were identified in colleges and universities outside of the United States, most through master's degree

and certificate programs. The number of programs has increased significantly from 2010 (with a small dip between 2014 and 2017), with the greatest growth in certificate and master's programs.

LaVelle (2020) also conducted a curriculum analysis of these university-based courses and found, consistent with previous research (LaVelle, 2014), that most programs emphasize quantitative skills, such as experimental and quasi-experimental designs, and basic and advanced statistics as well as evaluation theory. Much less common were qualitative evaluation courses, needs assessment, and data visualization. Similarly, a study of the curricula in evaluation courses in Council on Education for Public Health-accredited schools of public health and masters of public health programs (Hobson et al., 2019) analyzed the "essential competencies" for program evaluators as enumerated by Stevahn et al. (2005a, 2005b)[1] and found that the programs focus on professional practice (e.g., how to engage interest holders), systematic inquiry (e.g., design, methods), situational analysis (e.g., describing the program), and interpersonal skills (e.g., communication), but program management and reflective domains were much less covered. Also absent in most courses were **competencies** unique to evaluation, such as evaluability and meta-evaluation.

Program evaluation is also being taught at the undergraduate level (LaVelle et al., 2020). Ninety-one percent of the top 80 public and private U.S. universities offered at least one course, with an average of seven per institution and a total of 470 evaluation-specific or evaluation-related courses taught in the 2017–2018 academic year. Although still a small offering compared to other fields and disciplines, the findings indicate that evaluation principles and tools are being taught to undergraduates, though the extent to which they are being covered is less known.

Finally, evaluation courses are offered through professional development workshops, such as the Evaluators' Institute and the Minnesota Evaluation Studies Institute (LaVelle & Donaldson, 2021). These workshops, offered in person and online, span a variety of introductory and advanced topics. As with the university graduate and undergraduate courses, work is needed to assess whether these courses align with the needs of the field and with the competencies the field believes evaluators need to have.

EVALUATION COMPETENCIES

The growth in evaluation training and professional development is encouraging, but more in-depth understanding is needed of the extent to which evaluators are being adequately prepared for the range of roles in government, philanthropic, nonprofit, and contract research settings, among others. Over the last decade, in the United States and internationally, efforts have been dedicated to developing evaluation competencies (Tucker et al., 2023). Competencies are defined as the totality of knowledge, skills, attributes, behaviors, and attitudes needed to perform the role of an evaluator. In 2015, the AEA appointed a Competencies Task Force (CTF) to develop the competencies, with the aim of generating common language and criteria for distinguishing evaluation as a profession and a practice. As described by King (see Chapter 2, this volume), the CTF worked in a democratic, evolving manner to develop a set of competencies that were finalized in 2018 and published in *New Directions for Evaluation*. After the competencies project's completion, the CTF was sunsetted and, in 2021, a Professionalization and Competencies Working Group was developed to implement the task force recommendations. The work group is now gathering information on how the competencies work in practice and how they relate to the field's professionalization, defined as the process by which an occupation becomes a profession.

The competencies are intended to improve the quality of evaluation performance, the effectiveness of evaluation education, and the shaping of evaluation outcomes to be aligned with social justice (Tucker et al., 2023). The competencies developed to date are considered an initial set, building on foundational documents (the **Program Evaluation Standards**, **AEA Guiding Principles**, and the AEA Statement on Cultural Competence in Evaluation). Forty-nine compe-

[1]Prior to the summer of 2018 and before AEA endorsed professional competencies, these competencies were the most recently published peer-reviewed set of evaluation competencies.

tencies are in the AEA Evaluator Competencies, falling into five domains: professional practice, methodology, context, planning and management, and interpersonal. The domains are ordered according to how directly they are related to evaluation (King & Stevahn, 2020). As noted by those involved in the process, not all competencies apply to every evaluator. Evaluation is often a team endeavor, and others on a team may assume some of the tasks and roles. For example, teams may have specialists in qualitative evaluation and quantitative evaluation. In addition, some competencies may only apply in specific situations. As King notes (see Chapter 2, this volume), "the evaluator competencies provide a framework for thinking about evaluation practice, one that practicing evaluators must examine and shape to their own settings."

AN EVALUATOR'S COMPANION: THE WHAT AND WHY OF THE *HANDBOOK*

We have shaped the content and organization of the *Handbook* on the competencies AEA has identified as critical for evaluators regardless of where they work. These include those professional competencies that distinguish evaluators from other professionals; technical competencies, including designs and methods that are specific to evaluation as well as shared by other professions; practical skills, such as planning and budgeting for evaluation; and crosscutting competencies, such as communication and bringing these different roles for evaluators together in exemplary practice. Although specific sections in this handbook are not outlined for the context and interpersonal domains, many of these competencies are woven through the chapters in the examples provided. In addition, the chapters are written in a "how-to" style to provide pragmatic, apprenticeship-like guidance and often highlight the importance of cultural competence, communication, facilitation, and how to resolve differences and make decisions. All chapters have illustrations and examples, and for those chapters where it makes sense, a set of boxed examples provide more detail illustrating the design, method, or concept.

The *Handbook* is aimed at three audiences: the novice evaluator, professors who teach novice evaluators, and practicing evaluators. The novice evaluator, especially the accidental evaluator who has not had formal training in evaluation, can benefit from resources that prepare evaluators for the variety of roles and tasks they may need to assume. Evaluators report having anxiety in their roles, in part due to a lack of preparation and training in the areas they are being asked to work in (Renger & Donaldson, 2022). Although a handbook will not fully prepare an individual in an area they may have little background, it can provide them a foundation of knowledge to begin to navigate their way.

For professors who teach novice evaluators, the *Handbook* can also provide a guide for the span of competencies required by evaluators and ideally help infuse into their courses content that may have been missing or lean.

Finally, for the practicing evaluator, the *Handbook* provides up-to-date information on key topics they may be familiar with but want refreshing. As evaluators often move across different evaluation settings over time and in different areas, they may need additional information on competencies needed in these roles and contexts that are new to them.

OVERVIEW OF THE *HANDBOOK*

Part I. Foundation for the Volume

The first section is aimed at providing a conceptual grounding for the *Handbook*. This first chapter provides a rationale for a handbook on competencies and how it can be used by evaluation professionals at different stages in their career as well as by evaluation academics in ensuring that novice evaluators are provided instruction in the range of competencies expected of them.

In Chapter 2, Competencies for Program Evaluators, Jean King sets a framework for the *Handbook*. As she notes, "Having a set of carefully developed, comprehensive competencies that detail with some certainty what the practice of evaluation includes can help people focus on the knowledge, skills, abilities, and other characteristics that they need to become competent practitioners." King presents a background of the field of evaluation that captures why a focus on competencies is needed, what they are and what they are not, and how they can be used by novice evaluators as well as those who have

been practicing evaluation for many years. She provides a brief history of the discussion of competencies and certification in evaluation and the reasons why there has been reluctance to develop them. The bulk of King's chapter describes how the current AEA competencies have been developed, as well as some context for competencies in other networks.

The remaining sections of the *Handbook* organize the chapters into four main groupings of competencies: Part II: Theories, Foundations, Principles, and Purpose; Part III: Answering Evaluation Questions: Designs, Methods, and Analyses; Part IV: Planning, Managing, and Implementing Evaluations; and Part V: Cross-cutting Issues.

Part II. Theories, Foundations, Principles, and Purpose

In this section, we commissioned chapters that highlight features of evaluation that are unique to the practice of evaluation: fostering **evaluative thinking**, using theories specific to evaluation to guide our work, having expertise in systematically deriving the value of what we are evaluating, centering equity in our work so that our studies are fair and just, anticipating and addressing ethical challenges that emerge in the work, taking specific actions that can foster evaluation use, and guiding the ways in which we design our studies with certain principles of practice. The last chapter in this section, by Michael Quinn Patton, describes the variation in evaluation practice that this chapter began with, and illustrates how evaluator's competencies are applied differently depending on the context in which the evaluator is working.

Tom Archibald, Jane Buckley, and Guy O'Grady Sharrock begin this section with Chapter 3, Evaluative Thinking: Understanding and Applying the Foundations of Evaluation. Evaluators have increasingly recognized evaluative thinking as a foundational philosophical concept in evaluation. In this chapter, the authors provide several definitions of evaluative thinking and distinguish the role that evaluative thinking plays in evaluation and in **evaluation capacity building** (ECB). As they note, evaluative thinking is critical thinking, fueled by curiosity and the quest for evidence. It involves a process of "identifying assumptions; posing thoughtful questions; marshaling evidence to make judgments; pursuing deeper understanding; and making logically aligned, contextualized decisions in preparation for action." The authors illustrate how these five components of the definition can guide the application of evaluative thinking in evaluation, incorporating reflection and multiple perspectives as they are conducted. From their own work in applying evaluative thinking in ECB, the authors outline six guiding principles for promoting evaluative thinking, including seizing opportunities that naturally occur to engage learners, incorporating it incrementally, offering opportunities to intentionally practice it, providing opportunities for identifying and questioning assumptions, offering opportunities for all to develop as evaluative thinkers and incorporating it into critical conversations involving all levels of staff, and fostering a psychologically safe and trusting environment for evaluative thinking to occur.

Mel Mark follows in Chapter 4, Evaluation Theories: Guidance to Evaluating in Various Circumstances. Theory, as Mark convincingly describes, is the evaluator's navigational tool. It guides how we do evaluation and why we do it in particular ways. It provides a basis for us to choose the methods and techniques we use in each evaluation. Mark explores the role of evaluation theory, defining what it is and the importance of theory for evaluation practice. He examines individual theorists and theories, taking a deep dive on three theories, and presents two meta-models or frameworks designed to help make sense of a multitude of evaluation theories, and how these meta-models can guide the study of evaluation theory. Mark discusses how evaluators can draw on evaluation theories as helpful guides to evaluation practice, as well as how evaluation theory might be modified or added to in the future to further influence practice.

One of the competencies of professional practice is using systematic evidence to make **evaluative judgments**. As Emily Gates and Tom Schwandt describe in Chapter 5, Valuing in Evaluation, professional evaluators bring facts and values together to make these judgments. They define **valuing** as the process of reaching warranted conclusions about value and identify it as both a responsibility and an expertise claimed by professional evaluators. The authors

explain in detail this expertise and responsibility, with a particular focus on how the process of valuing should be systematic, transparent, and defensible. As they note, valuing is one of the features that distinguishes informal, day-to-day judgments of value from professionally conducted evaluation. The authors discuss a variety of values and how they influence evaluation practice, steps in valuing (i.e., selecting criteria, identifying sources of evidence, setting **standards**, and synthesis) and methods to inform them, and a checklist and questions for guiding the evaluator in navigating practical challenges in valuing.

Equity is at the center of societal discourse today and within evaluation. As Donna Atkinson describes in Chapter 6, Equity in Evaluation, evaluators have a responsibility to consider equity in the design of program evaluations to ensure that they are fair and just. To do that, we need to better understand what equity is and how to include it in our work. Atkinson aims to increase evaluators' awareness of different frameworks of equity and what it means to include it in evaluation, how to think about one's own perspectives on the topic and how those perspectives affect one's work, factors that may influence the overall approach to the evaluation when including equity, technical design topics that need to be addressed when including equity in evaluation, and strategies to incorporate equity to consider when planning an evaluation. As Atkinson states, "there is no single approach to including equity in an evaluation." As she notes, "The key to achieving equity in your evaluations is to think about it early and often, seek the support you may need, and conduct a technically solid evaluation that can determine the merits of a program *for all participants*."

In Chapter 7, Ethical Challenges, Michael Morris introduces the concept of **ethical practice** within the context of evaluation. He explores the major ethical challenges that evaluators report encountering throughout the various stages of evaluation (entry/contracting, designing the evaluation, data collection, data analysis and interpretation, reporting the results, and utilization of findings). Morris outlines professional resources evaluators can use to guide them ethically in their work (the AEA Guiding Principles, the Program Evaluation Standards), and offers strategies for planning and conducting evaluations in a way that facilitates the prevention of, and effective response to, ethical difficulties (such as establishing a foundation for dealing with ethical issues in the contracting stage, consulting with colleagues, etc.).

A foundational purpose for evaluation is to inform decision making. In Chapter 8, Fostering Evaluation Use, Marv Alkin and Anne Vo provide clear guidelines for fostering evaluation use, guided by the firm belief that evaluation use typically comes from the specific actions the evaluator takes to foster it. Extracting and examining the necessary elements for a comprehensive evaluation use theory, the authors develop guidelines for fostering use based on reviewing research on use, then identifying the evaluator actions that correspond to the research. A unique and strong "how-to" element of the chapter is the provision of examples of observable actions that both demonstrate that the evaluator action was taken and provide the novice evaluator with guidance on how to take those actions to foster change. The authors end with some practical advice to readers on how best to use the framework, either to guide their work, to teach others, or to stimulate additional research to enhance use.

Chapter 9, Illuminating Evaluation's Kaleidoscope: Beautiful, Diverse, Ever-Changing Manifestations, by Michael Quinn Patton, rounds out this section by focusing on the principles that guide evaluation practice. Through the metaphor of a kaleidoscope, Patton discusses the different roles or forms that evaluation takes: as social science method, profession, discipline, transdiscipline, science, technology, art, and evaluation practice. Using the kaleidoscope as an analogy, he illustrates how the elements of practice can combine in different ways with certain elements having more dominance than others depending on the context one is in. The kaleidoscope provides a basis for understanding the variation within evaluation discussed in the front section of this chapter and why certain aspects are more or less important in specific settings.

Part III. Answering Evaluation Questions: Designs, Methods, and Analyses

Evaluators are asked to address a wide range of evaluation questions that, in turn, need a va-

riety of systematic inquiry approaches. In this section, the chapters offer designs, methods, and analysis strategies to answer several questions: Is the program ready to be evaluated? How is the program being implemented? How do we involve interest holders in the design and implementation of the evaluation? What are the outcomes of the program and other changes that are taking place? What is the program's impact and cost-effectiveness?

David Fetterman, Liliana Rodríguez-Campos, and Ann Zukoski begin this section with Chapter 10, Interest Holder Involvement Approaches to Evaluation: **Collaborative**, **Participatory**, and **Empowerment** Evaluation. The authors focus on how to engage communities in an evaluation to improve programs, build capacity, and produce credible results, especially in culturally diverse contexts and communities. They describe varying approaches that fit different program contexts and circumstances and focus on ways to empower community members to have a voice, if not control, over the design and implementation of an evaluation. The authors describe techniques to achieve involvement that are both effective and practical and offer guidance to evaluators in selecting the approach (collaborative, participatory, or empowerment) best suited to the context.

One question an evaluator should ask before embarking on an evaluation is "Is the program ready for an evaluation?" As Debra J. Rog discusses in Chapter 11, Making the Most of Evaluations: Strategies for Assessing Program Evaluability and Evaluation Feasibility, if an evaluation is conducted too soon, the program may not be sufficiently implemented to have an effect even if the conceptual basis of the program is sound. **Evaluability assessment** provides a practical approach to judging whether a program is evaluable, based on several factors that influence a program's readiness for evaluation. This chapter provides readers with a stepwise approach to judging whether an evaluation could be conducted or if an evaluation would be premature and thus not able to establish the merit and worth of a program. Rog also describes that evaluators can use evaluability assessment for more than judging evaluation readiness, including to help develop programs, select sites to use in multisite evaluations, and provide quick information on a program.

Byron J. Powell, Leonard Bickman, and Kimberly E. Hoagwood provide a comprehensive treatment of studying implementation in Chapter 12, Monitoring Program Implementation. In the last decade, the key role of the quality of **program implementation** has become the focus of not only program evaluation but any field that attempts to intervene to solve a problem. Failure to find that a program is effective can be caused by four main factors: (1) poor implementation of the program, (2) poor theory of change or program theory, (3) poor design of the evaluation, and (4) poor implementation of the evaluation. This chapter provides the reader with the perspective and techniques necessary to distinguish among these factors in the design and implementation of the evaluation, focusing on the ability to judge whether the program was sufficiently well implemented to distinguish between a program implementation failure and a program theory failure. The chapters that follow in this section provide guidance on how to design and implement the evaluation to avoid evaluation design and implementation failures.

Chapter 13, Examining Outcomes and Impacts: Designs and Strategies in Theory and Practice, by Laura Peck and Brad Snyder, focuses on the nature of cause and effect as the key consideration of **summative** or **impact** evaluations seeking to determine changes in outcomes attributable to the program. The authors embed this discussion in a classic review of **threats to internal validity** and the importance of **counterfactual comparisons**. Evaluators will benefit from the practical examples the authors provide on design options and strategies for carrying out impact evaluations as well as their discussion of how impact evaluations relate to other concepts discussed in this handbook, such as logic models, resource constraints, and quantitative design and analysis.

Robert Shand, A. Brooks Bowden, and Henry Levin focus on an important, but often neglected, aspect of program evaluation in Chapter 14, Combining Costs and Results: Designs, Strategies, and Analysis. Not only do we as evaluators want to know whether a program was successful (had an impact), as discussed in the previous chapter, but we should know whether it was effective relative to cost. The authors provide evaluators with the basic knowledge to understand how costs are allocated to a pro-

gram and then the variety of methods used to analyze the cost data relative to the program. These methods include **cost-feasibility analysis**, **cost-effectiveness analysis**, **cost-utility analysis**, or **benefit–cost analysis** that all provide ways to help make decisions about the future of the program. The authors believe these methods, when combined with robust effectiveness studies, can increase the usefulness of the evidence for decision makers.

To address the variety of questions that are posed to evaluators, we have included chapters that offer multiple approaches to qualitative, quantitative, and mixed methods studies. In Chapter 15, Evaluation as Storytelling: Using a Qualitative Design, Sharon Rallis and Janet Usinger explain how to use a qualitative approach, **storytelling**, as a systematic, rigorous, criteria-guided evaluation process to develop a credible story about a program. The story can be told from different angles, including developmental, formative, and summative. As these authors illustrate, storytelling brings life to both the program and the theory underlying it, providing detail and context to generate results that enhance understanding of the program and create potential for change based on the evaluation. Rallis and Usinger base their approach on principles of systematic inquiry, which brings standards and discipline to this approach. Based on years of experience applying this approach, the authors guide evaluators in how to use storytelling to paint a compelling picture of the program using various perspectives.

Moving to a quantitative approach to evaluation in Chapter 16, Planning, Data Collection, and Data Preparation for Quantitative Analysis, Erica Harbatkin, Gary Henry, and Lam Pham provide valuable guidance on planning quantitative analysis, and collecting and preparing quantitative data for analysis. Like many of the other chapters in this handbook, this chapter focuses the evaluator on the planning that is necessary and critical before conducting an evaluation. In quantitative evaluations this preparation requires multiple steps. As the authors outline, planning a quantitative study includes the selection of a research design, sampling, power analyses, selection of measures, and other steps, such as data screening and cleaning. In addition, as the authors indicate, quantitative study planning, as with qualitative efforts (as described in Chapter 15), requires collaboration with interest holders to inform planning, including specifying the research questions, outlining the research design, organizing the data acquisition strategy, planning the data analysis approach, and determining how the findings will be communicated.

Preparing quantitative data for analysis is the last step before conducting the analysis. The same authors from Chapter 16, Lam Pham, Gary Henry, and Erica Harbatkin, offer three approaches to conducting the analysis in Chapter 17, Conducting a Quantitative Analysis. In the first approach, which they label "**exploratory**," the evaluator develops an understanding of the patterns of the data to better understand the program and to plan more rigorous approaches to analysis in later steps. The second step, or **descriptive** analysis, attempts to describe the program, the participants, and its context more precisely. In the third step, **correlational** and **causal** analysis, the evaluator analyzes the data to see whether it had the intended effect. The authors provide several examples of each step so that the evaluator has a practical and well-informed perspective on what actions to take at each step.

In the final chapter in this section, Chapter 18, Mixed Methods Design, Tarek Azzam and Natalie Jones discuss the integration of both qualitative and quantitative designs in one evaluation, a design that is preferred by many evaluators. The design's popularity may be based on its potential to answer questions that can't be answered easily in one method. For example, focus groups, a qualitative method, is good for understanding the feelings of participants in depth, but it can't answer questions about how widespread or representative those feelings are; that can best be estimated by representative surveys. Each approach provides strengths and weaknesses to balance any single approach. However, this combined approach requires both resources and skills in both approaches, which can be more complex and demanding than any single approach. The authors walk the reader through the **mixed methods** design process, highlighting the decision points to be considered, including the priority placed on each data source, the timing of when each type of data will be collected, and where in the process the integration of the data from the different sources will occur. Using the design options as a building block, the au-

thors then describe the ways in which different configurations of the decisions made translate into different mixed methods designs. A major strength of this chapter is the clarity in which it provides decision strategies to help the evaluator decide whether a mixed methods approach is desired and whether it is feasible.

Part IV. Planning, Managing, and Implementing Evaluations

Part IV focuses on two important aspects of program evaluation. First, we continue our focus on planning, but we consider factors that are not usually taught in graduate courses or presented in evaluation textbooks. Our focus here is on resources required to be considered in planning an evaluation. We start with an overview of the planning effort, then move to two detailed chapters on how to best use the resources available to plan an evaluation. Second, we examine roles that evaluators have in planning and conducting evaluations from two different but commonly occurring **contexts**. In the first role we consider the evaluator who is internal to the organization in which the evaluation is being conducted. Second, we carefully examine the role of the independent consultant as the evaluator. Chapters 19 and 20 show the strengths and challenges associated with conducting evaluations in these two contexts.

In Chapter 19, Designing and Planning an Evaluation: Beyond Methods, Darlene Russ-Eft provides a valuable overview of the evaluation planning process. According to Russ-Eft, the evaluator needs to consider several stages in planning an evaluation that includes but goes beyond study design. She provides important guidance for each stage, including planning and outlining the scope of the evaluation, the data collection and analysis, the work schedule, the personnel, the budget, and the risks that the evaluator may encounter. Russ-Eft provides some tools and detailed strategies for conducting these stages. Although all stages are important, attending to risks in the planning process may be the least written about but anticipating them in an explicit manner, with a risk matrix as Russ-Eft outlines, can help evaluators avoid disasters and maximize opportunities. The chapter dovetails well with two other chapters in this section, one on resource planning, by Len Bickman (Chapter 21), and the other on conducting evaluations under tight resource constraints, by Michael Bamberger (Chapter 24).

Chapter 20, Logic Models and Program Theory, by Joy Frechtling, provides a different theoretical perspective than what is found in Chapter 4 by Melvin Mark. Chapter 20 focuses on program theories, rather than evaluation theories. Frechtling describes how program theories can frame an evaluation and focus the evaluation questions. Logic models are tied to **program theory** as a way of depicting the causal relationships that are often implicit in the program developer's mind. It fell to the evaluator, almost by default, to develop the program theory to conduct a rigorous theory-based evaluation. Frechtling introduces critical tools that evaluators will need to plan and conduct evaluations based on an explicit understanding of how and why an intervention should lead to predicted outcomes. Importantly, using program theory moves an evaluation of what might be a single program being tested to a representative of the theory underlying the program being tested.

In Chapter 21, Resource Planning, Len Bickman demonstrates how careful planning for the implementation of an evaluation can be as important as the evaluation questions asked, the design and methods used, and the data collected. He discusses four types of resources to consider while planning an evaluation. These resources include the data that are needed, the time available to complete the evaluation, the people and the skills needed for the evaluation to be successful, and the funding needed to support the evaluation until completion. Because evaluations are conducted in dynamic, real-world contexts, the evaluator also must carefully consider the realistic constraints on the planning and implementation of an evaluation, many beyond the control of the evaluator. Before making decisions about the specific design to use and the type of data collection procedures to employ, the evaluator must consider the resources available and the limitations of these resources within the contextual constraints. Bickman details the steps that evaluators need to take to help ensure they have sufficient resources to complete the evaluation. Not only must these factors be considered in planning an evaluation, but they must be monitored during the evaluation for it is unlikely that no matter how carefully the

evaluation was planned, things will be missed and contemporaneous changes will occur that affect the implementation of the evaluation and the resources needed.

As described at the outset of this chapter, some evaluators conduct evaluations within their own organizations, as a member of the staff of the program or the organization delivering the program. In Chapter 22, The Role of an Internal Evaluator, Arnold Love describes a position that has grown in use across the world. Love provides an understanding of what **internal evaluation** is and how it compares with other roles (such as internal audits) and describes the strengths and limitations of the conduct of an evaluation by an internal evaluator in contrast to an external evaluator. He describes the various roles internal evaluators play, including those that are valued and rewarding (such as an honest broker, navigator, management consultant, change agent, information specialist, and evaluation facilitator) as well as those that are negative (spy, number cruncher, archaeologist, publicity agent, quibbler, and terminator). Love bases many of his observations and suggestions on almost 25 years of teaching about internal evaluations as well as his role as an evaluation consultant. The chapter highlights important differences between internal evaluation and other forms of evaluation and ties those differences to specific suggestions and tips on how to manage the conflicting role an internal evaluator may have in their own organization.

In Chapter 23, The Independent Consultant: An Insider's Guide to a Consulting Career, Gail Barrington describes a different role for an evaluator, as an **independent consultant**. Her chapter describes the requirements needed to be successful in this role, which include personality traits, ethics, and business planning and processes to achieve financial stability. Independent consultants, like internal evaluators, have unique pressures in this role. Barrington describes the characteristics of the independent consultant life, often working with small organizations. She notes the need for consultants to have the intellectual capacity to be quick studies and have self-confidence, courage to face many unknowns, adaptability, and resilience. Barrington provides excellent advice and tools for starting up a new consulting practice, finding the right market niche for the evaluator's skills and interests, obtaining new business, and writing proposals. She also provides insightful advice on how to be productive and survive this challenging environment.

The final chapter in this section by Michael Bamberger, Chapter 24, Conducting Evaluations under Budget, Time, and Data Constraints: An International Perspective, provides guidance to evaluators working in less-than-ideal situations, where designing and implementing methodologically sound, meaningful evaluations are challenged by client biases, politics, unexpected budget cuts, unrealistic client expectations, shortened time frames, and denial of access to data or participants. Although Bamberger has developed his "**real-world evaluation**" approach based on his experience in conducting evaluations outside of industrialized countries, the challenges faced are frequently found in evaluations more broadly. The chapter complements Bickman's chapter on resource planning (Chapter 21) under more typical conditions as well as Russ-Eft's chapter on evaluation planning (Chapter 19) by focusing on strategies for addressing budget and resource constraints, time constraints, and data access issues. The tradeoffs Bamberger suggests in dealing with very real constraints make his chapter a practical resource for evaluators grappling with these challenges.

Part V. Crosscutting Issues

In this section, the first two chapters describe competencies that cut across evaluation practice—communications and **data visualization**. The final chapter, serving somewhat as a capstone to the handbook, describes what it means to conduct exemplary evaluation practice; that is, what counts as "competent" practice in different settings and circumstances.

As described earlier, more attention is being placed on the intersection between communication and evaluation, especially in the philanthropic sector. Glenn O'Neil, in Chapter 25, Communicating with Interest Holders, supports and describes a systematic and strategic approach to communications for evaluators that can be tailored to each evaluation context. As he accurately notes, much of the discussion and research on communications focuses on communicating evaluation findings to decision makers

and other audiences. However, throughout an evaluation O'Neil notes that two-way dialogue, engagement, and interaction can heighten the effects of communication on use. He suggests a range of communication formats for presenting findings (text based, interpersonal, and audio–visual), tailoring the amounts and level of information to the needs of various interest holders and audiences. As O'Neil realistically notes, a well thought-out and implemented communication plan can foster evaluation use, but factors such as context, **organizational setting**, **decision-making characteristics**, and the **policy environment** can be countervailing forces also influencing whether results are used.

Data visualization also has emerged as a critical crosscutting skill that evaluators need to be effective, especially as clients of evaluation clamor for efficient, easy-to-follow diagrams and graphs of findings, study and program conceptual frameworks, and study designs and method. In Chapter 26, Information Visualization and Evaluation, Tarek Azzam and colleagues offer a thorough, how-to guide for developing and using different types of data visualization. Azzam and colleagues focus on first understanding the purpose and audience for the communication, and then ground the chapter in good data visualization principles that can guide effective communication. They provide the reader with an iterative process for creating visualizations, using the principles to select design choices and refine them. The authors also highlight other key topics for evaluators, such as developing **logic models**, and then outline future directions important for both novice and seasoned evaluators to consider in their work, including becoming familiar with more tools and types of data visualization, gaining skills to be sensitive to a range of audiences, understanding how to communicate qualitative and mixed methods results visually as well as quantitative findings, and using visualization as an analytic technique in addition to a vehicle for communication.

In Chapter 27, Exemplary Evaluations in a Multicultural World, Stewart Donaldson addresses what it means to conduct an exemplary evaluation and bookends this handbook with Jean King's chapter on evaluator competencies (Chapter 2). Given the diversity of what falls under the heading of evaluation, however, determining the features of what is considered exemplary practice is no small task. Donaldson uses the learnings from his year as president of AEA in 2015, the International Year of Evaluation involving over 80 evaluation conferences focused on strengthening evaluation capacity, as a springboard to outlining common themes to exemplary practice. Donaldson's query into exemplary practice found that our values determine what we see as exemplary, and that values differ by the evaluation context we are in, especially by regions of the world and across cultures. Some of the most common values for exemplary practice were consistent with the Program Evaluation Standards, AEA's Guiding Principles, and AEA's Statement on Cultural Competence. He notes that movements to develop evaluator competences, such as those developed by AEA, and the movement toward credentialing in a couple of countries signal greater emphasis in evaluation in developing the capacity for more exemplary evaluation practice. As evaluators strive for exemplary practice, Donaldson reminds us of the particular importance of evaluation theory and supports calls for more research on evaluation that can both develop more evidence-based theories to guide evaluator decisions and foster exemplary practice.

WHAT'S NEXT ON THE PLAYBILL: EMERGING DIRECTIONS IN EVALUATION

This handbook itself covers a range of topics that align with the AEA competencies. And we acknowledge that many more topics could have been included had we more space. Notable absences are **developmental evaluation**, use of **machine learning** and **natural processing language**, **ethnographic methods**, **complexity and systems evaluation**, **evaluation synthesis**, **performance measurement**, and **virtual data collection**, among others. Some topics are covered in one or two chapters but could have been written in several—for example, we have a chapter on interest holder approaches to evaluation and one on storytelling and qualitative evaluation, but many more chapters could be written on qualitative methods and approaches, analysis, and data collection. Similarly, Atkinson's chapter (Chapter 6) on equitable and just evaluation could have spawned more chapters digging

deeper into social justice and culturally responsive evaluation. We have chapters on randomized and quasi-experimental methods but could have expanded the focus on methods that rely on secondary data sources (such as interrupted time series and synthetic control methods) that are used more prominently in other fields, as well as methods such as single-case designs that show promise in education and other areas. We therefore hope that the *Handbook* provides a starting point for evaluators, giving them some foundational principles, key designs and approaches, and planning and management tools that they can build upon with other resources.

In this section, we look ahead, identifying a few emerging trends we believe are likely to become more dominant in the field of evaluation. We have identified these trends based on our own experiences and knowledge of where the field is, rather than on a comprehensive literature review or a consensus-based process. Several of these trends support one another (e.g., **artificial intelligence** [AI] and real-time evaluation). We view these areas as likely topics, among others, to be added in a second edition of the *Handbook*. We touch on each briefly.

Artificial Intelligence

In the next few years, we expect there will continue to be tremendous growth in AI. This growth will directly impact how evaluations are conducted and conceptualized. AI will affect how both qualitative and quantitative data are collected and analyzed. The way data are collected will be expanded to include chatbots as well as using the "internet of things," which can collect data passively from sensors in the environment. Data analyses will also be influenced by AI in classifying data for qualitative analyses as well as using voice and video data to score emotional state. AI will also accelerate the use of big data, which will require evaluators to become proficient in managing and analyzing larger data sets than they are usually accustomed to (Bickman, 2020; Mason & Montrose-Moorhead, 2023).

Individuals with Lived Expertise on Evaluation Teams

Incorporating **interest holders** in evaluation and participatory approaches to evaluation are not new. However, the U.S. federal government, the largest funder of evaluations, has been slower to take up the trend. Having individuals and communities participate in evaluating the services, systems, and other interventions that affect their lives is likely to continue and become a more traditional approach to evaluation.

Evaluator Involvement in Program Development

Evaluators typically do not have a formal role in the programs they evaluate, with the exception of evaluators involved in developmental evaluation. In some situations, evaluators in a sense "back into" the issue of program development because programs often are not fully conceptualized and ready to be evaluated against their goals. In this handbook, chapters that indirectly address program development focus on evaluability assessment (see Rog, Chapter 11), and logic models and program theories (see Frechtling, Chapter 20) but are not specifically aimed at developing programs from the outset. Program development should mature into its own field, taking advantage of the skills and perspectives evaluators can bring when program and policy initiatives are first being considered and created. In lamenting what he considered the sad state of program evaluation in 1994, one of the luminaries of program evaluation, Lee Sechrest, voiced similar concerns and attributed some of the lack of advances in program development to the absence of "vigorous commitment" of the evaluation community to champion program development.

Providing Real-Time Analyses

Real-time evaluation methods are not new (McNall & Foster-Fishman, 2007) but the interest in them appears to have grown, especially with COVID-19 (e.g., Buchanan-Smith & Morrison-Métois, 2021). These methods, characterized by short time frames of generally less than 6 months, are known by a variety of terms, including *real time*, *rapid cycle*, and *rapid appraisal*. Two common terms used are described below:

- Real-time evaluation gives insight into how a program or intervention is progressing, often as the evaluation is continuing in the field

(INTRAC, 2012). The approach is associated with emergency response or humanitarian interventions, especially during their early stages when there is more flexibility to make changes.

- **Rapid cycle evaluation** generally is aimed at quickly testing program changes or specific program components or modifications. It is a formative evaluation approach that can involve different methods that range in rigor (Atukpawu-Tipton & Poes, 2020).

Rapid, real-time methods support learning, another key direction we have seen evaluation taking that goes beyond accountability purposes to those that guide direction on the intervention under study and broader. Moreover, as we move to using AI and other technological approaches in collecting and analyzing data, expectations for quick turnaround analyses and findings will grow. Evaluators need to become adept with the tools that can help them collect, analyze, and report data efficiently.

Emphasis on Personalization of Services

We expect that the influence of precision medicine and the more general move to the personalization of services will shape the way in which we evaluate health and human services (Johnson et al., 2020). There is growing recognition that the effectiveness of services is limited by the inability of program developers to personalize or design services to properly fit services to the characteristics and needs of the recipient. Among other things, this means a shift in measuring average **effect size** to individualized outcomes. There will be additional emphases to identify causal factors before conducting a randomized experiment using new methods of causal data sciences (Bickman et al., 2016; Saxe et al., 2022).

Evaluation Responsibility in Environmental Sustainability

A number of prominent leaders in evaluation across the globe (e.g., Davidson et al., 2023) have called for evaluators and sponsors of evaluation to consider environmental sustainability in their work. In their *Footprint Evaluation Guide*, Davidson and colleagues provide guidance for feasible and useful ways to infuse environmental sustainability in the planning, managing, and/or conducting of evaluations. The *Guide* shares the hope that through the emerging set of practices and principles evaluators, organizations, and governments can move to having all evaluations become "sustainability inclusive."

FINAL NOTE

As these future trends indicate, evaluation is a growing field and profession that continues to shape and change. In this handbook, we have provided a fundamental set of principles, theories, values, and methods to guide your practice. Whether you are a novice evaluator, an evaluation professor, or a seasoned practicing evaluator, we hope that this handbook becomes a resource on your shelf to strengthen the competencies you need as you move through different evaluation contexts and topics.

We especially encourage those who are novice evaluators to expand on the knowledge provided in this handbook through journals in our field: *American Journal of Evaluation* and *New Directions for Evaluation* (both supported through the AEA), *Evaluation Review, Evaluation and Program Planning, Evaluation and the Health Professionals,* and *Evaluation* (the journal of the European Evaluation Society), and *Canadian Journal of Program Evaluation* (a journal of the Canadian Evaluation Society), as well as through the variety of textbooks and other resources available. We also encourage you to build on your knowledge and fellowship in the profession by joining the evaluation association or network in your country and any local affiliate as well as seek out workshops and trainings on specific topics and methods.

Finally, we launch this handbook with 27 chapters on a range of topics and methods to align with Donald Campbell's belief in the evolution of knowledge through learning, as Mel Mark (2016) reminded us in his blog post for AEA. Quoting Campbell, Mark noted,

> In science we are like sailors who must repair a rotting ship while it is afloat at sea. We depend on the relative soundness of all other planks while we replace a particularly weak one. Each

of the planks we now depend on we will in turn have to replace. No one of them is a foundation, nor point of certainty, no one of them is incorrigible.

REFERENCES

American Evaluation Association Evaluation Policy Task Force. (2022). An evaluation roadmap for a more effective government. *New Directions for Evaluation, 2022*(173), 17–28.

Atukpawu-Tipton, G., & Poes, M. (2020). *Rapid cycle evaluation at a glance* (OPRE Report #2020-152; produced by James Bell Associates). Office of Planning, Research, and Evaluation, U.S. Department of Health and Human Services.

Bickman, L. (2020). Improving mental health services: A 50-year journey from randomized experiments to artificial intelligence and precision mental health. *Administration and Policy in Mental Health and Mental Health Services Research, 47*(5), 795–843.

Bickman, L., Lyons, A., & Wolpert, M. (2016). Achieving precision mental health through effective assessment, monitoring, and feedback processes. *Administration and Policy in Mental Health and Mental Health Services Research, 43*, 271–276.

Buchanan-Smith, M., & Morrison-Métois, S. (2021). *From real-time evaluation to real-time learning: Exploring new approaches from the COVID-19 response*. ALNAP.

Center for Evaluation Innovation. (2023). *Benchmarking foundation learning and evaluation practices 2023*. Author.

Cioffi, C. C., Hibbard, P. F., Hagaman, A., Tillson, M., & Vest, N. (2023). Perspectives of researchers with lived experience in implementation science research: Opportunities to close the research-to-practice gap in substance use systems of care. *Implementation Research and Practice, 4*.

Davidson, J., Macfarlan, A., Rogers, P., Rowe, A., & Stevens, K. (2023). *Sustainability-inclusive evaluation: Why we need it and how to do it. A Footprint Evaluation Guide*. Available at *www.betterevaluation.org/tools-resources/sustainability-inclusive-evaluation-why-we-need-it-how-do-it*

Epstein, D., Zielewski, E., & Liliedahl, E. (2022). Evaluation policy and the federal workforce. *New Directions for Evaluation*, 2022, 85–100.

Hart, N., & Newcomer, K. (2018). *Presidential evidence initiatives: Lessons from the Bush and Obama administrations' efforts to improve government performance*. Bipartisan Policy Center. Available at *https://privpapers.ssrn.com/sol3/papers.cfm?abstract_id=3786814*

Hobson, K. A., Coryn, C. L. S., Fierro, L. A., & Sherwood-Laughlin, C. M. (2019). Instruction of evaluation competencies in Council on Education for Public Health (CEPH)-accredited master of public health (MPH) degree programs. *American Journal of Evaluation, 40*(4), 590–606.

INTRAC. (2012). *Real-time evaluation*. Available at *www.intrac.org/wpcms/wp-content/uploads/2019/05/Real-time-evaluation.pdf*

Johnson, K., Wei, W.-Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., . . . Snowdon, J. L. (2020). Precision medicine, AI, and the future of personalized health care. *Clinical and Translational Science, 14*(1), 86–93.

King, J. A., & Stevahn, L. (2020). Presenting the 2018 AEA evaluator competencies. *New Directions for Evaluation, 2020*(168), 49–61.

LaVelle, J. M. (2014). *An analysis of evaluation education programs and evaluator skills across the world* (Doctoral dissertation, Claremont Graduate University).

LaVelle, J. M. (2020). Educating evaluators 1976–2017: An expanded analysis of university-based evaluation education programs. *American Journal of Evaluation, 41*(4).

LaVelle, J. M., & Donaldson, S. (2021). Opportunities and challenges ahead for university-based evaluator education programs, faculty, and students. *American Journal of Evaluation, 42*, 428–438.

LaVelle, J. M., Sabarre, N., & Umans, H. (2020). An empirical examination of evaluation's presence in the undergraduate curriculum in the United States. *American Journal of Evaluation, 41*(2), 297–310.

Long, M. (2017). *Five trends in learning and evaluation* [Blog post]. Council on Foundations.

Mark, M. (2016, May 23). *Memorial week: Mel Mark on remembering Donald T. Campbell (1996–2016), pioneer in evaluation methodology* [Blog post]. AEA365.

Mason, S., & Montrose-Moorhead, B. (Eds.). (2023). Evaluation and artificial intelligence [Special issue]. *New Directions for Evaluation, 178–179*, 1–134.

McNall, M., & Foster-Fishman, P. (2007). Methods of rapid evaluation, assessment, and appraisal. *American Journal of Evaluation, 28*, 151–168.

Ramirez, G. G., Bradley, K., Amos, L., Jean-Baptiste, D., Ruggiero, R., Marki, Y., . . . Benton A. (2022). *What is lived experience?* Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services.

Renger, J., & Donaldson, S. I. (2022). The prevalence of evaluator anxiety in practice: An empirical examination. *American Journal of Evaluation, 45*(2).

Rugh, J. (2018, September 23). *What are VOPEs? And how do they relate to us as evaluators?* [Blog post]. AEA365.

Saxe, G. N., Bickman, L., Ma, S., & Aliferis, C. (2022). Mental health progress requires causal di-

agnostic nosology and scalable causal discovery. *Frontiers in Psychiatry*, 2471.

Sechrest, L. (1994). Program evaluation: Oh what it seemed to be! *Evaluation Practice, 15*(3), 359–365.

Skelton-Wilson, S., Sandoval-Lunn, M., Zhang, X., Stern, F., & Kendall, J. (2021). *Methods and emerging strategies to engage people with lived experience: Improving federal research, policy and practice*. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services.

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005a). Establishing essential competencies for program evaluators. *American Journal of Evaluation, 26*(1), 43–59.

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005b). Evaluator competencies in university-based evaluation training programs. *Canadian Journal of Program Evaluation, 20*(2), 101–123.

Tucker, S. A., Stevahn, L., & King, J. A. (2023). Professionalizing evaluation: A time-bound comparison of the American Evaluation Association's foundational documents. *American Journal of Evaluation, 44*(3), 495–512.