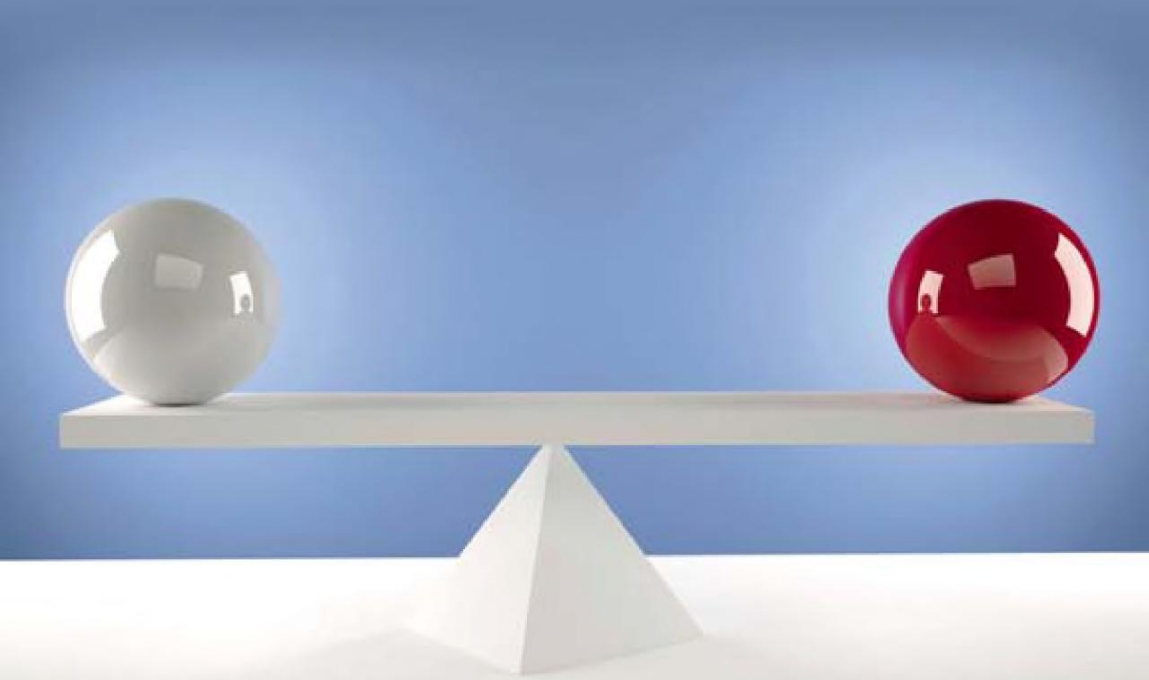
EVALUATION FOUNDATIONS REVISITED

CULTIVATING A LIFE OF THE MIND FOR PRACTICE



THOMAS A. SCHWANDT

EVALUATION FOUNDATIONS REVISITED

Cultivating a Life of the Mind for Practice

Thomas A. Schwandt

STANFORD BUSINESS BOOKS
An Imprint of Stanford University Press
Stanford, California

Stanford University Press
Stanford, California

Special discounts for bulk quantities of Stanford Business Books are available to corporations, professional associations, and other organizations. For details and discount information, contact the special sales department of Stanford University Press.
Tel: (650) 736-1782, Fax: (650) 736-1784

Printed in the United States of America on acid-free, archival-quality paper

Library of Congress Cataloging-in-Publication Data

Schwandt, Thomas A., author.
Evaluation foundations revisited : cultivating a life of the mind for practice / Thomas A. Schwandt.
pages cm
Includes bibliographical references and index.
ISBN 978-0-8047-8655-3 (pbk. : alk. paper)
1. Evaluation. I. Title.
AZ191.S35 2015
001.4—dc23
2014049326
ISBN 978-0-8047-9572-2 (electronic)

Typeset by Bruce Lundquist in 10.9/13 Adobe Garamond

Contents

Acknowledgments

Sincere thanks to members of the QUERIES Seminar in the Department of Educational Psychology at the University of Illinois, Urbana-Champaign, for reading drafts of chapters and offering thoughtful comments. I am especially grateful to Jennifer Greene, Emily Gates, Gabriela Garcia, and Nora Gannon-Slater from that group for sharing their reactions to the overall concept of the book as well as for suggesting ways to improve various chapters. Michael Morris and Leslie Goodyear reviewed an earlier draft and offered valuable suggestions for sharpening the focus of the book. My wife was a very patient and engaged listener as I mulled over ideas and fretted over ways to express myself; I have learned much from her. A very special thanks to my editor at Stanford University Press, Margo Beth Fleming, for her encouragement, editorial advice, and her careful attention to this project since it was first proposed. Many of the ideas expressed here were developed, improved, or refined in conversations with colleagues, far too numerous to mention, over many years. What readers may find of value here I owe in large part to those conversations and to the insights and wisdom of my colleagues. Errors, blunders, and ideas that confound rather than enlighten are of course my own doing.

Evaluation Foundations Revisited

Prologue

Evaluation is the act of judging the value, merit, worth, or significance of things. Those "things" can be performances, programs, processes, policies, or products, to name a few. Evaluation is ubiquitous in everyday life. It is evident in annual performance appraisals, the assessment of a good golf swing, scholarly reviews of books, quality control procedures in manufacturing, and the ratings of products provided by agencies such as Consumer's Union. Forensic evaluations by psychiatrists are a common aspect of many court proceedings; teachers evaluate the performance of their students; expert judges assess and score athlete performances in high school diving competitions; drama critics for local newspapers appraise the quality of local theatre productions; foundation officers examine their investments in social programs to determine whether they yield value for money. Evaluation is an organizational phenomenon, embedded in policies and practices of governments, in public institutions like schools and hospitals, as well as in corporations in the private sector and foundations in the philanthropic world. Evaluation and kindred notions of appraising, assessing, auditing, rating, ranking, and grading are social phenomena that all of us encounter at some time or another in life.

However, we can also identify what might be called a "professional approach" to evaluation that is linked to both an identity and a way of practicing that sets it apart from everyday evaluation undertakings. That professional approach is characterized by specialized, expert knowledge and skills possessed by individuals claiming the designation of "evaluator." Expert knowledge is the product of a particular type of disciplined inquiry, an organized, scholarly, and professional mode of knowledge production.

The prime institutional locations of the production of evaluation knowledge are universities, policy analysis and evaluation units of government departments or international organizations, and private research institutes and firms. The primary producers are academics, think tank experts, and other types of professionals engaged in the knowledge occupation known as "evaluation," or perhaps more broadly "research and evaluation."[1] Evaluation is thus a professional undertaking that individuals with a specific kind of educational preparation claim to know a great deal about, on the one hand, and practice in systematic and disciplined ways, on the other hand. This knowledge base and the methodical and well-organized way of conducting evaluation sets the expert practice apart from what occurs in the normal course of evaluating activities in everyday life. Expert evaluation knowledge, particularly of the value of local, state, national, and international policies and programs concerned with the general social welfare (e.g., education, social services, public health, housing, criminal justice), is in high demand by governments, nongovernmental agencies, and philanthropic foundations.

Focus of the Practice

The somewhat unfortunately awkward word used to describe the plans, programs, projects, policies, performances, designs, material goods, facilities, and other things that are evaluated is *evaluands*; a term of art that I try my best to avoid using. It is the evaluation of policies and programs (also often referred to as social interventions) that is of primary concern in this book. A program is a set of activities organized in such a way so as to achieve a particular goal or set of objectives. Programs can be narrow in scope and limited to a single site—for example, a health education program in a local community to reduce the incidence of smoking—or broad and unfolding across multiple sites such as a national pre-school education initiative that involves delivering a mix of social and educational services to preschool-aged children and their parents in several locations in major cities across the country. Policies are broader statements of an approach to a social problem or a course of action that often involve several different kinds of programs and activities. For example, the science, technology, engineering, and mathematics (STEM) education policy, supported by the U.S. National Science Foundation (NSF) (among other agencies), is aimed in part at increasing the participation of

students in STEM careers, especially women and minorities. That aim is realized through a variety of different kinds of programs including teacher preparation programs; informal, out-of-school education programs; educational efforts attached to NSF-funded centers in science and engineering; and programs that provide research experience in STEM fields for undergraduates.

However, even if we restrict our focus to the evaluation of programs, there remains an incredible variety including training programs (e.g., job training to reduce unemployment; leadership development for youth programs); direct service interventions (e.g., early childhood education programs); indirect service interventions (e.g., a program providing funding to develop small- and medium-size business enterprises); research programs (e.g., a program aimed at catalyzing transdisciplinary research such as the U.S. National Academies of Science Keck Future's Initiative); surveillance systems (e.g., the U.S. National Security Agency's controversial electronic surveillance program known as PRISM or, perhaps in a less sinister way of thinking, a program that monitors health-related behaviors in a community in order to improve some particular health outcome); knowledge management programs; technical assistance programs; social marketing campaigns (e.g., a public health program designed to encourage people to stop smoking); programs that develop advocacy coalitions for policy change; international development programs that can focus on a wide variety of interventions in health, education, food safety, economic development, the environment, and building civil society; strategy evaluation;[2] and programs that build infrastructure in or across organizations at state or local levels to support particular educational or social initiatives.

Aims of the Practice

Evaluation is broadly concerned with quality of implementation, goal attainment, effectiveness, outcomes, impact, and costs of programs and policies.[3] Evaluators provide judgments on these matters to stakeholders who are both proximate and distant. Proximate stakeholders include clients sponsoring or commissioning an evaluation study as well program developers, managers, and beneficiaries. Distant stakeholders can include legislators, the general public, influential thought leaders, the media, and so on. Moreover, policies and programs can be evaluated at different times in their development and for different purposes as shown in Table 1.

Table 1. Types of Evaluation

Implementation evaluation (also called operational evaluation)[a]	Focusing on what happens in the program—its characteristics, participants, staff activities, and so on—to determine what is working and what is not; determining whether the program is operating according to design. Can be undertaken for several purposes including monitoring, accountability (i.e., compliance with program specifications), and program improvement.
Program monitoring	Monitoring/tracking continuously key program data (e.g., participation levels, program completion rates, program costs); often accomplished through the design and operation of a management information system. (A type of implementation evaluation.)
Process evaluation	Examining how the program produces an outcome; searching for explanations of successes and failures in the way the program works; typically requires an examination of how program staff and participants understand and experience the program. (A type of implementation evaluation.)
Outcome evaluation	Focusing on the kinds of outcomes that the program produced in the program participants such as a change in status, behavior, attitudes, skills, and so on; requires the development and specification of indicators that serve as measures of outcome attainment. Outcome monitoring involves examining whether performance targets for program outcomes are being achieved.[b]
Impact evaluation	Examining whether results are actually attributable to the program or policy in question; often focuses not only on the expected and unexpected effects on program participants or beneficiaries attributable to the program (or policy) but also, in the longer run, on the community or system in which the program operates.[c]
Efficiency assessment	Determining whether the benefits produced by the program justify its costs; often determined by using cost-benefit or cost-effectiveness analysis.[d]
Social impact assessment	Measuring the social or public value created by policies, projects, and social interventions.

[a] See Khandker, Koolwal, and Samad (2010: 16–18).

[b] The literature is not in agreement on the definitions of outcome and impact evaluation. Some evaluators treat them as virtually the same, others argue that outcome evaluation is specifically concerned with immediate changes occurring in recipients of the program while impact examines longer-term changes in participants' lives.

[c] Khandker, Koolwal, and Samad (2010); see also the InterAction website for its Impact Evaluation Guidance Notes at http://www.interaction.org/impact-evaluation-notes

[d] Levin and McEwan (2001).

Scope of the Practice

Evaluation of policies and programs in the United States is an important part of the work of legislative agencies and executive departments including the Government Accountability Office, the Office of Management and Budget, the Centers for Disease Control and Prevention, the Departments of State, Education, and Labor, the National Science Foundation, the U.S. Cooperative Extension Service, the Office of Evaluation and Inspections in the Office of the Inspector General in the Department of Health and Human Services, and the Office of Justice Programs in the U.S. Department of Justice. At state and local levels evaluations are undertaken by legislative audit offices, as well as by education, social service, public health, and criminal justice agencies (often in response to federal evaluation requirements).

Professional evaluation of programs and policies is also an extensive global undertaking. Governments throughout the world have established national-level evaluation offices dealing with specific public sector concerns such as the Danish Evaluation Institute charged with examining the quality of day care centers, schools, and educational programs throughout the country or the Agency for Health Care Research and Quality in the U.S. Department of Health and Human Services that investigates the quality, efficiency, and effectiveness of health care for Americans. At the national level, one also finds evaluation agencies charged with oversight for all public policy such as the Spanish Agency for the Evaluation of Public Policies and Quality of Services and the South African Department of Performance Monitoring and Evaluation. These omnibus agencies not only determine the outcomes and impacts of government policies but in so doing combine the goals of improving the quality of public services, rationalizing the use of public funds, and enhancing the public accountability of government bodies.

Evaluations are both commissioned and conducted at think tanks and at not-for-profit and for-profit institutions such as the Brookings Institution and the Urban Institute, the Education Development Center, WestEd, American Institutes of Research, Weststat, and the International Initiative for Impact Evaluation (3ie). Significant evaluation work is undertaken or supported by philanthropies such as the Bill & Melinda Gates Foundation and the Rockefeller Foundation. Internationally, evaluation is central to the work of multilateral organizations including the World Bank, the

Organisation for Economic Co-operation and Development (OECD), the United Nations Educational, Scientific, and Cultural Organization (UNESCO), and the United Nations Development Programme (UNDP); as well as government agencies around the world concerned with international development such as the U.S. Agency for International Development (USAID), the International Development Research Centre in Canada (IDRC), the Department for International Development (DFID) in the United Kingdom, and the Science for Global Development Division of the Netherlands Organization for Scientific Research (NWO-WOTRO).

The scope and extent of professional evaluation activity is evident from the fact that twenty academic journals[4] are devoted exclusively to the field, and there are approximately 140 national, regional, and international evaluation associations and societies.[5] This worldwide evaluation enterprise is also loosely coupled to the related work of professionals who do policy analysis, performance measurement, inspections, accreditation, quality assurance, testing and assessment, organizational consulting, and program auditing.

Characteristics of the Practice and Its Practitioners

The professional practice of evaluation displays a number of interesting traits, not least of which is whether it actually qualifies as a profession—a topic taken up later in this book. Perhaps its most thought-provoking trait is that it is heterogeneous in multiple ways. Evaluators examine policies and programs that vary considerably in scope from a local community-based program to national and international efforts spread across many sites. They evaluate a wide range of human endeavors reflected in programs to stop the spread of HIV/AIDS, to improve infant nutrition, to determine the effectiveness of national science policy, to judge the effectiveness of education programs, and more. This diversity in focus and scope of evaluation work is accompanied by considerable variation in methods and methodologies including the use of surveys, field observations, interviewing, econometric methods, field experiments, cost-benefit analysis, and network and geospatial analysis. For its foundational ideas (for example, the meaning of evaluation, the role of evaluation in society, the notion of building evaluation capacity in organizations), its understanding of the fields in which it operates, and many of its methods, evaluation draws on other disciplines.

The individuals who make up the community of professional practitioners are a diverse lot as well. For some, the practice of evaluation is a specialty taken up by academic researchers within a particular field like economics, sociology, education, nursing, public health, or applied psychology. For others, it is a professional career pursued as privately employed or as an employee of a government, not-for-profit or for-profit agency. There are three broad types of evaluation professionals: (1) experienced evaluators who have been practicing for some time either as privately employed, as an employee of a research organization or local, state, federal, or international agency, or as an academic; (2) novices entering the field of program evaluation for the first time seeking to develop knowledge and skills; and (3) "accidental evaluators," that is, people without training who have been given responsibility for conducting evaluations as part of their portfolio of responsibilities and who are trying to sort out exactly what will be necessary to do the job.[6]

Preparation for the Practice

With only two exceptions (evaluation practice in Canada and Japan), at the present time there are no formal credentialing or certification requirements to become an evaluator. Some individuals receive formal training and education in evaluation in master's and doctoral degree programs throughout the U.S. and elsewhere in the world.[7] However, at the doctoral level, evaluation is rarely, if ever, taught as a formal academic field in its own right but as a subfield or specialty located and taught within the concepts, frameworks, methods, and theories of fields such as sociology, economics, education, social work, human resource education, public policy, management, and organizational psychology.

Perhaps the most significant sources of preparation for many who are "accidental evaluators," or those just learning about the practice while employed in some other professional capacity, are short courses, training institutes, workshops, and webinars often sponsored by professional evaluation associations; certificate programs offered by universities in online or traditional classroom settings;[8] and "toolkits" prepared by agencies and available on the Internet.[9] The rapid development and expansion of these types of training resources are, in part, a response to the demand from practitioners in various social service, educational, and public health occupations to learn the basics of evaluation models and techniques because

doing an evaluation has become one of their job responsibilities (perhaps permanently but more often only temporarily).[10]

This trend to "train up" new evaluators is also consistent with broader cultural efforts to scientize and rationalize the professional practices of education, health care, and social services by focusing directly on measurable performance, outcomes, and the development of an evidence base of best practices. In this climate, evaluators function as technicians who, when equipped with tools for doing results-based management, program monitoring, performance assessment, and impact evaluation, provide reliable evidence of outcomes and contribute to the establishment of this evidence base. However, while providing evidence for accountability and decision making is certainly an important undertaking, the danger in this way of viewing the practice of evaluation is that assurance is substituted for valuation. Evaluation comes to be seen for the most part as one of the technologies needed for assuring effective and efficient management and delivery of programs as well as for documenting achievement of targets and expected outcomes.[11] The idea that evaluation is a form of critical appraisal concerned with judging value begins to fade from concern.

At worst, this trend to narrowly train contributes to the erosion of the ideal of evaluation as a form of social trusteeship, whereby a professional's work contributes to the public good, and its replacement with the notion of technical professionalism, where the professional is little more than a supplier of expert services to willing buyers.[12] Focusing almost exclusively on efforts to train individuals to do evaluation can readily lead to a divorce of technique from the calling that brought forth the professional field in the first place. Furthermore, while learning about evaluation approaches along with acquiring technical skills in designing and conducting evaluation are surely essential to being a competent practitioner, a primary focus on training in models and methods (whether intending to do so or not) can create the impression that evaluation primarily requires only "knowing how."

Rationale for the Book

Taking some license with an idea borrowed from the organizational theorist Chris Argyris, we might characterize evaluation training in "knowing how" as a type of single-loop learning. It is learning problem

solving, how to apply procedures and rules to be sure we are, in a phrase, "doing things right."[13] In single-loop learning we take the goals, values, norms, and frameworks that underlie why we do what we do for granted. Argyris contrasted the reactive mindset of single-loop learning with the productive mindset of double-loop learning. The latter asks, "Are we doing the right things?" and involves examining and reassessing goals, values, and frameworks by asking "why" and "so what" that help improve understanding. Triple-loop learning goes even further by challenging existing learning frameworks as well as mental models and assumptions. The core question is "What makes this the right thing to do?"[14] This kind of learning involves examining one's professional identity and the aims of one's profession as well as how one's frame of reference, style of thinking, and behavior produce both intended and unintended consequences. It is toward the cultivation of double-loop and triple-loop learning in evaluation that this book is addressed.

This aim reflects my longstanding concern that training in technique in evaluation must be wedded to education in both the disposition and the capacity to engage in moral, ethical, and political reflection on the aim of one's professional undertaking. A few years ago, I encountered a wonderful phrase that captures this idea. In 2002–2003, the Carnegie Foundation for the Advancement of Teaching convened a seminar as part of its continuing examination of education for the professions. In that seminar, participants were invited to discuss pedagogical approaches across the disciplines that would help students develop a "life of the mind for practice." Participants defined this notion as "critical thinking through lived situations that demand action with and on behalf of others."[15] Developing "a life of the mind for practice" certainly involves all three types of learning sketched above. However, the phrase particularly draws our attention to the fact that professional practice requires both a capacity for discretionary judgment (critical thinking that issues in appropriate actions) and a clear realization of the social purpose of the practice (action with and on behalf of others).

To possess a life of the mind for evaluation practice is to recognize that professional practice is a matter of designing means for investigating and rendering a judgment of value that will be effective in light of contextual circumstances, political and ethical considerations, client expectations, and resource constraints. It is to realize that assumptions that evaluation practitioners and the public make about the purpose and roles of evaluation,

how and why it is of service to society, the uses to which evaluation knowledge is put, and the evidence employed in arguments about value are almost always contested or contestable and must constantly be revisited in professional practice. A "life of the mind for practice" is required because central to the nature of what we call professional work is the notion that it lacks uniformity in the problems it must contend with and the solutions it must invent. Donald Schön argued that the problems of real-world professional practice are really not problems at all but messy, undefined situations that he called "indeterminate zones of practice—uncertainty, uniqueness, and value conflict—[that] escape the canons of technical rationality."[16] By that he meant that approaches to technical problem solving, the clear-cut application of theories or techniques from the store of professional knowledge, and unambiguous means-end reasoning simply fail to be useful in addressing such problems. To face these kinds of situations and act appropriately, the professional must learn practical reason or the capacity for discretionary judgment.[17]

A vital aspect in the development of a life of the mind for practice is thoughtful, perceptive, and continual engagement with what, for lack of a better term, could be called the contemporary landscape of evaluation issues. These are problems, questions, disputes, and concerns that the professional practice of evaluation, as a particular kind of disciplined inquiry, must always contend with. As one experienced evaluator put it, these issues, "by their very nature, are never finally solved, only temporarily resolved . . . [and they] influence the character of evaluation theory, method, practice, and, indeed, the profession itself."[18]

The landscape encompasses assumptions, principles, values, and ways of reasoning that inform how evaluators define and position their professional activity and argue for the usefulness of their work to society. It includes notions about what comprises evaluative evidence and how that evidence is used to make claims about the merit, worth, or significance of policies and programs. It incorporates perspectives on the ethical conduct, political stance, and professional obligations of evaluation professionals. Moreover, this landscape is inescapably multi- and interdisciplinary with concepts, theories, and empirical findings from social psychology, political science, organizational theory, philosophy, and other fields informing perspectives on key issues. Evaluators inescapably engage features of this landscape as they define and navigate their practice and interact with evaluation funders, clients, and users.

How best to portray this landscape is no small challenge. Several handbooks and other resources provide more encyclopedic coverage than is offered here.[19] My goal is to offer a responsible, but modest, overview in a way that is accessible to novice and accidental evaluators yet inviting of further reflection by seasoned ones. I have taken a cue from an earlier attempt to argue what the field is and ought to be concerned with. In 1991 William Shadish, Thomas Cook, and Laura Leviton offered the first comprehensive effort to map what they referred to as the foundations of program evaluation. They held that their book was "meant to encourage the theoretical dispositions of practitioners by expanding their repertoire of methods, challenging the assumptions behind their methodological and strategic decisions, and creating a broader conceptual framework for them to use in their work."[20] Their expectation was that, upon engaging their analysis of several major theorists' perspectives, readers would be able to better address key questions related to how social programs contribute to social problem solving, why and when evaluations should be done, the evaluator's role, what constitutes credible evaluation knowledge, how judgments of value are to be made, and how evaluation knowledge is used.

While this book is not as thorough an appraisal as theirs, nor organized around the work of evaluation theorists, it shares a similar interest in broadening the conceptual framework of evaluators by portraying a range of key ideas, concepts, and tasks that foreground the professional practice of evaluation as a particular form of disciplined inquiry. Accordingly, I have opted to survey the landscape in such a way that readers can consider the following questions:

- How are we to understand the variability and heterogeneity that characterizes the practice?
- What is the role of theory in evaluation?
- How are notions of value and valuing to be understood?
- What is evaluative evidence and how is it used in evaluative arguments?
- How is evaluation related to politics?
- What does the use of evaluation entail?
- What comprises professional conduct in evaluation, and what might the professionalization of the practice involve?

The term *professional practice* is used intentionally to signify that evaluation is not simply a technical undertaking. Undoubtedly, effective practice requires a wide range of skills for the competent execution of all manner of tasks involved in evaluating including negotiating a contract, preparing an evaluation budget, designing an evaluation, managing an evaluation, choosing or designing instruments and means of data collection, analyzing data, reporting, and so forth. However, in addition to knowing how, evaluation practice requires knowing why. Examining "knowing why" is critical to a life of the mind for practice. The skill of knowing why is cultivated through an engagement with conceptual, practical, and theoretical knowledge that offers resources for answering the questions listed above. Engagement means not simply familiarity with this knowledge, but using it to develop reasoned viewpoints and warranted actions.

The term *disciplined inquiry* is defined much in the same manner as Lee Cronbach and Patrick Suppes used it more than forty years ago. It refers to a systematic, empirical process of discovery and verification, wherein the argument is clear, and the logical processes that link evidence to credible conclusions are apparent.[21] Evaluation is a form of disciplined inquiry concerned with the determination of value. This particular trait distinguishes it from forms of disciplined inquiry that employ social science methods in service of diagnosing, explaining, and solving social problems. Evaluation is concerned with the merits of various approaches to social problem solving and thus shares with those types of inquiry a broad ambition to contribute to social betterment.

Plan of the Book

The chapters aim to present material on the nature and practice of evaluation in such a way as to appeal to two audiences. The expectation is that that those new to the practice might consider what is presented here an invitation to explore the field, while those more seasoned hands might find some new perspectives on familiar material. Chapter 1 briefly introduces the issue of the heterogeneity and variability of evaluation practice by highlighting several fundamental disagreements about the meaning of evaluation, its purpose, methods, and the role and responsibility of the evaluator. Subsequent chapters discuss other ways in which the practice displays significant variations in both its understanding of key concepts and in its self-conception. Chapters 2 through 7 are organized around

three broad objectives: to describe the significant dimensions of the topic in question and indicate key issues; to introduce, but not necessarily resolve, controversies and debates in each area; and to suggest, where appropriate, what appears to be gaining traction as a new issue or direction in professional evaluation practice.

Chapter 2 presents a viewpoint on how theory is related to practice in evaluation and explores ways in which theoretical knowledge can be useful to practice. Chapter 3 discusses issues related to values and valuing in evaluation. It seeks to clarify the multiple ways in which the term *values* figures in evaluation (e.g., as the criterion for determining merit, worth, significance; as the perspective of the individual evaluator; as broadly held social values); it examines the common problem of values disagreement among stakeholders to an evaluation; and it discusses ways in which the activity of valuing is performed in evaluation. Chapter 4 is concerned with how evidence and argument are central to evaluation practice. It discusses the properties of evidence and the character of evaluative judgments. Politics in and of evaluation is the topic of Chapter 5, which explores the multiple ways in which relations between evaluation, politics, and policymaking are understood. Evaluation is inherently a political activity in multiple ways. For example, the very activity of conducting an evaluation involves one in the micro-politics of negotiating contracts, access to and control of data, and navigating among the competing views of stakeholders. At a macro-level, the decision to evaluate is a political act and the value of evaluation as a social enterprise is linked to conceptions of how it best serves a democratic society. Chapter 6 addresses the complicated matter of what it means to use evaluation. It expands upon current ways in which use is discussed in the evaluation literature by bringing to bear what we know from a broader literature on knowledge utilization and the relationship between experts and citizens. Chapter 7 is concerned with issues related to the professional obligations and conduct of evaluators. It takes up issues surrounding professional integrity and discusses efforts of the profession to evaluate its own work. The book concludes with some final thoughts about educational preparation for professional practice.

1 Variability in Evaluation Practice

Best practices have become the most sought after form of knowledge.
Not just effective practices, or decent practices, or better practices—but best.[1]

A focus on best practices, toolkits, practice guidelines, and the like arises fairly naturally in the context of concerns about variation in professional practice. The terms *variation* and *variability*, as commonly used, signify unevenness, deviation, or divergence from norms or standards, discrepancy, or inconsistency.

Every profession is concerned about variability both for reasons of quality assurance and from the standpoint of ensuring that practitioners consistently deliver on the fundamental purpose of the practice regardless of the particular circumstances in which they work.

There are two general views on the nature of variability in professional practice:[2] One holds that variation is a challenge to the rational basis of practice and could be eliminated if practitioners had clear guidelines, including protocols and rules for decision making. Researchers as well as some practitioners who argue that decisions made in practice are too often based on habit and intuition and lack a firm grounding in empirical evidence endorse this view. (One contemporary example of this idea is evident in the book *Moneyball*, the story of Billy Beane, who pioneered the use of sabermetrics—objective knowledge about baseball players' performance based on players' statistics—versus the tradition of relying on the intuitions of a team's scouts to evaluate players.)[3] Concerns about restoring the rational basis of practice are also informed, in part, by nearly fifty years of empirical studies examining whether predictions made by expert clinicians are superior to those made by simple statistical rules or algorithms—about 60% of the studies have shown significantly better accuracy for the statistical rules.[4] Finally, support for this way of thinking

also comes from those who argue that practice ought to be primarily technically based; that is, it should consist of the application of scientifically validated knowledge.[5]

A second view holds that variability is inherent in multiple dimensions of the environment where a practice is performed (think of variation in both resources and the composition of patient populations in hospitals in rural versus metropolitan areas, for example) and thus it is always an aspect of normal practice. In this view, by definition, practice involves flexibility and constant adjustments and modifications. Generally, those who hold this view subscribe to a judgment-based view of practice as comprised of actions informed by situated judgments of practitioners.[6] Rather than encouraging the development of practices that are protocol-driven and rule following, advocates of this view of variability support the idea of developing practical wisdom. They also challenge the idea that intuition often employed by practitioners is an irrational, unscientific process and cannot be improved or sharpened, so to speak.[7] However, this perspective on variability as a normal dimension of practice does not necessarily mean that addressing and managing variation in practice is not a problem.

Responses to practice variation encompass a range of different actions. Evidence-based approaches to practice promote the use of protocols, practice guidelines, and in some cases rules (consider rules regarding nursing care for patients with dementia, for example) for how practitioners should provide services.[8] Another response involves developing performance measures for practitioners based on protocols. In order to identify behaviors or actions considered outside practice norms, a practitioner's performance is measured and compared to standards or targets; in other words, professional practice is audited and the results of the audit fed back to the practitioner to change behavior.[9] Still other responses involve developing lists of best practices and toolkits that, while perhaps not intended to achieve complete standardization of practice, aim to help practitioners operate on some common ground with shared understandings of concepts, methods, ethical guidelines, and so on.[10]

As suggested by the broad description of the field in the Prologue, variation in evaluation practice is common. No doubt, heightened awareness of this state of affairs of the practice and the motivation to address it have been fueled by the evidence-based movement that has developed across the professions of nursing, social work, teaching, counseling, and clinical medicine. At the heart of this movement is the idea that practitioners

ought to use models and techniques that have been shown to be effective based on scientific research. (However, whether what are often touted as best practices are actually backed by scientific evidence is another matter.) Exactly where and when the idea of best practices that originated in the business world migrated to the field of evaluation is not clear, yet the term is no longer a buzzword confined to business enterprises. The literature is full of best practice approaches for evaluating just about everything, including leadership development programs, faculty performance, think tanks, public health interventions, and teacher education programs, to name but a few targets.[11] Moreover, there is a growing sense in the field that although evaluators operate in a world marked by complex contextual conditions, that world "is not so fluid that meaningful patterns cannot be appreciated and used as a basis for action."[12] Hence, in recent years we have witnessed efforts to develop practice guidelines for matching methods to specific evaluation circumstances, as well as guidelines for choosing appropriate means for determining program value in different contexts.[13]

Multiple sources of variability in evaluation practice will be discussed throughout this book. Here, I focus on four primary sources: how evaluation is defined, what methods an evaluator ought to employ, how the professional evaluator relates to and interacts with parties to an evaluation, and how the purpose of the practice is understood.

Defining "Evaluation"

There is no universally agreed upon definition of evaluation, although there are two primary points of view. The first emphasizes that evaluation is an activity concerned with judging value; the second views evaluation as a form of applied research.

In a precise sense—what one would find in dictionary definitions—evaluation refers to the cognitive activity of determining and judging the value of some object, which could be an activity, event, performance, process, product, policy, practice, program, or person. Evaluation is a matter of asking and answering questions about the value of that object (its quality, merit, worth, or significance).[14] The four-step logic involved in doing an evaluation defined in this way is as follows:

1. Select criteria of merit (i.e., those aspects on which the thing being evaluated must do well on to be judged good).